

Chapter 1

Modeling in systems biology

1.1 Introduction

An important aspect of systems biology is the concept of modeling the dynamics of biochemical networks where molecules are the nodes and the molecular interactions are the edges. Due to the size and complexity of these networks, intuition alone is not sufficient to fully grasp their dynamical behavior. Instead an explicit mathematical description of the network and its interaction dynamics is used, which allows for testing and predicting the behavior in computer simulations.

This text starts with an introduction to dynamical systems. It then describes building-blocks used when modeling molecular interactions, and introduces how these 'bricks' are combined into models of large biochemical networks. Then different parameter estimation and analysis methods are discussed. In the end there is a discussion on diffusion and the combination of reactions and diffusion into one model. The text is very sparse and it is meant to be used as lecture notes for both teacher(!) and students.

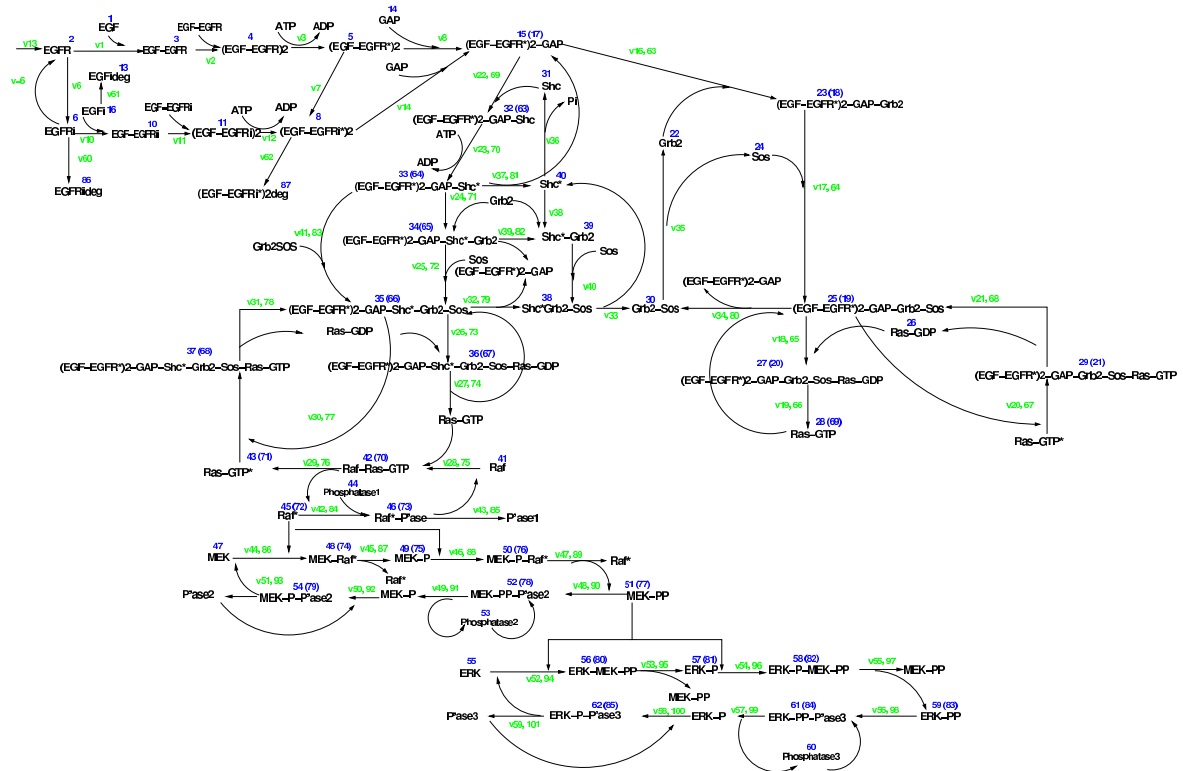
Aims

The main goals for this part of the course are to

1. Understand the concept of modeling dynamical systems. Be able to create a mathematical model of a dynamical system and to do simple analysis of behavior.
2. Learn about some basic building-blocks for describing biochemical interactions (reactions, transcription, ...). Be able to explicitly formulate these interactions as ordinary differential equations.
3. Use the building blocks to create models of a complete biochemical network. Be able to do simulations of such a network in the computer exercise.

4. Be aware of methods for estimating model parameters and tools for model analysis of properties such as robustness.
5. Understand the concept of diffusion and why it can be important when creating models in systems biology. Be introduced to reaction-diffusion models.

Items 2 and 3 will constitute the bulk of these lectures. An important goal is to understand how a model of a large-scale network (as in the figure below) is developed.



Literature

This document is the lecture notes for the dynamics part of the systems biology course, and it is also the course literature. Additional suggested literature and articles will be available in pdf-format at <http://www.thep.lu.se/~henrik/bnf079/literature.html>. The compilation consists of a number of introductory texts and scientific publications, and can be used as references for the interested reader to clarify concepts and to learn more about specific examples.

Contact

Henrik Jönsson, henrik@thep.lu.se, Ph 046-2220663.

1.2 Model building in systems biology

Before building a mathematical model of a biological system, it is important to make some basic decisions on how the model should be defined. Examples of such decisions are:

- **Resolution.** What should be the resolution of the model? What are our model variables representing? Given the recent development of experimental techniques resulting in quantitative molecular data it is now possible to create models describing molecular contents (numbers or concentrations) and compare directly with experiments. In this course we will look at dynamical models of molecular contents describing biochemical networks within single cells and briefly extend the approach to multicellular systems. This choice of resolution would for example not be applicable to explain the evolution of the human population (for which it would be far too detailed) or describe a single molecular reaction mechanism in detail (for which the resolution is too low and a quantum mechanical approach would be needed).
- **Continuous vs. discrete.** Molecules are individual objects, and the quantitative measure of molecular content is in principle number of molecules. On the other hand, the number of molecules (of the same type) within for example a cell is often large and a continuous variable for the concentration (number of variables per unit volume) is then applicable to describe the system behavior. In this course we will use continuous concentrations as variables. The limit for when the concentration is sufficient to describe a system depends on the details of the system but typically when the number of molecules are more than $10^1 - 10^2$ it is safe to use concentration as a measure of molecular content.
- **Deterministic vs. stochastic.** This point is somewhat related to the previous. In principle there is a probability connected to an individual reaction to occur. This can be taken into account using a stochastic update of the system variables (reactions happen with a specific probability). Again, within this course we will assume systems with a large number of molecules where it is applicable to use a deterministic description of the system update.

A modeling approach for a biochemical network includes several different steps or tasks to be solved. The theoretical modeling has to be combined with biological experiments for an effective and useful approach. Important steps within a modeling approach are

1. **Define the molecular players and interactions.** It is a slow and hard process to do experimental work for elucidation of which molecules are involved in different biological processes, and how these interact. The genome projects, where the complete genome of different species are sequenced, have increased the knowledge

about the molecular (protein) players. This provides a list of components, and molecular genomics research are contributing to the knowledge of what biological processes individual molecules are involved in and how these interact. This results in the biological networks that you have come across earlier in the course. Here we assume that this is the input to our modeling approach and will not discuss this part further. It should be noted though, that one main purpose of a modeling approach is to be able to guide these kinds of experiments for an increased understanding of the biological system at hand.

2. **Describe the molecules and interactions in a mathematical model.** To be able to do a quantitative model of a biochemical network, a system has to be defined where molecular concentrations are the variables and their interactions are described explicitly by mathematical functions. These functions depend on the type of interactions that are described, and a main goal in these lectures is to be familiar with common 'translations' for reactions, transcription, etc. into equations, with the ultimate goal of a complete quantitative model for the network.
3. **Estimate parameter values for the model.** The mathematical description includes a number of parameters defining for example reaction rates. Different values of these parameters can result in different behaviors of the model. Hence it is crucial to estimate the parameter values that are relevant for a specific biological network. One possibility is to *experimentally measure* the parameter for a specific reaction, which will result in an optimal single estimated value for the parameter. A potential drawback is that it is hard to do such measurements within a biological organism, and if it is measured elsewhere that specific condition might lead to a different value compared to within the organism. Another approach is *reverse engineering*, where model parameters are estimated by fitting model output to available experimental data. This will exclude most parameter values but still it is possible that this approach will find different values for a single parameter that equally well describe the biological behavior. Within this part of the course we will see how parameter estimations can be done in practice.
4. **Analyse the dynamical behavior.** A final step in a modeling approach is to analyse the behavior of the defined model. Many molecular networks and modules show very high robustness. This should then also be accounted for in the model and can be tested by a sensitivity analysis, where the changes of behavior is tested when parameters are perturbed. Also, the model can be tested for perturbations where molecules or interactions are removed from the system, which then can be compared with knock-out experiments, or provide biological predictions from the model. The analysis can provide feedback into the previous three steps improving the description and knowledge of the biological system.

1.3 Dynamics

Dynamics deals with changes; the evolution in time of a system. It can concern more or less anything from e.g. classical mechanics with an apple falling to the ground, or the growth of the human population. Within systems biology dynamics typically refer to the changes in molecular concentrations (or numbers) within a cell.

A system is defined by i) a set of variables defining the state of the system, and ii) the rules for how the variable values change in time. Variables can be discrete where the state of the variable can be described by a distinct set of values, or continuous where any real value is allowed. The update rules can depend on the time and on the state of all variables. It can be deterministic where the time and variable states uniquely defines the state at next time point, or it can be stochastic where the time and variable state defines the probability of how the variable values changes over time.

The goal when dealing with a dynamical system is to describe and analyse the behavior of the individual variables and also of the complete system, and to be able to make predictions. A dynamical system can be in equilibrium where variables do not change, it can oscillate in a repeating fashion, or it can be more complicated and even chaotic. We will only touch on these subjects briefly, and the interested student can learn more in introductory courses or text books of the subject (e.g. Fys244, System theory, which is given at the Department of Theoretical Physics).

1.3.1 Ordinary differential equations

A fundamental tool for studying dynamics of a continuous system is ordinary differential equations (ODEs). Within this course, we will deal with systems defined as

$$\begin{aligned}\frac{dx}{dt} &= f_x(x, y, \dots, t) \\ \frac{dy}{dt} &= f_y(x, y, \dots, t) \\ &\dots\end{aligned}\tag{1.1}$$

x, y, \dots are the state variables which in our case typically are molecular concentrations, and f_x, f_y, \dots are the functions describing the molecular interactions. The dimension of a system is defined by the number of variables. If the differential equations are given and the initial states (values) of the variables are known, the future behavior of the system is completely defined.

Numerical integration of ODEs

The systems of ODEs for molecular networks are most often too complex to solve analytically and numerical integration is used to simulate the behavior on a computer. There

are many sophisticated algorithms for doing this, but almost all are built from discretizing the differential equation and step forward in time with small steps. The simplest variant of this stepping is the Euler step

$$\begin{aligned}\frac{\Delta x}{\Delta t} &= \frac{x(t + \Delta t) - x(t)}{\Delta t} = f_x(x, y, \dots, t), \quad \text{or} \\ x(t + \Delta t) &= x(t) + \Delta t f_x(x, y, \dots, t).\end{aligned}$$

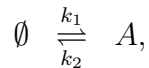
The error introduced by this step is of the order Δt^2 at each step. More accurate solvers will be discussed within a computer exercise.

1.3.2 Behavior of a dynamical system

For one dimensional systems the only possible behavior is that the variable value approaches a specific value, which is defined as a fixed point. The variable might also approach plus (or minus) infinity.

Example: creation and degradation of a molecule

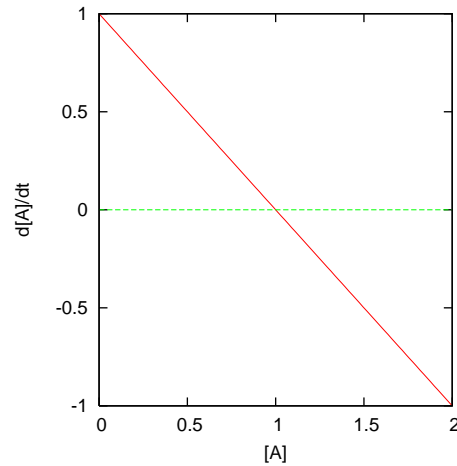
Assume a molecule A which is produced and degraded at a constant rate.



where k_1 is the production rate and k_2 is the degradation rate. The production is assumed to be constant in time (or depend on variables that do not change in time and hence are left outside the model). The degradation rate is assumed to be constant for each individual molecule of A . A differential equation describing this system is

$$\frac{d[A]}{dt} = k_1 - k_2[A], \tag{1.2}$$

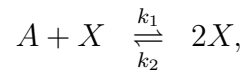
where $[A]$ is the concentration of molecule A . Fixed points of the system can be found by solving the algebraic equation $d[A]/dt = 0$ (i.e. if the system is in such a state it will stay in this state). k_1 and k_2 are assumed to be positive constants, and the only solution is $[A] = k_1/k_2$. A closer look at the time derivative as a function of the concentration of $[A]$ (see figure) resolves more of the dynamical behavior.



Since the derivative is positive when $[A] < k_1/k_2$ and negative when $[A] > k_1/k_2$ the system will always approach k_1/k_2 for infinite times. Since any initial concentration $A(t_0)$ eventually will lead to the fixed point, it is called globally stable. \square

Example: autocatalysis

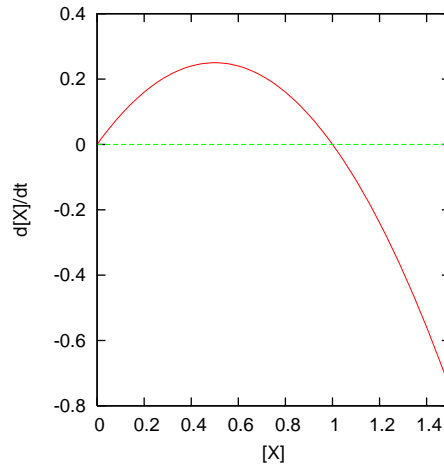
In this example there is a molecule X which induces it's own production mediated by a molecule A .



The law of mass action (which will be discussed in more detail later) states that the rate of a reaction is proportional to the concentrations of the reactants. In this case we assume that there is a surplus of molecule A resulting in that its concentration can be assumed to be constant.

$$\frac{d[X]}{dt} = k_1[A][X] - k_2[X]^2 = K[X] - k_2[X]^2 = [X](K - k_2[X]), \quad (1.3)$$

where the constant $K = k_1[A]$ is introduced. $d[X]/dt$ equals zero for either $[X] = 0$ or $[X] = K/k_2$. A closer look at $d[X]/dt$ as a function of $[X]$ reveals that the fixed points are of different kinds. The $[X] = 0$ fixed point is instable, while $[X] = K/k_2$ is stable (Figure).



The conclusion is that only if we start the system exactly at $[X] = 0$ it will stay there. For any other initial value, the system ends up in $[X] = K/k_2$. A quite interesting note to make is that the equation in this example is exactly the logistic equation used in population dynamics. \square

The two examples show that it is possible to analyse the behavior of a dynamical system without solving the differential equation. We can still predict what will happen in those examples.

For one dimensional systems we can formalize the approach. Given a differential equation

$$\frac{dx}{dt} = f(x) \quad (1.4)$$

1. Find all fixed points x^* by solving $dx/dt = 0$.
2. Investigate the sign of dx/dt around each fixed point to determine the stability. This can be done by plotting it as in the examples, but also by looking at $df(x)/dx$ in the fixed points, where

$$\frac{df(x^*)}{dx} < 0 \rightarrow \text{Fixed point stable}$$

$$\frac{df(x^*)}{dx} > 0 \rightarrow \text{Fixed point instable}$$

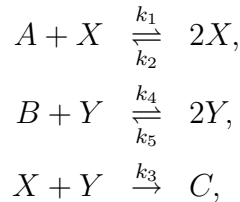
while if the derivative is zero at the fixed point further analysis is needed (it is typically semistable).

It is quite important to note that the behavior (and analysis) depends on the parameter values. Different values can result in different stabilities (e.g. a change from stable to unstable). What will for example happen if d in our first example is negative?

Systems of higher dimensionality can have more elaborate behaviors including oscillations and chaotic behavior. Rigorous analysis of higher dimensional systems are out of scope for this course, but we will briefly address their dynamical behavior by analysing phase plots and nullclines.

Example: two autocatalysing molecules that form a complex

This example is an extension of the previous example, where we now have two molecules X, Y which induces their own production mediated by molecules A and B . X and Y can also form a complex C ($= XY$).

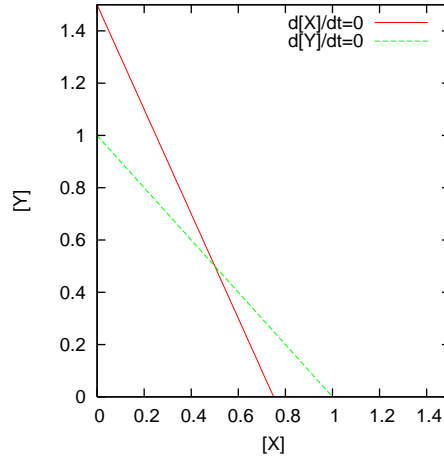


We again assume that there is a surplus of A and B , resulting in their concentrations being constant. Since the dynamics of X and Y does not depend on the complex C , it will also be left out of the analysis.

$$\begin{aligned} \frac{d[X]}{dt} &= k_1[A][X] - k_2[X]^2 - k_3[X][Y] = K_1[X] - k_2[X]^2 - k_3[X][Y] \\ &= [X](K_1 - k_2[X] - k_3[Y]), \\ \frac{d[Y]}{dt} &= k_4[B][Y] - k_5[Y]^2 - k_3[X][Y] = K_2[Y] - k_5[Y]^2 - k_3[X][Y] \\ &= [Y](K_2 - k_5[Y] - k_3[X]), \end{aligned}$$

where the constants $K_1 = k_1[A]$ and $K_2 = k_4[B]$ are introduced. $d[X]/dt$ equals zero for either $[X] = 0$ or $[X] = (K_1 - k_3[Y])/k_2$. These expressions no longer defines specific points but rather defines lines which are defining nullclines. The nullclines for Y are similarly defined by $[Y] = 0$ and $[Y] = (K_2 - k_3[X])/k_5$.

An informative way of representing this system is by plotting the nullclines in the phase space (Figure) which is a plot where $[X]$ and $[Y]$ defines the axes. Now it is easy to see that for example the $[X] = 0$ null cline corresponds to all points on the $[Y]$ axis. Fixed points of the system are found where the nullclines intersect where both $d[X]/dt$ and $d[Y]/dt$ are zero. In the regions in between the nullclines there are non-zero time derivatives (for both $[X]$ and $[Y]$) and by looking at the signs of the derivatives it is possible to analyse the dynamics. It can for example be seen that $d[X]/dt$ is positive beneath the nullcline defined by $[X] = (K_1 - k_3[Y])/k_2$ and negative above.



The conclusion of this analysis is that there are four fixed points $(0, 0), (K_1/k_2, 0), (0, K_2/k_5)$, and $'([X]^*, [Y]^*)'$. The only stable fixed point is at $'((K_1k_5 - K_2k_2)/(k_3(k_5 - k_2)), [Y]^*)'$.

Again it must be noted that this is for the parameter set used, and the behavior can change if for example the nullclines for $[X]$ and $[Y]$ overlaps (the two not defined by the axes). \square

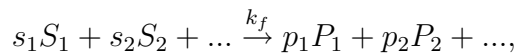
1.4 Biochemical rate equations

In a deterministic continuous formulation, molecular reactions are described by differential equations defining the rate of change in molecular concentrations. Molecular concentrations are most often measured in molar which is defined by mole per liter, where one mole is 6.02×10^{23} molecules. Typical molecular concentrations within a cell are from $0.1nM$ to $1\mu M$ (with lots of exceptions of course).

1.4.1 Mass action formalism

Despite its simplicity, the mass action formalism has been validated in many experimental settings. The law of mass action states that the rate of an elementary chemical reaction is proportional to the product of the concentrations of the reactants. It is based on the assumptions of i) a well stirred solution and ii) low molecular concentrations, where the probability of diffusing molecules to get close enough, for a reaction to occur, is proportional to the concentrations. A rate parameter is used to define the 'probability' of a reaction to occur if two molecules approach each other.

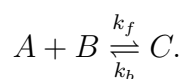
Generally a mass action reaction can be written as



where the variables S_1, S_2, \dots are defining the reactants and the P_1, P_2, \dots are defining the products. The parameters $s_1, s_2, \dots, p_1, p_2, \dots$ are called stoichometric coefficients and k_f is the rate parameter. The stoichometric coefficients are typically chosen such that the total mass is conserved in the reaction (or such that atom numbers are the same before and after the reaction).

Example: a simple mass action reaction

Consider the simple reaction of species A and B forming complex C .



k_f is the rate of the forward reaction of unit $[time]^{-1}[conc]^{-1}$, while k_b is the rate of the backward reaction of unit $[time]^{-1}$. Note that reaction rate units are not uniquely defined, but rather depends on the reaction. In a differential equation formalism the equations are defined by

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -\frac{d[C]}{dt} = -k_f[A][B] + k_b[C], \quad (1.5)$$

which will have an equilibrium point (fixed point) for $[C]/[A][B] = k_f/k_b$ where $K = k_f/k_b$ defines a relation between concentrations of reactants and products which is independent on initial concentrations. K is often defined as the *reaction constant*. \square

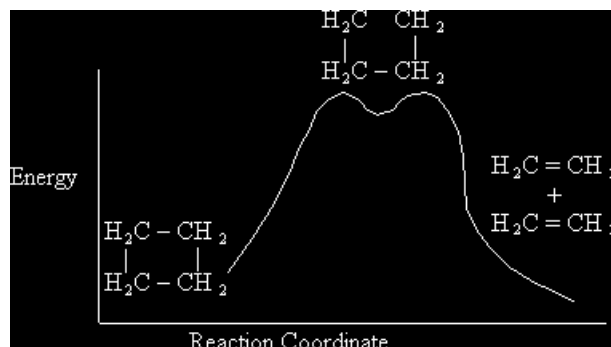
1.4.2 Thermodynamics and rate constants

In experiments it can be seen that the logarithm of the rate constant, $\ln k$, is linearly related to the inverse temperature $1/T$. The parameters for the slope and intercept is formulated in Arrhenius law

$$k = Ae^{-E_a/TR} \quad (1.6)$$

where E_a is the activation energy, R is the gas constant and A is the steric factor, a constant measuring the efficiency of a molecular collision leading to a reaction.

In transition state theory the energy is replaced by the Gibbs free energy, $G = E + PV - TS$, where P is the pressure V is the volume, T is the temperature and S is the entropy. The idea is that the a molecule is in a local minima in a “reaction space”, and that for a reaction to happen, it has to find a path to the product within this space, and a maxima needs to be passed (see figure below). Values for the Gibbs free energy for different molecules can be found in the literature and the reaction constant of a bidirectional reaction can be related to the difference in G .

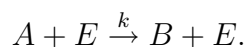


1.4.3 Enzyme kinetics

Many reactions have a far too high activation energy to ever occur spontaneously. A common type of reaction is an enzyme reaction, where a helper molecule (the enzyme) facilitate a reaction to occur. The enzyme is not used up in the reaction itself.

Example: a simple enzymatic reaction

Consider the simple reaction of species A forming compound B with the help of enzyme E .



k is the rate of the reaction of unit $[time]^{-1}[conc]^{-1}$. Using a differential equation formalism the equations are defined by

$$\frac{d[A]}{dt} = -\frac{d[B]}{dt} = -k[A][E], \quad (1.7)$$

$$\frac{d[E]}{dt} = 0. \quad (1.8)$$

The problem with this formulation is that there is no upper limit on how much a single enzyme molecule can facilitate the reaction. Often there is an upper limit on the rate due to the fact that the enzyme is occupied during the reaction, and a model accounting for this is described in the next section. \square

1.4.4 Enzyme kinetics, Michaelis-Menten

A more proper description of an enzyme reaction is to let the enzyme E bind to the substrate S and letting the substrate turn into a product P while the enzyme is released



The rate equations for this system can be written as

$$\begin{aligned}
\frac{d[S]}{dt} &= -k_1[S][E] + k_2[SE] \\
\frac{d[E]}{dt} &= -k_1[S][E] + k_2[SE] + k_3[SE] \\
\frac{d[SE]}{dt} &= k_1[S][E] - k_2[SE] - k_3[SE] \\
\frac{d[P]}{dt} &= k_3[SE]
\end{aligned} \tag{1.10}$$

The first reaction is assumed to be fast (and in equilibrium) and we assume that $d[SE]/dt \approx 0$. Solving the fixed point equation gives $K = k_1/(k_2 + k_3) = [SE]/[S][E]$. If we also assume a constant amount of total enzyme, $[E] + [SE] = E_0$, the complex concentration can be written as a function of the substrate concentration,

$$\begin{aligned}
[SE] = K[S][E] &= K[S](E_0 - [SE]) \\
[SE](1 + K[S]) &= KE_0[S] \\
[SE] &= \frac{KE_0[S]}{1 + K[S]} = \frac{E_0[S]}{(1/K + [S])}.
\end{aligned} \tag{1.11}$$

The production of P as a function of the substrate concentration is then

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]} \tag{1.12}$$

where the constants $V_{max} = k_3E_0$ and $K_m = 1/K$. The choice of parameters is due to the fact that V_{max} is the saturated maximal rate of production and K_m is the amount of substrate that corresponds to half the maximal rate (Fig. 1.1). A problem with the Michaelis-Menten equation is the “slow” response to substrate concentration compared with what is often seen in experiments. To get the rate $0.1V_{max}$ a substrate concentration of $S_{0.1} = K_m/9$ is needed and to get a rate of $0.9V_{max}$, the substrate concentration needs to be $S_{0.9} = 9K_m$. Hence an 81-fold change in concentration is needed between ‘on’ and ‘off’ states. This is often handled by using a Hill-type kinetics as will be discussed in more detail later.

It should also be noted here that the dependence on the enzyme concentration is built into the V_{max} parameter and assumed to be constant. The amount of enzyme is often also a dynamic variable and the reaction can then be described by

$$\frac{d[P]}{dt} = \frac{V'_{max}[S][E]}{K_m + [S]} \tag{1.13}$$

where it is assumed that the concentration of the enzyme changes slowly compared to the change in P .

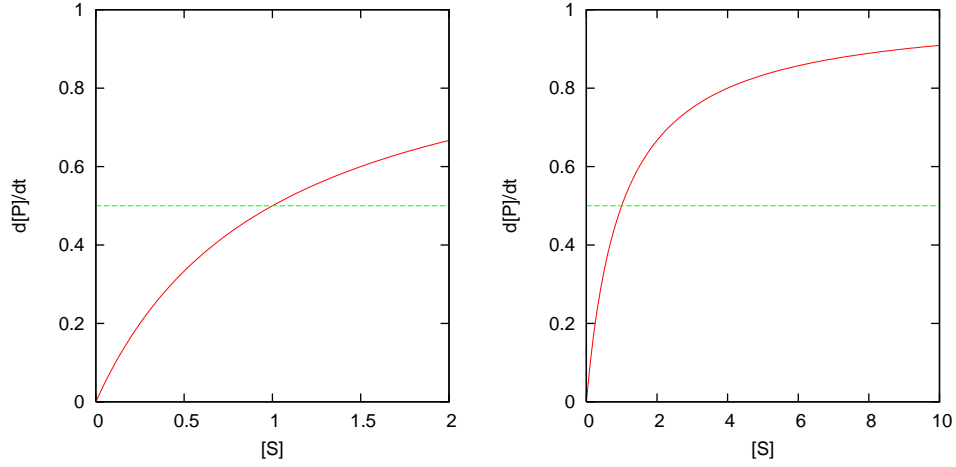
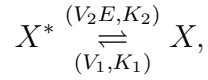


Figure 1.1:

Example: protein activation/deactivation cycle

Previously in the course you have seen the example of a protein that can be activated and deactivated



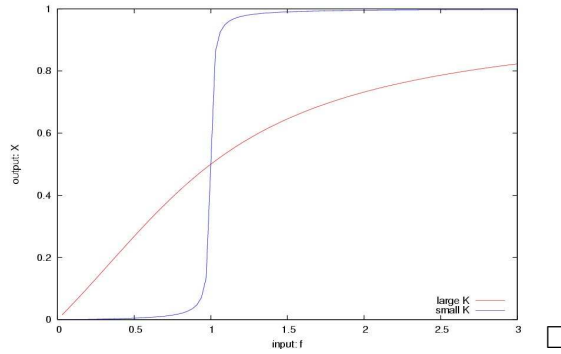
where the total concentration is constant $X^* + X = X_{\text{tot}} = 1$ (or $X^* = 1 - X$). Assuming that both the activation and deactivation are dependent on other molecules (enzymes), and that the activation enzyme is dynamic, result in the following Michaelis-Menten description

$$\frac{d[X]}{dt} = -\frac{V_1[X]}{K_1 + [X]} + \frac{V_2[E](1 - [X])}{K_2 - (1 - [X])}. \quad (1.14)$$

Setting the parameters $K_1 = K_2 = K$ and $f = V_2[E]/V_1$ and investigating the system at equilibrium ($\frac{d[X]}{dt} = 0$) results in the equation

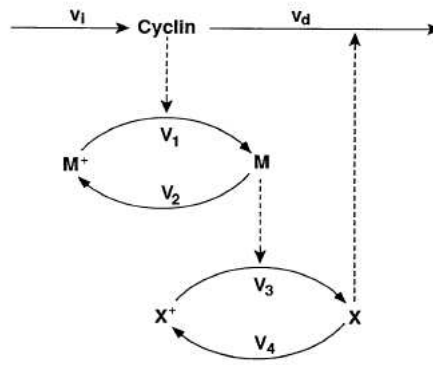
$$\frac{[X]}{K + [X]} = f \frac{(1 - [X])}{K + (1 - [X])}. \quad (1.15)$$

When studying how the activation, $[X]$, is dependent on the input, f , it was shown to behave either as an analogue amplifier or a digital switch depending on the K value, as shown in the figure.



Example: cell cycle

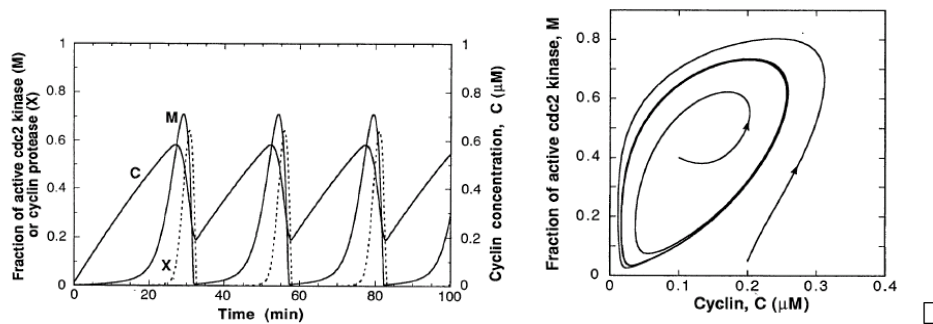
A minimalistic model for the cell cycle was introduced by Goldbeter 1991. It has only three state variables and the interactions are shown in the figure.



In the model, cyclin (C) is produced and degraded at constant rates. The cyclin induces a cyclin kinase (M) to be activated, which in turn activates a cyclin protease (X). Finally the protease induces degradation of the cyclin closing a feedback loop in the system. All reaction kinetics used is in the Michaelis-Menten format. A minor simplification of the equations leads to the following model.

$$\begin{aligned}
 \frac{dC}{dt} &= v_i - v_d X \frac{C}{K_d + C} - k_d C \\
 \frac{dM}{dt} &= V_1 C \frac{(1 - M)}{K_1 + (1 - M)} - V_2 \frac{M}{K_2 + M} \\
 \frac{dX}{dt} &= V_3 M \frac{(1 - X)}{K_3 + (1 - X)} - V_4 \frac{X}{K_4 + X}
 \end{aligned} \tag{1.16}$$

Simulation of the network shows that, for some ranges of parameter values, an oscillatory solution is possible (which also exhibit limit cycle behavior) as can be seen in the figures below.



1.4.5 Models within a cell

The mathematical formulations described in previous sections are simplified and assumes idealized conditions. For example the assumptions of low molecular concentrations and of well-stirred solutions are very unlike the situation in a cell (Fig.1.2).

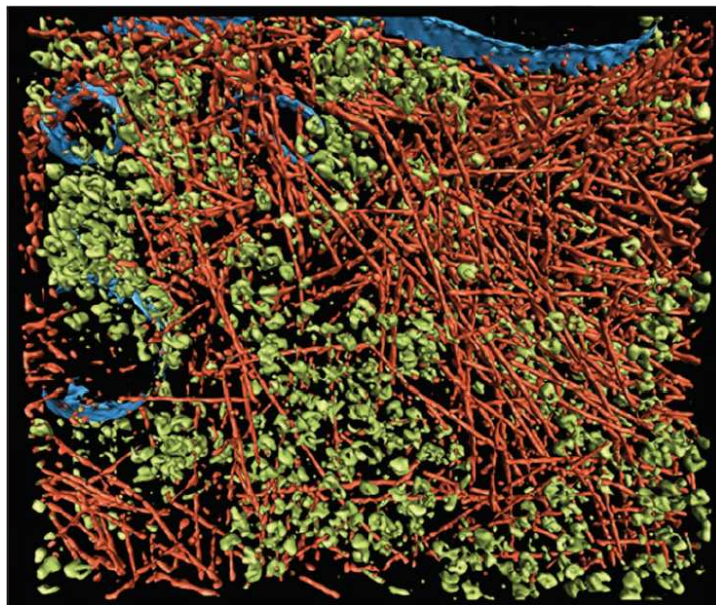
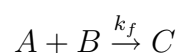


Figure 1.2: Visualization of actin network, membranes, and cytoplasmic macromolecular complexes in a volume of 815 nm by 870 nm by 97 nm. Colors were subjectively attributed to linear elements to mark the actin laments (reddish); other macromolecular complexes, mostly ribosomes (green); and membranes (blue). From Mendalia et. al. (2002), Science 298, 1209-1213. Copyright 2002 AAAS.

Example: generalized mass action

It is often the case that the mass action dynamics deviate from *in vivo* experiments. It might then be useful to “extend” the reaction models to better correlate with experiments. In the generalized mass action approach the concept of activity is introduced. The idea is that the effective concentrations for a reaction can be different from the absolute concentration. Without going into details, the generalized mass action formalism for a simple reaction



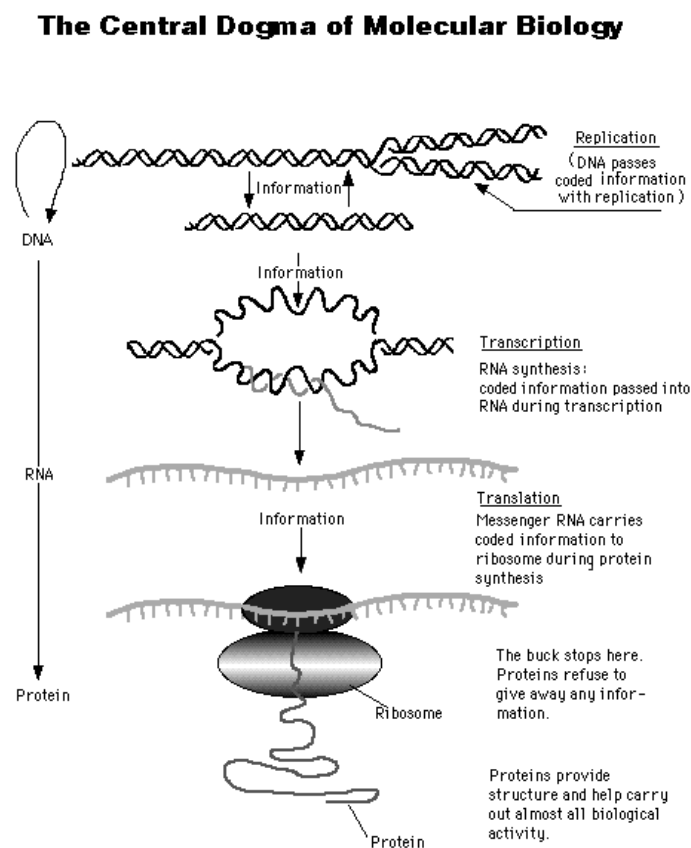
uses a differential equation of the form

$$\frac{d[A]}{dt} = k_f a [A]^\alpha b [B]^\beta \quad (1.17)$$

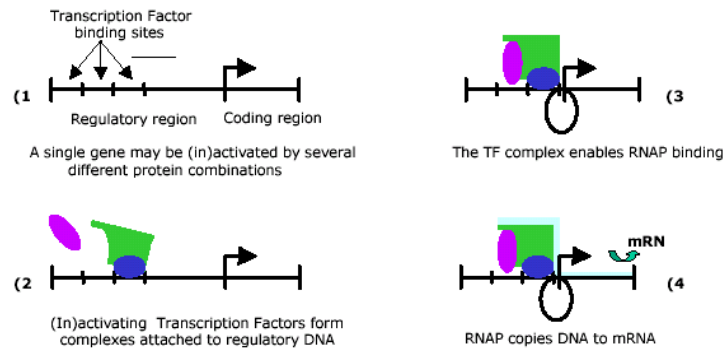
where a , α , b , and β are (real-valued) parameters. The generalized mass action hence allow for additional possibilities of dynamical behavior compared to classical mass action. \square

1.5 Gene regulation

The central dogma of molecular biology concerns the information flow within cells. It states that the information is translated between different molecular types as follows:



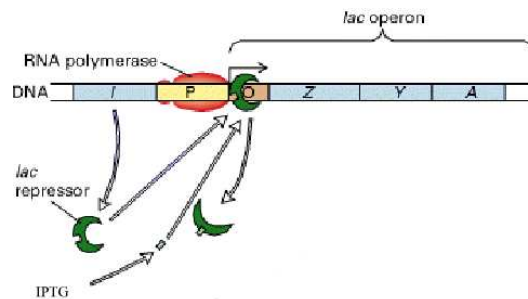
For gene regulation the important steps are the transcription ($\text{DNA} \rightarrow \text{RNA}$) and translation ($\text{RNA} \rightarrow \text{Proteins}$). As have been discussed previously in the course, also the ability of specific proteins (transcription factors) to affect the transcription rate is essential (see figure below). This allows for a network of proteins regulating each others production (or a network of genes regulating each others activity).



The biological processes involved in transcription and translation are complex, and the mathematical descriptions we will discuss here are simplified approximations. This is most often sufficient due to the lack of detailed experimental data, and allows for using them in a large network setting. It is also often convenient to model transcription and translation within a single equation, and due to the complex input-output relations for these processes, nonlinear descriptions are required.

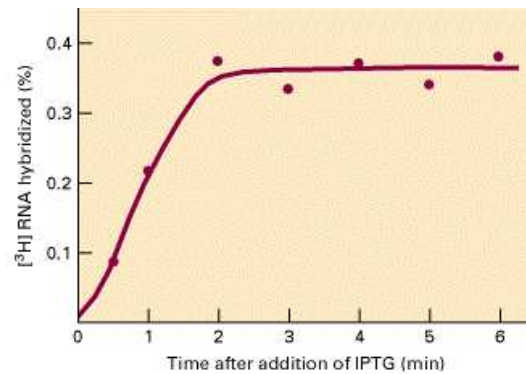
Example: the lac-operon

The idea that transcription factors (proteins) bind to the DNA and regulate the transcription rate of genes was first introduced by Jacob and Monod in 1961. They used the lac operon in *E. coli* and their model is shown in the figure.



In the model a transcription factor, lac-repressor, binds to the DNA and prevents transcription of the lac-operon. The repressor can form a complex with IPTG, which results in that the repressor is released from the DNA and transcription is activated.

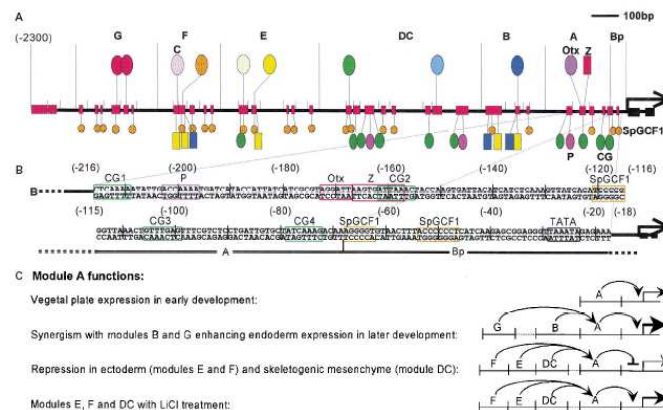
In an experiment where IPTG is introduced to the cells and the lac-operon activity is measured, a quick response can be seen (figure)



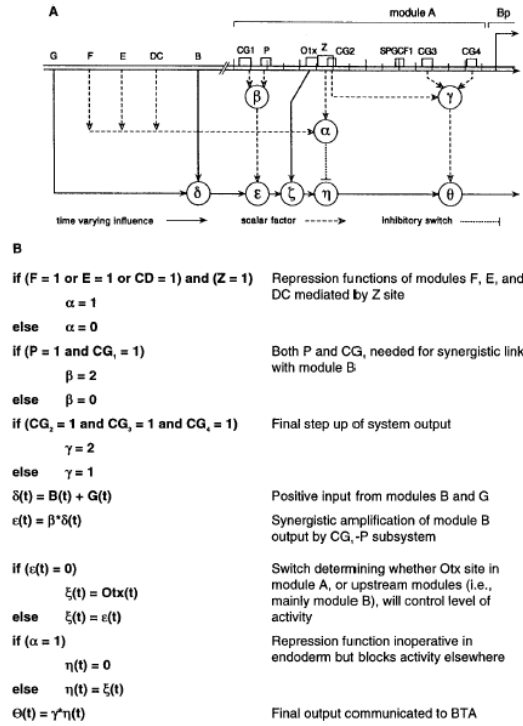
This simple gene regulation system has features which are common for gene expression. It is highly nonlinear, and it has a saturated behavior with a maximal value of the production rate.□

Example: sea urchin gene *Endo16*

When it comes to genetic regulation in multicellular organisms, one of the most studied species is the sea urchin. This example shows the complexity of a single promotor with a manifold of modules which in turn is regulated by a manifold of molecules (figure).



The authors have also created a model of the transcription activity and use a combination of logical rules and continuous equations (figure below). Fortunately(?), this complex regulation is beyond the scope of the course, but one should be aware of that the simple models introduced later in this section have limitations on how accurately they describe the transcription/translation processes.



□

1.5.1 Boolean model with logical rules

The simplest assumption for a gene regulatory network is the boolean approximation, where genes can be either active or inactive (on/off). This can also be interpreted as proteins being present/absent in the cell. Boolean rules (e.g. AND, OR) of the input nodes are defined for determining the state of a node at the next time point. This results in a model with discrete variables and discrete updates in time. The description has the advantage with an enumerable number of possible states for the network, and hence allows for a global exploration of states and dynamics.

Example: boolean description of the lac-operon

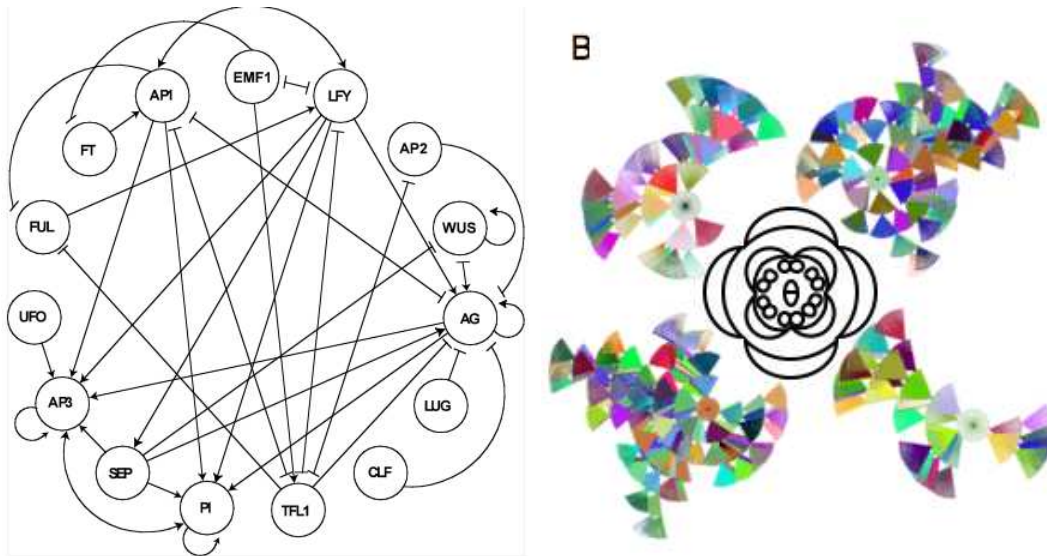
In the simple Jacob-Monod model for the lac-operon from the previous example, activity of the operon was determined by presence/absence of lac-repressor and IPTG. In a boolean description the logic of the lac-operon can be described by the following rule

input		output
lac-repressor	IPTG	lac-operon
0	0	1
0	1	1
1	0	0
1	1	1

The only case when the lac-operon is inactive is when the repressor and not the IPTG is present. The repressor is normally expressed. Adding IPTG then causes the lac-operon to switch from inactive to active (as is seen in this model and in previous experiment).□

Example: boolean description of flower development

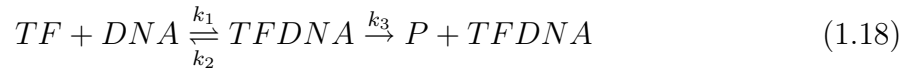
An example of an investigation of the complete state space in a boolean model is the work of Alvarez-Buylla et al. Here the ABC-model for plant flower development is investigated by defining a transcriptional network of genes known to be important along with known and hypothesised interactions. The authors were able to show that the network dynamics resulted in 10 fixed points (out of 139968 states), which then were correlated with known expression profiles for different organs such as petals, stamen, and carpel, as well as for earlier tissues in flower development (Figure).



□

1.5.2 Michaelis-Menten

The transcription/translation process can be modeled as a transcription factor (TF) binding to DNA (creating a complex) which activates or represses the production of a protein P . A model describing an activator is



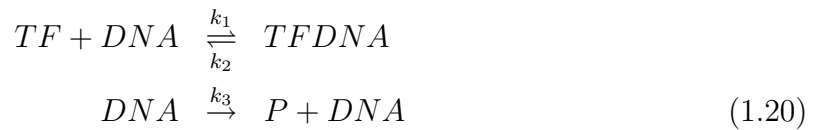
Assuming that the binding/release of the transcription factor is fast compared to the production of the protein allows for a Michaelis-Menten formalism to be used. The 'enzyme' in this case is the DNA, and it can be assumed to exist as a single copy within a cell ($DNA + TFDNA = 1$). Solving for the equilibrium of the left part of the reaction leads to $TFDNA = TF/(K + TF)$ where $K = k_2/k_1$. This can be interpreted as the relative occupation of the binding site or the fraction of time the transcription factor TF is bound. The production of P can then be seen as this fraction times the rate of production when the regulation is active (given by $k_3 = V_{max}$), which results in

$$\frac{d[P]}{dt} = V_{max} \frac{[TF]}{K + [TF]} \quad (1.19)$$

Note that the reactions described in Eq. 1.18 is not exactly the same as in the Michaeli-Menten enzyme reaction Eq.1.9. How are the parameters V_{max} and K_m defined in this transcription version? When is there no difference compared to the enzymatic case?

Example: Michaelis-Menten repressor

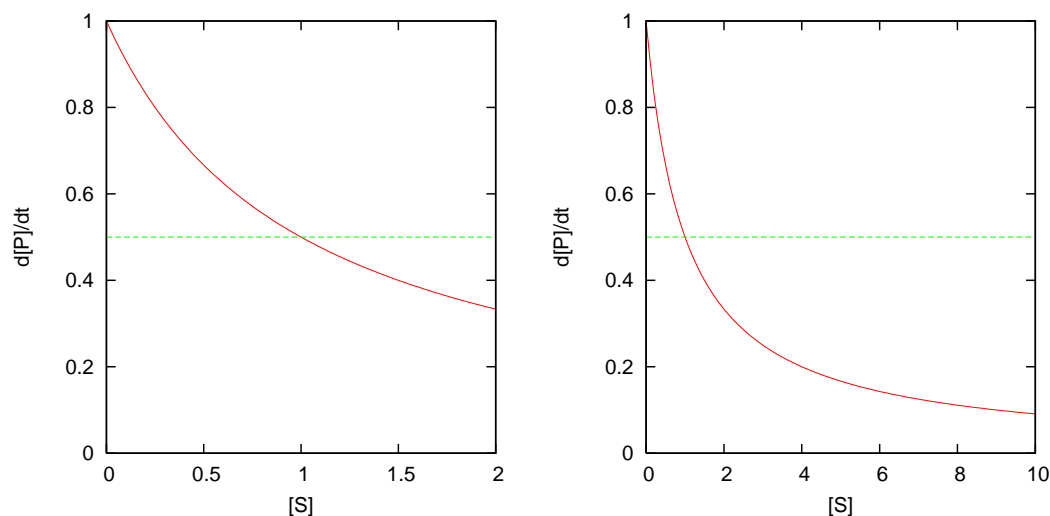
Assume instead that transcription is active if no transcription factor is bound to the DNA , and inactive when the transcription factor (TF) binds



This leads to a repressor model and working out the Michaelis-Menten formalism (try it!) leads to a production of P described by

$$\frac{d[P]}{dt} = \frac{V_{max}K}{K + [TF]} \quad (1.21)$$

which have the behavior shown in the figure below



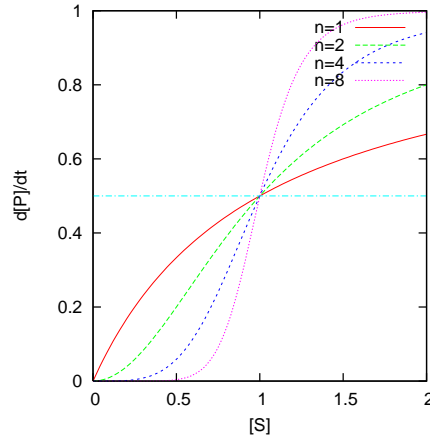
Again this can be seen as the fraction of time the DNA binding site is unoccupied ($K/(K + [TF])$) times the production rate, $k_3 = V_{max}$, when inactive (unoccupied). \square

1.5.3 Hill-equation

As mentioned in the Michaelis-Menten section on enzyme kinetics, a problem with this formalism is the slow response to changes in substrate concentrations (≈ 81 -fold change needed for switching between on/off). For transcription this becomes even more evident, and a common extension of the Michaelis-Menten formalism is the Hill equation. Often it is written in the form

$$\frac{dP}{dt} = V_{max} \frac{S^n}{K^n + S^n} \quad (1.22)$$

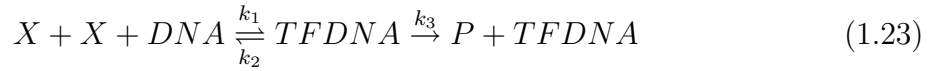
where the parameters n and K are called the Hill coefficient and Hill constant, respectively. The Hill constant corresponds to the substrate concentration that results in 50% response, and the Hill coefficient is determining the steepness of the response. The figure below shows the dependance on n given a fixed K .



The Hill-equation can be deduced from a model where a transcription factor can bind to DNA at multiple sites. Hill himself regarded the equation as a model that better fitted experiments, which is not an uncommon standpoint among modelers (i.e. the parameter values are defined by fitting to experiments, rather than from a transcription factor binding model).

Example, Hill from a complex

Assume that two molecules of a single protein type, X , activates the transcription/translation of another protein, P . The reactions can be formulated as



From the equilibrium of the left reaction (together with the assumption $DNA + TFDNA = 1$), the fractional occupancy of the binding site is given by $TFDNA = X^2/(K + X^2)$, where $K = k_2/k_1$ (show this!). The production rate is then determined by ($k_3 = V_{max}$)

$$\frac{dP}{dt} = V_{max} \frac{X^2}{K + X^2}. \quad (1.24)$$

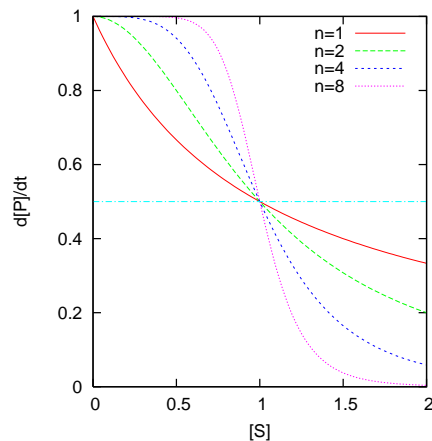
□

Example, Hill repressor

In the case of a repressor S deactivating the transcription of P , the Hill-equation looks like

$$\frac{dP}{dt} = V_{max} \frac{K}{K + S^n} \quad (1.25)$$

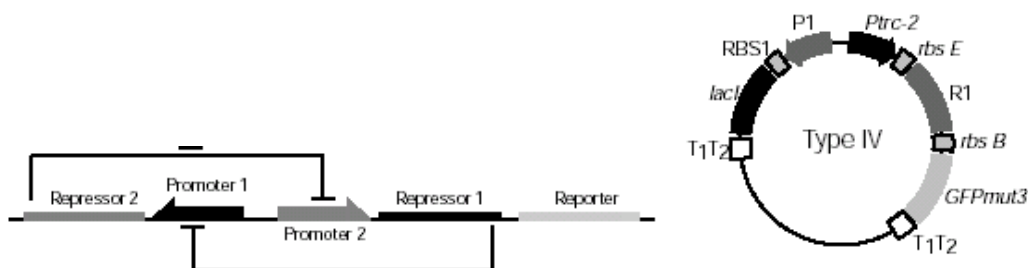
which shows a n dependance as in the figure below.



□

Example, bistable switch

In a beautiful work by Gardner et.al. a genetic switch is created by direct manipulation of the DNA in *E. coli* (figure below). A network of two genes repressing each other is constructed, and this novel technique allows for creating simple systems where direct comparisons between models and experiments are more tractable.

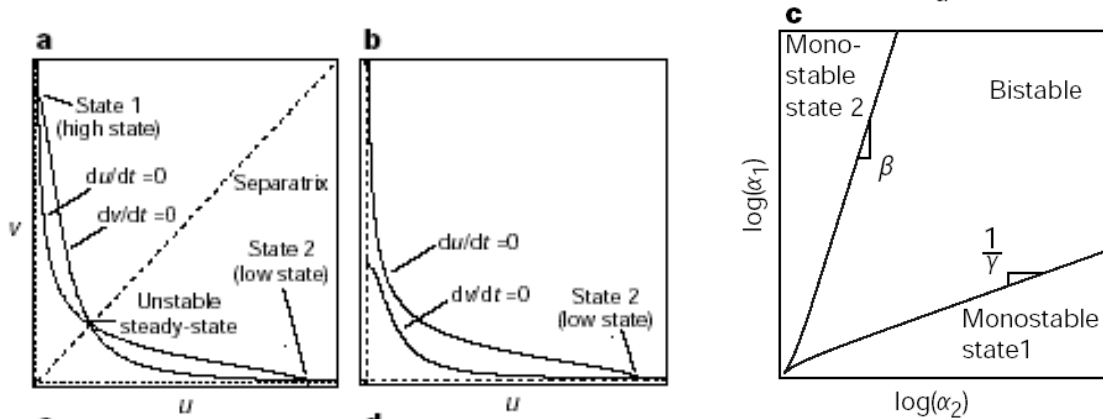


The equations used in this model are of Hill-type plus addition of a constant degradation term.

$$\begin{aligned} \frac{du}{dt} &= \frac{\alpha_1}{1 + v^\beta} - u \\ \frac{dv}{dt} &= \frac{\alpha_2}{1 + u^\gamma} - v \end{aligned} \quad (1.26)$$

The model can behave as a bistable switch where two stable fixed points are defined by $(u, v) = (\text{high}, \text{low})$ and $(\text{low}, \text{high})$ respectively. A phase plane plot with the nullclines

(calculate them!) are shown in the figure below, and quite interestingly, either β or γ needs to be larger than one to get the bistable behavior. Otherwise the system has a single stable fixed point. This model will be examined during the computer exercise.



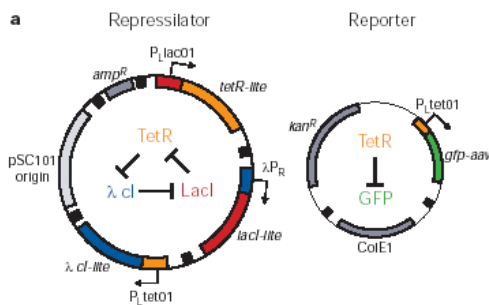
□

1.5.4 Models accounting for both transcription and translation

Sofar, we have only looked at models describing the transcription and translation in a single equation. It is of course also possible to divide these into two different processes, and also treat the mRNA as a dynamical variable.

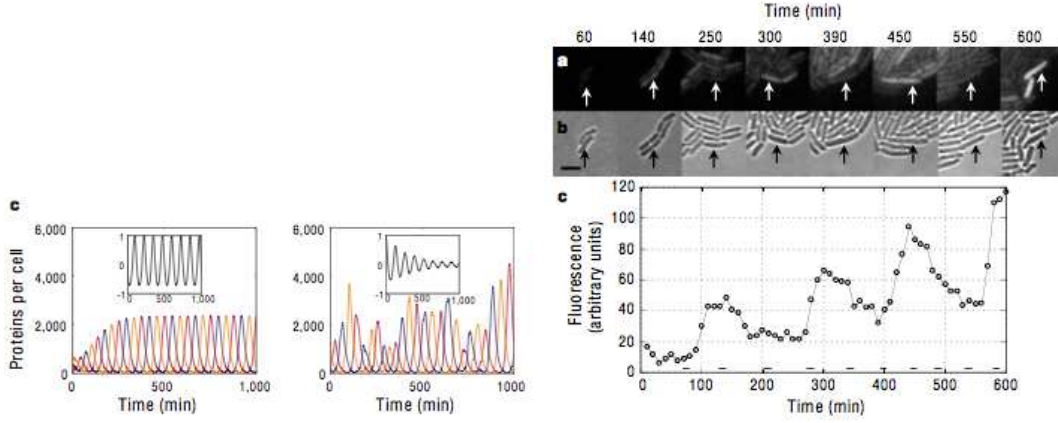
Example, the repressilator

In a similar effort as described in the bistable switch example, Elowitz et.al. constructed a network of three repressing genes (figure). A computer exercise is devoted to modeling of this system, and details are left for then, but the equations used are presented below as an example of a transcription/translation model.



$$\begin{aligned} \frac{dm_i}{dt} &= -m_i + \frac{\alpha}{(1 + \rho_i^n)} + \alpha_0 \\ \frac{dp_i}{dt} &= -\beta(p_i - m_i) \end{aligned} \quad \begin{aligned} (i = lacI, tetR, cl) \\ (j = cl, lacI, tetR) \end{aligned}$$

The m variables represent mRNA and the p variables represent proteins. The transcription is modeled by a Hill-type equation, and translation is modeled by a linear equation. In addition to this, constant degradation of all molecules are modeled. The figure below show the oscillating behavior achieved both in the simulations, and in the experiments. The left simulation plot shows the deterministic model described above, and the right plot shows a stochastic version.



□

1.5.5 Combining contribution from several transcription factors

As has been seen in the single transcription factor examples the rate limiting part of gene expression is typically the initiation of transcription. The models were based on the assumption that the binding and unbinding of transcription factors were fast and could be assumed to be in equilibrium, which resulted in a probability for a bound and unbound state respectively. Then each of these states were connected to a rate for transcription. This idea can easily be extended to multiple transcription factors where the combined probabilities are used.

Example: A combined activator/repressor rule

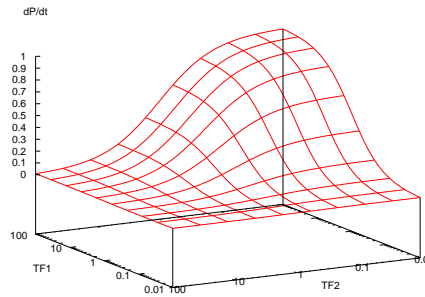
A combined activator and repressor in a Michaelis-Menten formalism results in individual probabilities

$$\begin{aligned}
 P_{TF1bound} &= \frac{[TF1]}{K_1 + [TF1]} = \frac{[TF1]/K_1}{1 + [TF1]/K_1} \\
 P_{TF2notbound} &= \frac{K_2}{K_2 + [TF2]} = \frac{1}{1 + [TF2]/K_2}
 \end{aligned} \tag{1.27}$$

If these probabilities are assumed to be independent, the probability that TF1 is bound and TF2 is not is given by

$$\begin{aligned} P_{TF1boundANDTF2notbound} &= P_{TF1bound}P_{TF2notbound} = \\ &= \frac{[TF1]/K_1}{1 + [TF1]/K_1 + [TF2]/K_2 + [TF1][TF2]/K_1K_2} \quad (1.28) \end{aligned}$$

This probability can then be multiplied with a maximal rate for transcription resulting in a function as shown in the figure below



□

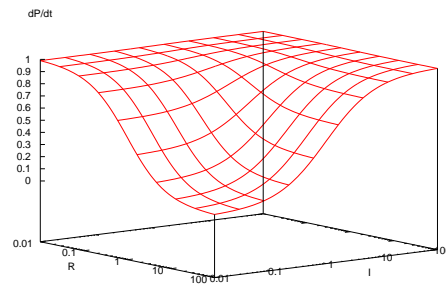
In the previous example only one specific bounding pattern resulted in transcription, but this can be generalized to transcription for more than one combination, as e.g. for the lac-operon as discussed previously.

Example: Michaelis-Menten version of the lac-operon

A simplified model for lac-operon regulation using a Michaelis-Menten formalism for a lac-repressor (R) and IPTG (I) could be assumed by letting transcription occur as soon as the repressor is not the only molecule present (compare with the boolean rule in the earlier example). Show that this leads to

$$\frac{dP}{dt} = \frac{V_{max}(1 + k_2[I] + k_3[R][I])}{1 + k_1[R] + k_2[I] + k_3[R][I]}. \quad (1.29)$$

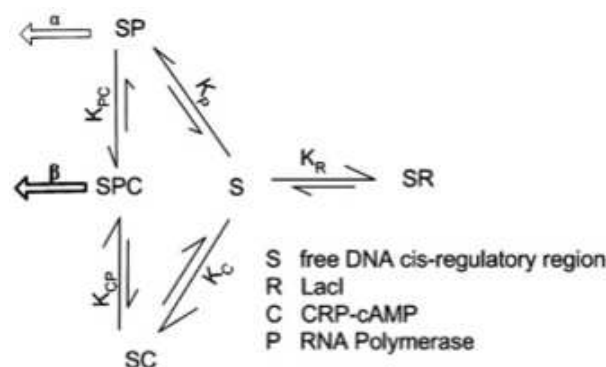
This function is shown in the figure below, and it can be seen that when I is not present R represses the activity, and that the activity increases with increasing concentration of I . Note that all active states leads to the same maximal production (V_{max}) in this example.



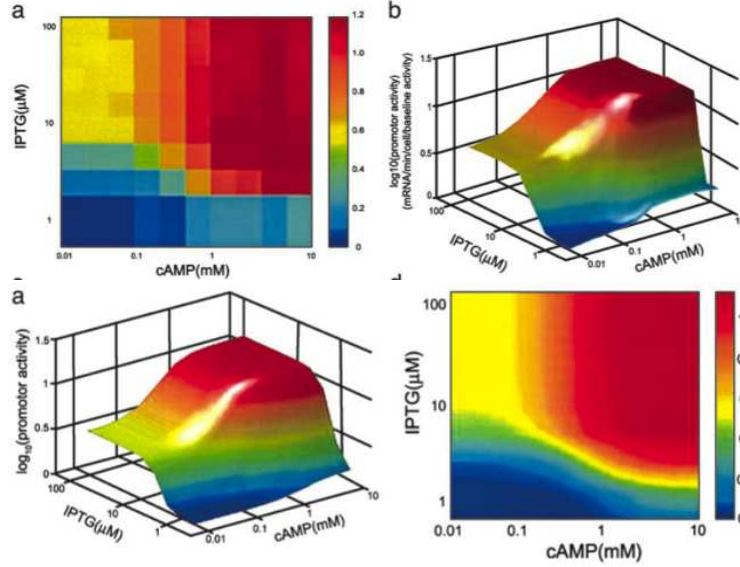
Note that a similar function is the result of a model where complex formation of R and I is assumed together with the single R -repression when R binding to DNA and complex formation is assumed to be fast. \square

Example: experimental comparison for the lac-operon

A more detailed model of the lac-operon has been presented by Setty et. al. It includes the lac repressor and IPTG as well as a second inducer (CRP-cAMP) and the RNA Polymerase. The model assumes different rates of production (α, β) for different states of the promoter and also some leakiness. Finally it uses Hill-formalism for the IPGT and cAMP binding. An illustration of the model interactions is shown in the figure below.



Transcription was measured at a number of concentration combinations of IPTG and cAMP concentrations. Interestingly the transcription rates were given by different plateaus and was more elaborate than a simple AND function (figure below, top), something that was also correctly described by the model (figure below bottom).



□

An alternative view of the transcription rates for different transcription factor binding states is given by the approach of Shea and Ackers (1985). In this statistical physics view, the combination of all possible states are defining a partition function (which is given by the denominator in the expressions). Transcribing states are then given in the nominator, which can be interpreted as the cases where the RNA Polymerase is bound to the DNA. By relating each combination of transcription factor states with a free energy dependancies of binding can be accounted for (e.g. recruitment and overlapping binding sites). The partition function can be written as

$$Z = \sum_{\sigma_1 \dots \sigma_n} \prod_i^n [TF_i]^{\sigma_i} e^{-\Delta G_{\sigma}/RT} \quad (1.30)$$

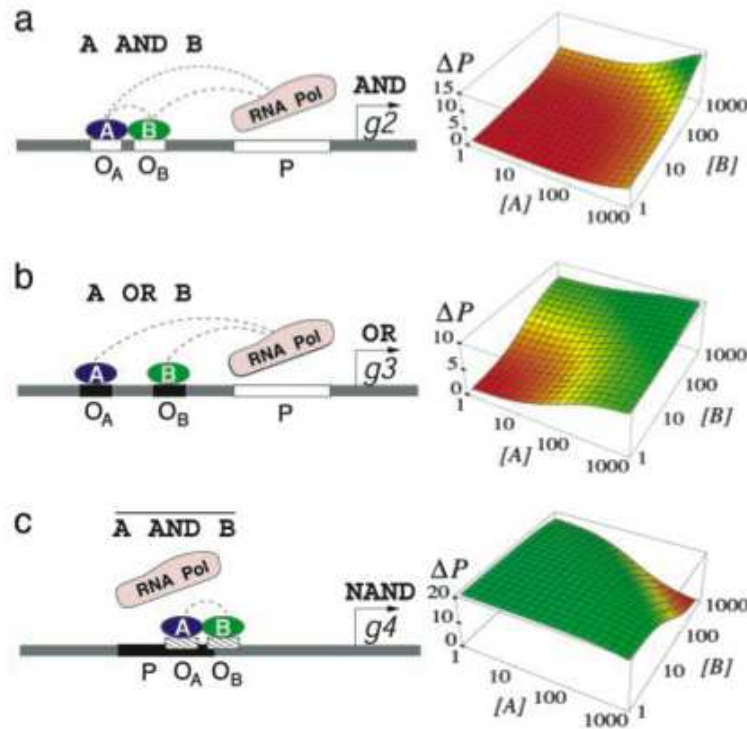
where each transcription factor TF_i can be either bound $\sigma_i = 1$ or not bound $\sigma_i = 0$, and all possible states are accounted for. The transcription rate is proportional to the probabilities of the transcriptionally active states

$$P = \frac{Z_{active}}{Z_{inactive} + Z_{active}} \quad (1.31)$$

Example, transcription logic

Buchler et. al. (2003) used the Shea-Ackers methodology to investigate how different logical rules could be implemented for regulating transcription, and its relation to tran-

scription factor binding mechanisms. The figure below shows example of some of the rules for two transcription factors.



□

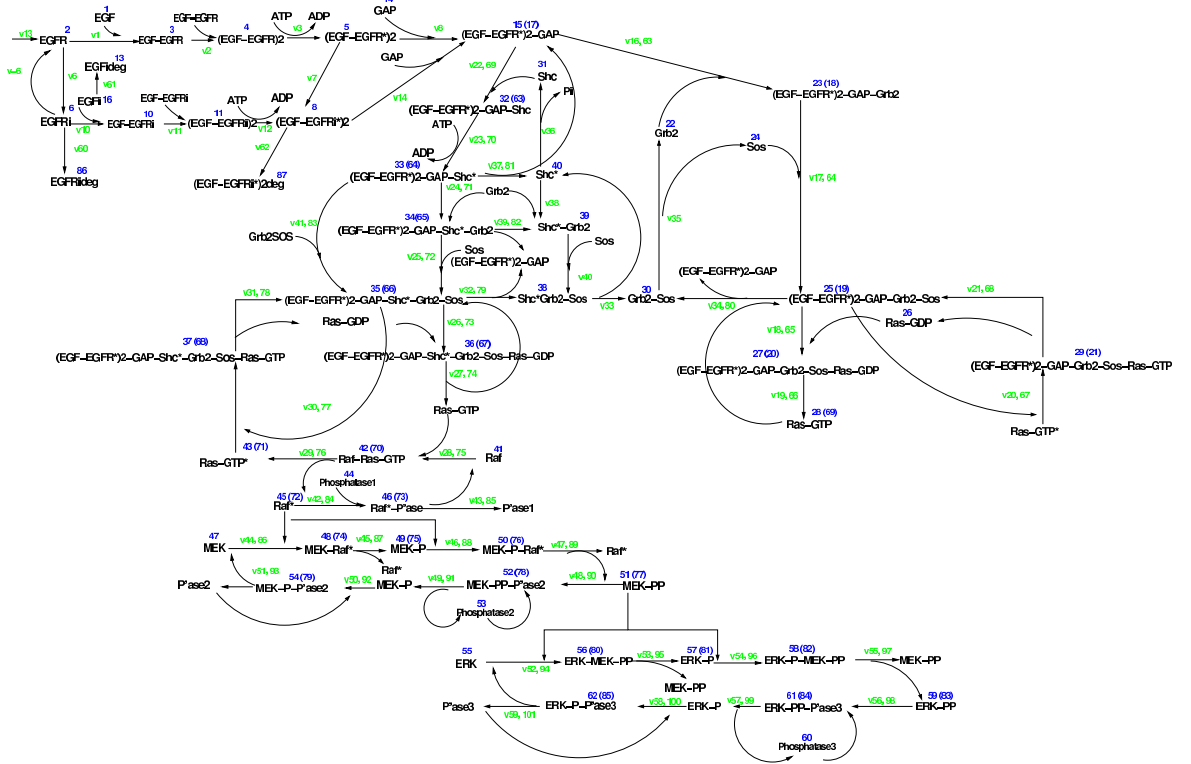
1.6 Large molecular networks; systems biology in a nutshell

Using the building blocks of mass action and enzymatic reactions, and transcription/translation descriptions, models of large biochemical networks can be developed. In these cases analytical solutions are unreachable, and computer simulations of the systems are necessary.

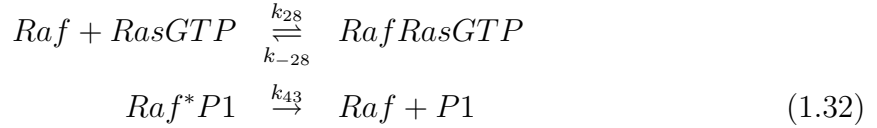
Example: EGF-pathway simulation

The receptor to the epidermal growth factor (EGF) ligand belongs to the tyrosine kinase family of receptors and is expressed in virtually all organs of mammals. EGF receptors play a complex role during development and in the progression of tumors. Schoeberl *et.al.* have created a model of the pathway as shown in the figure below.

1.6. LARGE MOLECULAR NETWORKS; SYSTEMS BIOLOGY IN A NUTSHELL³³



This might look like a far too advanced example for our purposes, but let's look at the reaction for a single molecule, e.g. the *Raf*. It is directly involved in two reactions



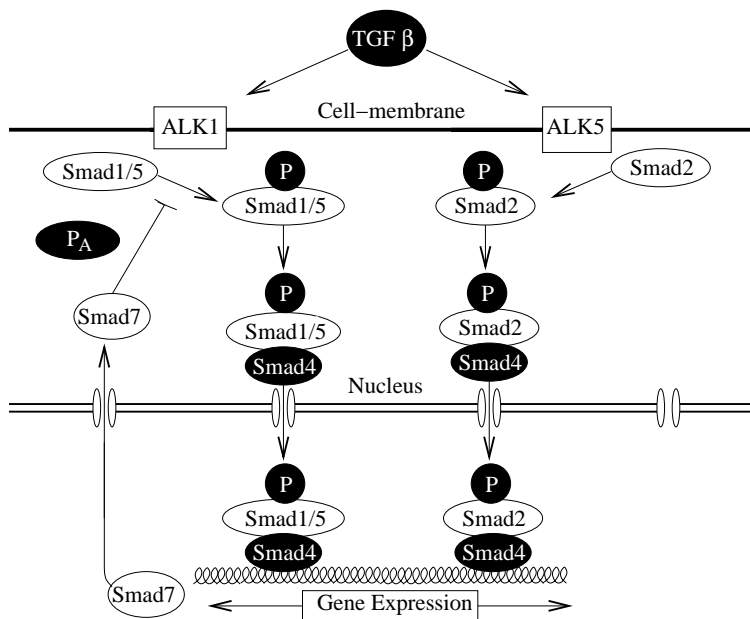
and the formulation of the differential equation for *Raf* is straightforward using the mass action formalism

$$\frac{d[Raf]}{dt} = -k_{28}[Raf][RasGTP] + k_{-28}[RafRasGTP] + k_{43}[Raf^*P1] \quad (1.33)$$

□

Example: TGF- β pathway

The TGF- β pathway plays a prominent role in inter- and intracellular communication and subversion can lead to cancer, fibrosis vascular disorders and immune diseases.



This network includes both molecular reactions and transcriptional regulation. A model for the pathway can be defined by the reactions in Table 1.1.

\emptyset	$\xrightleftharpoons[p_0 p_1]{p_0}$	ALK1	(1)	\emptyset	$\xrightleftharpoons[p_2 p_3]{p_2}$	Smad1	(9)
\emptyset	$\xrightleftharpoons[p_4 p_5]{p_4}$	Smad4	(2)	\emptyset	$\xrightleftharpoons[p_6 p_7]{p_6}$	Smad2	(10)
\emptyset	$\xrightleftharpoons[p_8 p_9]{p_8}$	ALK5	(3)	\emptyset	$\xrightleftharpoons[p_{10}]{PS14N (p_{11}, p_{12})}$	Smad7	(11)
TGF β + ALK1	$\xrightleftharpoons[p_{14}]{p_{13}}$	TA1	(4)	Smad1	$\xrightleftharpoons[p_{17}]{(p_{15}, p_{16}) TA1}$	PSmad1	(12)
PSmad1 + Smad4	$\xrightleftharpoons[p_{19}]{p_{18}}$	PS14	(5)	Smad2	$\xrightleftharpoons[p_{24}]{(p_{22}, p_{23}) TA5}$	PSmad2	(13)
TGF β + ALK5	$\xrightleftharpoons[p_{21}]{p_{20}}$	TA5	(6)	PSmad2 + Smad4	$\xrightleftharpoons[p_{26}]{p_{25}}$	PS24	(14)
P_A + TA1	$\xrightleftharpoons[p_{28}]{Smad7 p_{27}}$	TA1P	(7)	PS14	$\xrightleftharpoons[k_{30}]{p_{29}}$	PS14N	(15)
P_B + TA5	$\xrightleftharpoons[p_{32}]{Smad7 p_{31}}$	TA2P	(8)				

Table 1.1: The different reactions in the TGF- β pathway model, where p_i ($i = 0, 1, \dots, 32$) are the rate constants. Reactions with the symbol \emptyset model production and degradation. In reactions (11), (12) and (13) Michaelis-Menten dynamics is used.

1.6. LARGE MOLECULAR NETWORKS; SYSTEMS BIOLOGY IN A NUTSHELL35

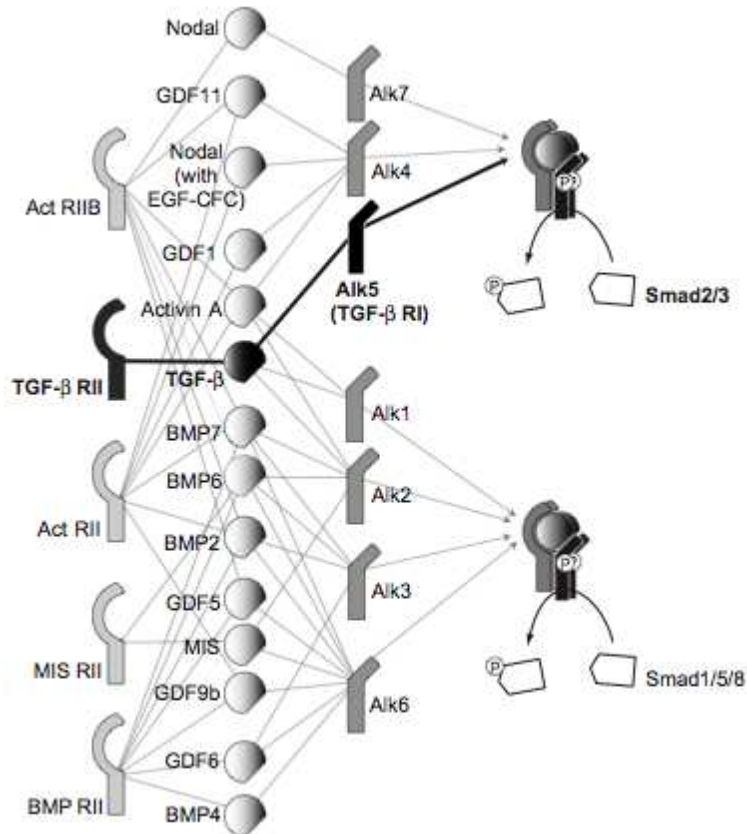
As an example, the model equation for the Smad1 concentration is given by

$$\frac{d[\text{Smad1}]}{dt} = p_2 - p_2 p_3 [\text{Smad1}] + p_{17} [\text{PSmad1}] - \frac{p_{15} [\text{Smad1}] [\text{TA1}]}{p_{16} + [\text{Smad1}]}, \quad (1.35)$$

which is extracted from reactions 9 and 12 above. Try to extract the model equation for another molecule! \square

Example: The TGF- β family of ligands and their receptors

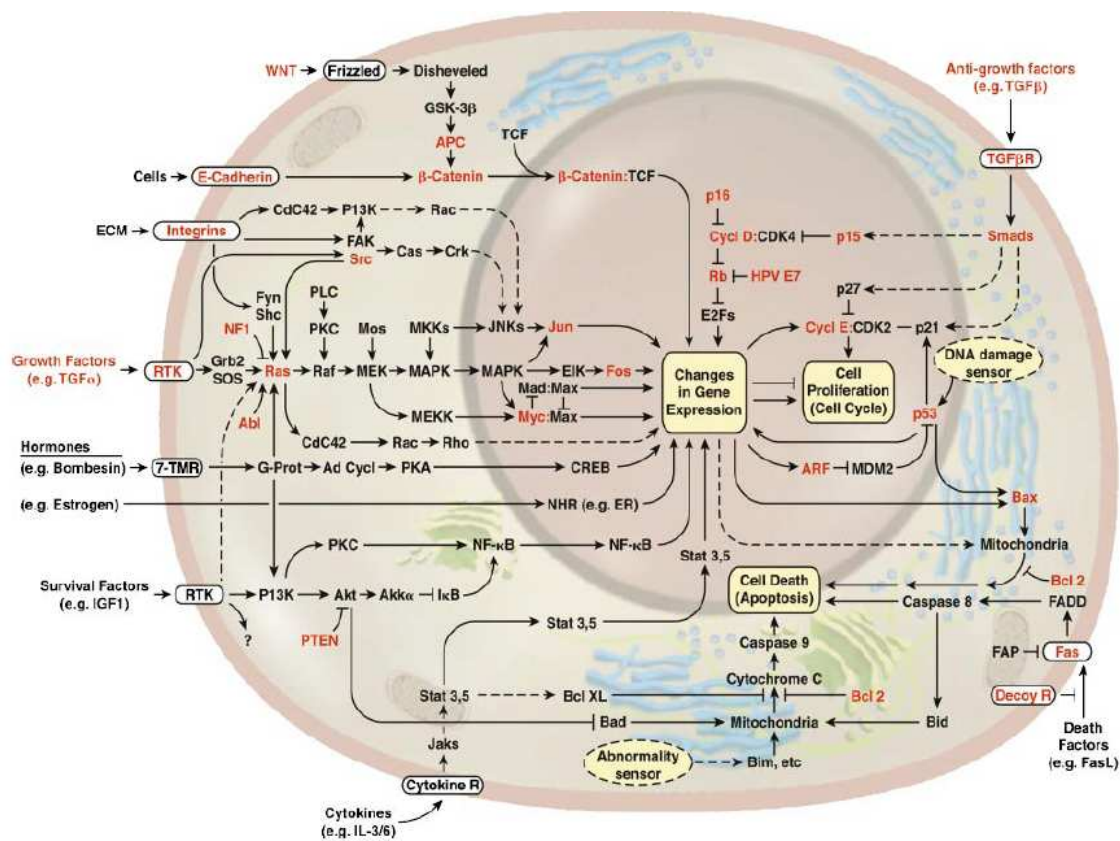
In the previous example, a module of the TGF- β signalling pathway was presented. In idealized experiments, this module can be investigated. A problem that might have to be accounted for in a modeling approach is crosstalk between a model and its surrounding (all molecules left out of the model). Hence the presented model might not correctly describe the behavior within a living organism. For example, TGF- β is only one member of a whole family of ligands, that binds to a number of different receptors and each ligand-receptor combination can activate/deactivate the same pathway (see figure).



\square

Example: Pathways of relevance for cancer

The complexity within living cells are even larger than shown in the previous examples. Both the EGF and the TGF- β pathways are important in cancer progression. As shown in the figure below (from Carstens introduction), these pathways are only two of multiple pathways that are important in this case.



This is an example of a number of modules (the specific pathways with robust behavior and 'output') that interact with each other. \square

1.7 Estimation of parameter values

Even when the mathematical description of a model is defined (as in the previous section) the dynamical behavior can change due to different values of the parameters. A main task within a modeling approach is to find or estimate parameter values that are relevant for the biological system at hand. Here we will discuss two different approaches for estimating parameter values; experimental measurements, and reverse engineering.

1.7.1 Experimentally measuring parameter values

If it is possible, a good way to find parameter values is to measure the dynamics of a single reaction. From this it is then possible to estimate the rate parameters.

Example: ALK1 internalization rate

In an experiment, the ALK1 receptor at the cell membrane is labeled with an antibody, and after 15 minutes the amount of labeled ALK1 receptor is measured. At this time only 5% of the labeled ALK1 molecules are still present.

Assume a reaction $X \xrightarrow{k} \emptyset$ as the receptor disappears from the membrane, which leads to an equation

$$\frac{dX}{dt} = -kX. \quad (1.36)$$

The solution to this equation is $X(t) = X_0 e^{-kt}$ where X_0 is the initial concentration. (This is easily checked by taking the time derivative of $X(t)$.) The kinetic parameter can be estimated by

$$\begin{aligned} e^{-kt} &= \frac{X(t)}{X_0} = 0.05 \\ k &= -\frac{1}{t} \ln \frac{X(t)}{X_0} = -\frac{1}{15} \ln 0.05 = 0.2 \text{ min}^{-1} \end{aligned} \quad (1.37)$$

This estimate could be improved further by fitting a curve $X = X_0 e^{-kt}$ to a dynamical measurement of the labeled ALK1 receptors. \square

1.7.2 Reverse engineering

Even if parameter values are not known from experiment it can be possible to do a reverse engineering to find parameters for the model that result in an agreement of model and some biological features of the system. The first thing needed is an objective function (error measure) that is a quantitative measure of how well the model behavior (for a given parameter set) corresponds to the biological feature at hand. Then an

optimization method is needed to find parameters that result in an optimal value of the objective function. Typically this is a hard optimization problem in a high dimensional parameter space, and one has to rely on iterative heuristic algorithms to find 'good' solutions.

Objective function

The objective function, $R(p)$, is a function of the model parameters p . If the system of differential equations for the model is not analytically solvable, a simulation of the model for specific parameters is needed for evaluating the objective function value. The most common type of objective function assumes that there are some quantitative experimental data available for molecular concentrations allowing for a direct comparison with the model variables. If for example the concentration of protein X has been measured at N time points t_1, t_2, \dots, t_N , a mean square error can be defined as

$$R(p) = \frac{1}{N} \sum_t^N (X_t^{\text{exp}} - X(p)_t^{\text{model}})^2 \quad (1.38)$$

where X_t^{exp} are the measured concentrations and $X(p)_t^{\text{model}}$ are the model variable values at different time points.

Optimization algorithms

Iterative algorithms are often used when optimizing an objective function. When the functional form of the model output is known, function fitting can be used, and when the model is linear and the parameters are confined by linear constraints, linear programming can be used.

Example: experimental lac-operon revisited

In a previous example where a model for transcription was compared to experimental data, the transcription rate, f , was described by the function

$$f = V_1 \frac{1 + V_2 A + V_3 R}{1 + V_4 A + V_5 R} \quad (1.39)$$

where

$$\begin{aligned} A &= \frac{X^n}{1 + X^n}, (X = [cAMP]/K_{cAMP}) \\ R &= \frac{1}{1 + Y^m}, (Y = [IPTG]/K_{IPTG}) \end{aligned}$$

The parameters were optimized using a nonlinear root mean square fit, and the resulting parameters (same for multiple runs with different initial conditions) are showed below.

Table 1. *lac* model parameters that best fit the measurement using the GFP reporter plasmid (wild type) and putative mutants that have purer AND-like and OR-like gates

Parameter	Wild type	AND	OR
m	4 ± 0.6	4	4
n	2 ± 0.4	2	2
$K_{IPTG}, \mu\text{M}$	1.2 ± 0.2	1.2	1.2
K_{cAMP}, mM	1.8 ± 0.5	1.8	1.8
V_1	3.5 ± 0.7	1	10
V_2	70 ± 10	70	1,700
V_3	170 ± 30	2,000	15
V_4	17 ± 3	17	400
V_5	540 ± 100	7,000	50

Errors are parameter variations that give 15% deviation from the best-fit results. The values of the parameters that were changed to produce AND- and OR-like gates are in boldface. V_1, V_2, V_3, V_4 , and V_5 represent combination of the biochemical parameters $a, b, c, d, \alpha, \beta$, and γ (see *Methods*).

In the table, also parameter values leading to different other logical rules are provided. \square

Often iterative heuristic algorithms are needed. These iterative procedure consists of three steps: 1) Solve the differential equation and calculate the objective function value, 2) Adjust model parameters and resimulate, 3) Accept or reject the new parameters (or construct a new set of parameter values) depending on the difference in objective function value. Three examples of iterative optimization algorithms that could be used for parameter estimations are

- **Local search.** This is the naive way of trying to find a good value for the objective function. Here you start with a parameter set for which the model is simulated and the objective function is evaluated. After adjusting parameters a new objective function value is evaluated and it is accepted if this value is lower than the previous one. This means that we will only go downhill in the objective function 'landscape' and we will end up in the closest local minimum.
- **Simulated annealing.** Again, you start with a parameter set for which the objective function is evaluated, then do a parameter adjustment and reevaluate the objective function. Now the new parameter set is accepted with a probability one if $\Delta R = R_{new} - R_{old}$ is negative, and with probability $e^{-\Delta R/T}$ if ΔR is positive. T is a parameter (fictitious temperature) which tunes the probability. The first thing to note is that the algorithm can allow for accepting new parameter sets with a higher objective function value, which means that it can escape from local minima. The

second thing to note is that at high values of T , almost all parameter adjustments are accepted and we get something like a random walk in the parameter space (searching large regions). At low values of T almost only decreased objective function values are accepted. The algorithm starts at high values of T and then slowly decreases T until no more updates are accepted.

- **Genetic algorithms.** This type of algorithm is developed from an evolutionary fitness principle. It starts with an ensemble of parameter values for which the objective function is evaluated. Then 'good' parameter sets are kept, 'bad' ones are removed. The bad solutions are replaced by forming new parameter sets from two principles; mutation, where the parameters of a good solution are slightly adjusted, and mating, where the new parameter set is some kind of combination of two good solutions.

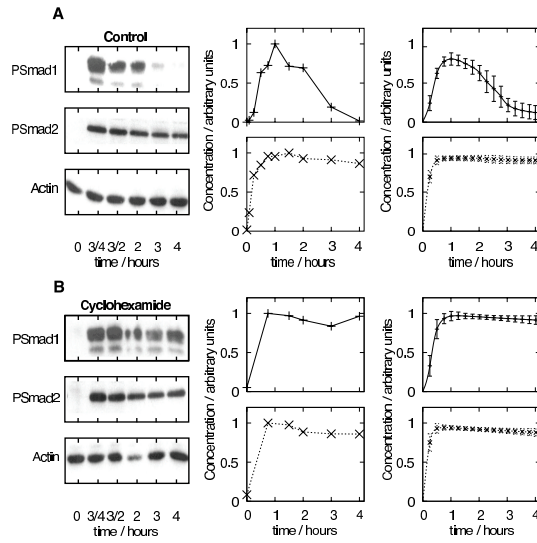
Example: TGF- β model

For the TGF- β pathway, PSmad1 and PSmad2 concentrations are measured at different times after TGF- β stimulation. The concentrations are measured at N discrete time points t_1, t_2, \dots, t_N for two experiments. The model is optimized using simulated annealing type of algorithm and the mean square error is used as an objective function:

$$R(\mathbf{p}) = \frac{1}{N} \frac{1}{M} \sum_{t=t_1}^{t_N} \sum_{i=1}^M (x_i(t) - \tilde{x}_i(t))^2, \quad (1.40)$$

where $x_i(t, \mathbf{p})$ and $\tilde{x}_i(t)$ denote model points and experimental points respectively and the index i denotes the different molecules ($M = 2$ in total). (The sum of the R values from the two experiments is used as objective function.)

The figure below shows experimental data, and the model output for optimized parameters. In this case multiple good solutions were found (the average model behavior is plotted with errorbars).



□

1.8 Model analysis in systems biology

1.8.1 Robustness

Biological systems have evolved and survived for millions of years. They typically inherit a stability towards fluctuations in parameters, and the same modules (e.g. pathways) exist in many different species with varying environment. A good model should also reflect this and hence a test for robustness can be an important test of the model. Robustness analysis can also pinpoint which reactions/parameters that are important for obtaining a specific biological behavior.

A simple measure for sensitivity is to measure the relative change of a system feature due to a change in a parameter. For example the feature can be the equilibrium concentration of a compound, C for which the sensitivity (S) to a parameter p is

$$S_p = \frac{\frac{dC}{C}}{\frac{dp}{p}} = \frac{dC}{dp} \frac{p}{C} \approx \frac{\Delta C}{\Delta p} \frac{p}{C} \quad (1.41)$$

It should be noted that this sensitivity measure is local and depends on the current system “topology” and most often on parameter values. When applying a sensitivity measure, there are often summation laws appearing, as for example in the case of measuring sensitivity on equilibrium values $\sum_i S_{p_i} = 0$. Features often used in robustness analysis are e.g. the time integral of a variable, the duration or amplitude of a peak, etc.

Example: creation and degradation revisited

Let's go back to our first example where a molecule A is produced and degraded at constant rates.



where k is the production rate and d is the degradation rate. We calculated that this system had a fixed point for $A^* = k/d$. This system is so simple that it is possible to calculate the sensitivity of the fixed point with regard to the two parameters. The derivative form leads to

$$\begin{aligned} \frac{dA^*}{dk} \frac{k}{A^*} &= \frac{1}{d} \frac{k}{k} = 1 \\ \frac{dA^*}{dd} \frac{d}{A^*} &= -\frac{k}{d^2} \frac{dd}{k} = -1 \end{aligned} \quad (1.43)$$

The difference version relies on that a parameter value is changed with a fraction f ($p \rightarrow p + fp$), and that the fixed point is calculated (or measured in a simulation) for the new parameter value. Changing the parameters a fraction f leads to new fixed points

$$\begin{aligned} A^*(k + fk, d) &= \frac{k + fk}{d} = (1 + f) \frac{k}{d} \\ A^*(k, d + fd) &= \frac{k}{d + fd} = \frac{1}{(1 + f)} \frac{k}{d} \end{aligned} \quad (1.44)$$

and the sensitivity measures are given by

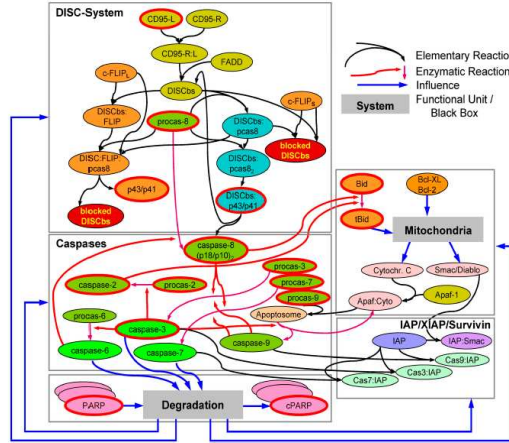
$$\begin{aligned} \frac{A^*(k + fk, d) - A^*(k, d)}{fk} \frac{k}{A^*(k, d)} &= \frac{fA^*(k, d)}{fk} \frac{k}{A^*(k, d)} = 1 \\ \frac{A^*(k, d + fd) - A^*(k, d)}{fd} \frac{d}{A^*(k, d)} &= \frac{-\frac{f}{1+f}A^*(k, d)}{fd} \frac{d}{A^*(k, d)} = -\frac{1}{1 + f} \approx -1 \end{aligned} \quad (1.45)$$

where in the last equation f is assumed to be small.

We can see that if the two parameter parts are summed we get zero (summation law), and that when using the difference version f needs to be small not to introduce errors. The conclusion is that the fixed point is directly increased with the same fraction as k is changed. For the d parameter there is an decrease of the same fraction as d is varied. The system is sensitive to changes in the parameters which is obvious since the parameters are determining the dynamics (and the fixed point) directly. \square

Example: CD95-induced apoptosis

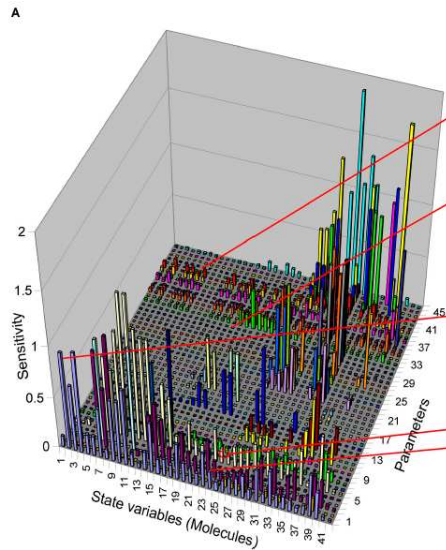
This model developed by Bentele et.al. describes a pathway that regulates apoptosis (programmed cell death). Defects in the regulation of apoptosis result in serious diseases such as cancer, autoimmunity and neurodegeneration. The model components are shown in the figure below.



A local sensitivity analysis is applied to a single solution (parameter set). The measure used is the integral of the protein concentration $c_i = \int_t x_i dt$ where x_i is a concentration. In the figure below the absolute value of the sensitivities,

$$s_{ij} = \frac{dc_i/c_i}{dp_j/p_j}, \quad (1.46)$$

are shown for all molecules i and parameters j .



□

A problem with the local sensitivity measure is that it can be very dependent on the parameter values. One way to improve the sensitivity measure is to measure the local sensitivity in multiple points spanning a region in the parameter space.

Example: Circadian clock

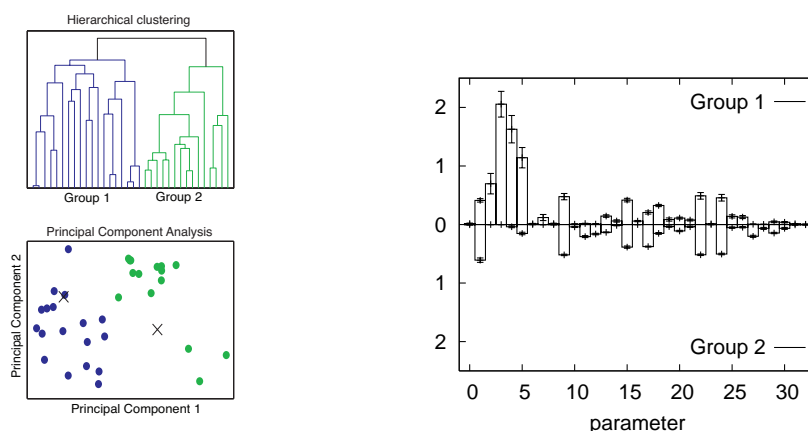
Doyle et. al. (2004) compared robustness in different models for the circadian clock in *Drosophila*. To avoid only using a local sensitivity, they calculated the sensitivity in the parameter space surrounding the optimal values by for each parameter pair scan the parameter space by measuring in points where the parameters where varied 10-fold up and down in 21 steps (resulting in 21*21 measurments for each pair and 703 parameter pairs combined from the 38 parameters). Since it is a oscillating system they measured sensitivity of amplitude and period of the oscillations.

□

Example: TGF- β model

For the TGF- β model the optimization provided multiple solutions that could explain the experimental data (as shown in a previous example). These solutions can be grouped into those that utilizes the Smad7 feedback and those that do not (left figure below shows a clustering of the solutions).

The figure below shows average sensitivity measures calculated from multiple solutions for each group. The sensitivity is measured on the integral of PSmad1 and PSmad2 concentrations for each parameter. The solutions in group 2 (those using Smad7 feedback) are more robust.



□

There are other means to measure more global robustness, which will be discussed in the last lecture.

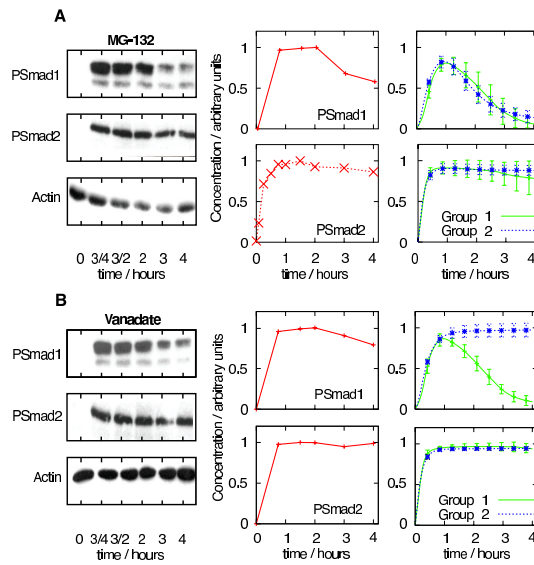
1.8.2 Perturbations

Another way of analysing a model is by applying perturbations. The model behavior could then be compared to the same perturbation in experiments, or predict new biology. The main benefit of having a model in this case is that perturbations are easy to do in the model, while it is often long and hard work to do it experimentally. Multiple perturbations can be tested in a model framework, and those that results in interesting behavior could then be tested in experiments.

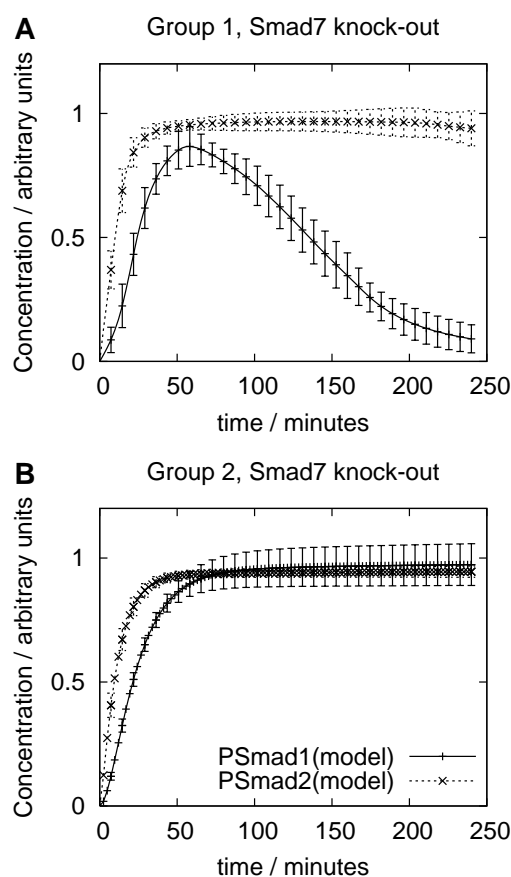
Examples of perturbations are genetic manipulations where genes can be either knocked out or overexpressed, and also silencing techniques such as short interference RNA. Another type of perturbations are environmental changes, where nutrient levels or temperature are examples. Also chemicals can be introduced for perturbing for example protein synthesis or degradation.

Example: Perturbation in the TGF- β model

Chemicals are introduced in the cells before TGF- β stimulation, where either degradation (MG-132) or phosphatases (orthovanadate) are blocked.



The figure below shows the model prediction of a local perturbation where Smad7 is removed from the TGF- β model. Again it is shown for two groups of solutions, where the two groups provide different predictions.



□

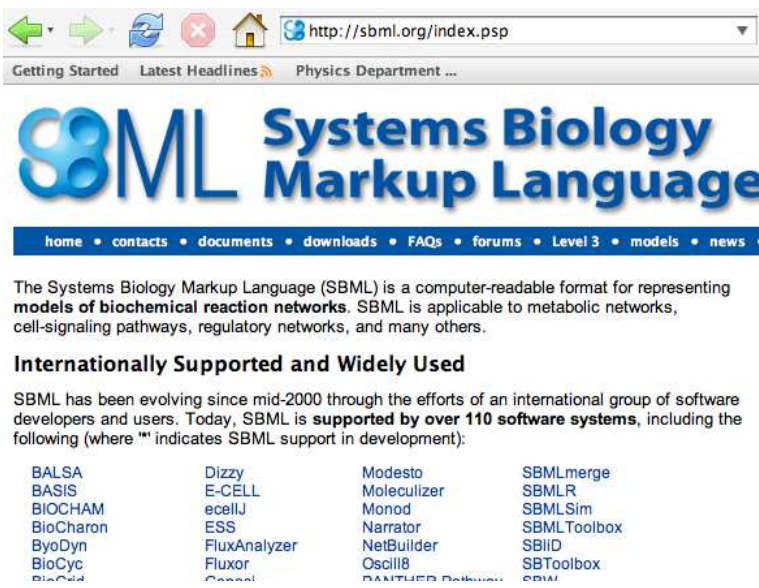
1.9 Systems Biology Tools

A number of computer tools are available for modeling biochemical networks. For simulation, analysis, and optimization of models several packages exist, both utilizing tools such as matlab and mathematica as well as stand-alone applications with graphical user interfaces. These application often have graphical tools for designing models, and there also exist specialized tools for this.

Many of these tools can import and export models in specific format for easier transfer of models, which also simplifies result reproducibility of modelling results. One of these formats is *systems biology markup language (SBML)* (figure).

Finally there are model databases where multiple models are stored, where one example is the *biomodels database* (figure).

Links to many of these tools can be found on the web page <http://sbml.org> (figure), and the biomodels database can be found at <http://www.biomodels.net>.



The Systems Biology Markup Language (SBML) is a computer-readable format for representing **models of biochemical reaction networks**. SBML is applicable to metabolic networks, cell-signaling pathways, regulatory networks, and many others.

Internationally Supported and Widely Used

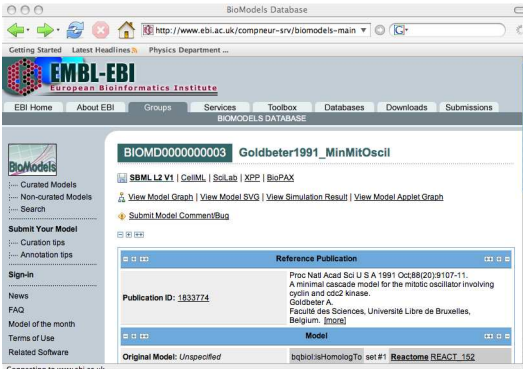
SBML has been evolving since mid-2000 through the efforts of an international group of software developers and users. Today, SBML is **supported by over 110 software systems**, including the following (where "*" indicates SBML support in development):

BALSA	Dizzy	Modesto	SBMLmerge
BASIS	E-CELL	Molecularizer	SBMLR
BIOCHAM	ecellJ	Monod	SBMLSim
BioCharon	ESS	Narrator	SBMLToolbox
ByoDyn	FluxAnalyzer	NetBuilder	SBID
BioCyc	Fluxor	Oscill8	SBToolbox
BioGrid	Genel	SBToolbox	SBML

```

<species metaid="_230515" id="X" name="Cyclin Protease" compar
initialConcentration="0.01" substanceUnits="substance" spatialSizeU
/>
</listOfSpecies>
<listOfParameters>
<parameter id="V1" name="V1" constant="false"/>
<parameter id="V3" name="V3" constant="false"/>
<parameter id="VM1" name="VM1" value="3"/>
<parameter id="VM3" name="VM3" value="1"/>
<parameter id="Kc" name="Kc" value="0.5"/>
</listOfParameters>
<listOfRules>
<assignmentRule metaid="rule1" variable="V1">
<math xmlns="http://www.w3.org/1998/Math/MathML">
<apply>
<times/>
<ci> C </ci>
<ci> VM1 </ci>
<apply>
<power/>
<apply>
<plus/>
<ci> C </ci>
<ci> Kc </ci>
</apply>
</apply>
<cn type="integer"> -1 </cn>
</apply>
</math>
</assignmentRule>

```



1.10 Transport

Reactions within a cell occur at different spatial locations. For example, a signal transduction network usually have reactions at the cell membrane, in the cytoplasm, and in the nucleus. Hence spatial dynamics of molecules might also be important for the behavior of a biochemical network within a cell. Spatial considerations become even more important when modeling multicellular systems, where it is known that signalling molecules (often termed morphogens) can be produced at specific positions, move out in the surrounding tissue, and regulate development.

1.10.1 Diffusion

Molecules are constantly moving and bouncing into each other due to thermal effects. This Brownian motion leads to molecular diffusion. Consider a microscopical model for diffusion that describes number of molecules on a one dimensional lattice discretized in time (x_i, t_k) , where $x_{i+1} - x_i = \Delta x$ and $t_{k+1} - t_k = \Delta t$. The number of molecules in position x_i at time t_k is denoted n_i^k . Assume that each molecule moves Δx either to the right or to the left during a time Δt with probabilities $P_l = P_r = 1/2$. Also assume that consecutive moves are uncorrelated. The average change in molecular number at a spatial point x_i in a time step Δt is given by

$$\begin{aligned}
 n_i^{k+1} - n_i^k &= \Delta n_i^k = P_r n_{i-1}^k - (P_l + P_r) n_i^k + P_l n_{i+1}^k \\
 &= \frac{1}{2} n_{i-1}^k - n_i^k + \frac{1}{2} n_{i+1}^k = \frac{1}{2} (n_{i-1}^k - 2n_i^k + n_{i+1}^k) \\
 &= \frac{\Delta x^2}{2} \frac{n_{i-1}^k - 2n_i^k + n_{i+1}^k}{\Delta x^2}
 \end{aligned} \tag{1.47}$$

This leads to a change per Δt as

$$\frac{\Delta n_i^k}{\Delta t} = \frac{\Delta x^2}{2\Delta t} \frac{n_{i-1}^k - 2n_i^k + n_{i+1}^k}{\Delta x^2} = D \frac{n_{i-1}^k - 2n_i^k + n_{i+1}^k}{\Delta x^2} \tag{1.48}$$

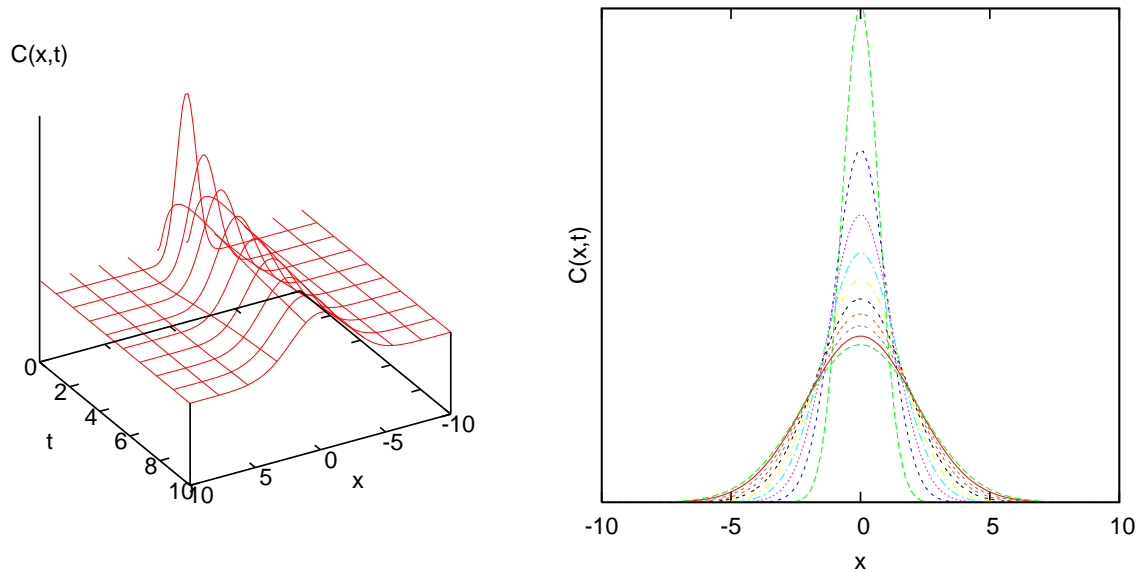
where $D = (\Delta x)^2/(2\Delta t)$ is defined as the diffusion constant. The experienced reader can recognize that the right hand side of the equation corresponds to a discrete version of the second derivative in x ($\approx d^2n/dx^2$). Letting $\Delta x \rightarrow 0$ and $\Delta t \rightarrow 0$ while keeping D constant, and transforming number of molecules into concentrations, C ($C = n/vol$) leads to

$$\frac{dC}{dt} = D \frac{d^2C}{dx^2} \tag{1.49}$$

which is Fick's law. This is a partial differential equation in time and space and describes diffusion in a continuous setting. Solving it is beyond the scope of this course.

Example: diffusion from a peaked distribution

A concentration peaked at a single point in space will diffuse as shown in the figure



□

Example: diffusion times

The time it takes for a diffusive substrate to “reach” a distance L can be approximated by

$$t = \frac{L^2}{2D} \quad (1.50)$$

The value of the diffusion constant, D , for a small molecule (e.g. glucose) is in the order of $10^{-9} \text{ m}^2/\text{s}$. Given a cell size of $L \approx 50 \text{ } \mu\text{m}$ the diffusion time within a cell is approximately

$$t = \frac{(50 \times 10^{-6})^2}{2 \times 10^{-9}} \approx 3 \text{ s} \quad (1.51)$$

while a macroscopic length as $L = 1\text{m}$ would give

$$t = \frac{(1)^2}{2 \times 10^{-9}} \approx 5 \times 10^8 \text{ s} \approx 16 \text{ years !} \quad (1.52)$$

□

If diffusion is included in a model, it can be integrated as an ordinary differential equation

on a discretized space, where the formulation is

$$\frac{dC_i}{dt} = D \left(\sum_j^{N_{neigh}} (C_j - C_i) \right) \quad (1.53)$$

where i is a compartment index and the sum over j is the N_{neigh} neighbors. Here the distances and cross section areas between compartments are assumed to be equal and included in the diffusion constant D .

Example: diffusion between two compartments

The diffusion rate is proportional to the molecular concentration, similar to what is given for a mass action reaction. This is particular apparent in the case of two compartments where the spatial factors are incorporated in the diffusion constant. Assume diffusion of molecule A between compartments i and j .

$$A_i \xrightleftharpoons[D]{D} A_j \quad (1.54)$$

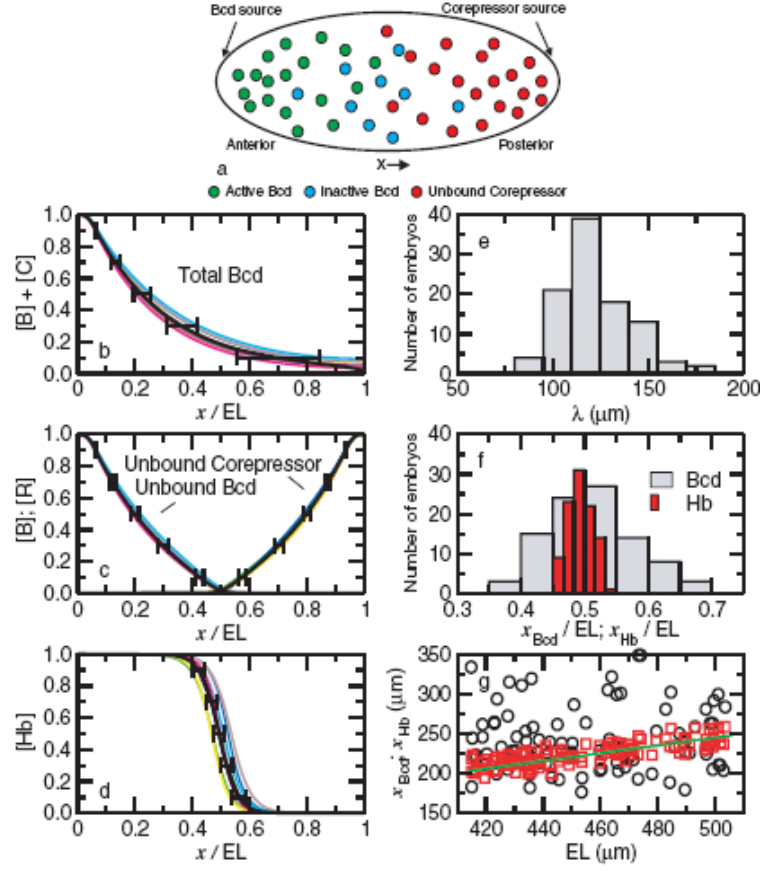
The resulting differential equations are given by

$$\frac{dA_i}{dt} = -\frac{dA_j}{dt} = -DA_i + DA_j \quad (1.55)$$

□

Example: early patterning in *Drosophila*

Diffusing signalling molecules (morphogens) are important for regulating development in multicellular organisms. In the *Drosophila* embryo, bicoid mRNA is deposited at the anterior pole (a localized source). This model by Howard *et.al.* (2005) discuss how this robustly can lead to a very precise gene expression pattern.



For the interested reader, the model equations are provided.

$$\begin{aligned}
 \frac{\partial [B]}{\partial t} &= D \frac{\partial^2 [B]}{\partial x^2} - \mu[B] - \nu[B][R] + J_B \delta(x - x_B(t)) \\
 \frac{\partial [R]}{\partial t} &= D \frac{\partial^2 [R]}{\partial x^2} - \mu[R] - \nu[B][R] + J_R \delta(x - x_R(t)) \\
 \frac{\partial [C]}{\partial t} &= D \frac{\partial^2 [C]}{\partial x^2} - \mu[C] + \nu[B][R] \\
 \frac{\partial [hb]}{\partial t} &= D_{hb} \frac{\partial^2 [hb]}{\partial x^2} - \mu_{hb}[hb] \\
 &\quad + \frac{\beta[B]^3(\eta[Hb] + \gamma K)}{K^4 + [B]^3(\eta[Hb] + K)} \\
 \frac{\partial [Hb]}{\partial t} &= D_{Hb} \frac{\partial^2 [Hb]}{\partial x^2} - \mu_{Hb}[Hb] + \alpha[hb].
 \end{aligned}$$

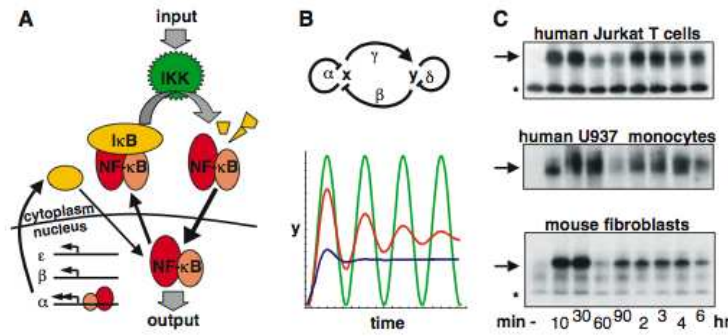
□

1.10.2 Membrane transport

Transport across membranes are important and can be both between cytosol and nuclei, or in and out of cells. Molecule transport across membranes can be both passive and active (mediated by helper molecules). Passive transport resembles diffusion in character i.e. driven by concentration gradients. Active transport is typically modeled similar to enzyme reactions, where a helper molecule (enzyme) does not change in concentration but can be saturated. A main difference compared to reactions is that one has to take into account that the number of molecules leaving from one side of the membrane has to be the same as reaching the other side. This typically means that the change in concentration is not the same on both sides of the membrane.

Example: NF κ B

Hoffmann et al (2002) presented a model of the NF κ B pathway where an integral investigation was on the transport in and out of the nuclei for different isoforms of a molecule complex (figure).



The cross-membrane transport was assumed to be passive and was modeled by terms defined by e.g.

$$\frac{dNF\kappa B_{nucl}}{dt} = k_1 NF\kappa B_{cyt} - k_{01} NF\kappa B_{nucl} \quad (1.56)$$

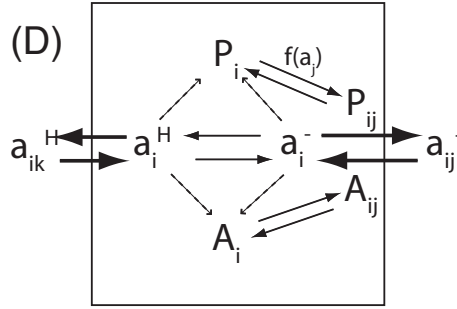
Note that the author did not account for the difference in volume between cytosol and nuclei, something that has been corrected for in later versions of the model. □

Example: polarized auxin transport

The plant hormone auxin is important for several developmental features in plants. It has been showed that the polar (directed) transport is a main regulator of the auxin location. Auxin can be in a charged (anion) form, which passivly pass through membranes, and a uncharged (protenated) form, which requires helper molecules for membrane crossing. A model for auxin flux (from cell to wall) is given by (see figure)

$$\begin{aligned}
 J_a = & p_a^H (f_a^{cell} a_i - f_a^{wall} a_{ij}) \\
 & + p_{PIN} p_{ij} \left(N(\Phi) \frac{f_a^{cell} a_i}{K_p + f_a^{cell} a_i} - N(-\Phi) \frac{f_a^{wall} a_{ij}}{K_p + f_a^{wall} a_{ij}} \right) \\
 & + p_{AUX} p_{ij} \left(N(-\Phi) \frac{f_a^{cell} a_i}{K_A + f_a^{cell} a_i} - N(\Phi) \frac{f_a^{wall} a_{ij}}{K_A + f_a^{wall} a_{ij}} \right)
 \end{aligned}$$

(plus additional spatial factors).



□

1.10.3 Reaction-Diffusion models

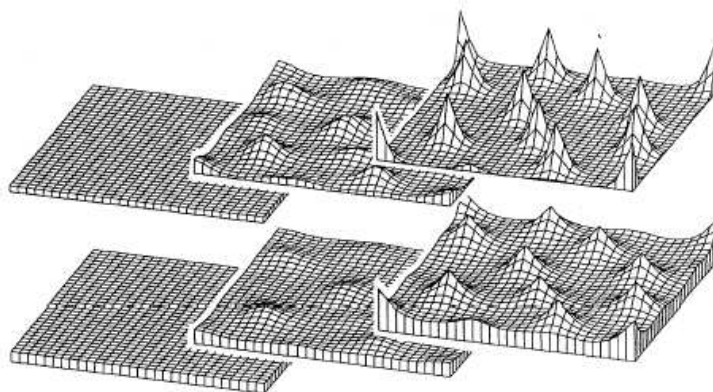
Models combining biochemical reactions and diffusion have the ability to create spatial patterns in molecular concentrations. This was first noted by Turing in the 1950s.

Example: the activator-inhibitor model

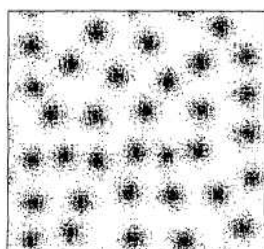
Meinhardt introduced an activator (a) inhibitor (h) reaction-diffusion model. The one dimensional version of the equations look like

$$\begin{aligned}
 \frac{da}{dt} &= \rho_a \left(\frac{a^2}{h} - a \right) + D_a \nabla^2 a \\
 \frac{dh}{dt} &= \rho_b (a^2 - h) + D_h \nabla^2 h
 \end{aligned} \tag{1.57}$$

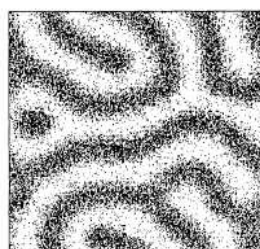
The figure shows the spontaneous pattern formation in activator (top) and inhibitor (bottom) concentrations when starting in a close to homogeneous state.



Different types of patterns of the activator, generated from different parameter sets, are shown in the figure below.



(b)

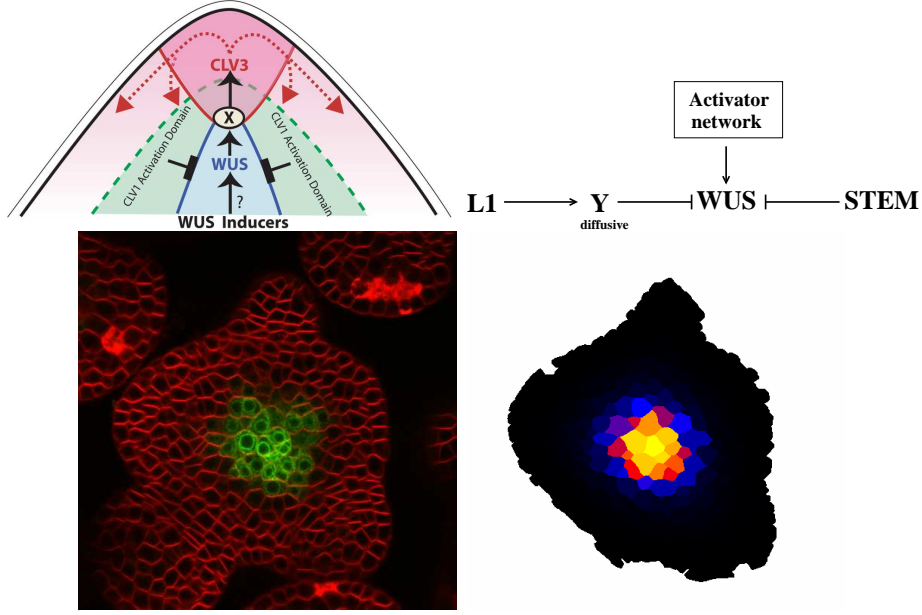


(c)



Example: stem cell regulation in plants

At the tip of a plant shoot, there is a pool of stem cells throughout the adult life of the plant. These cells are in part regulated by the WUS protein which is expressed in the interior of the shoot (see figure). This expression is very robust, and even removal of the shoot will lead to a new WUS domain forming. A model in which WUS is assumed to be induced by an activator network is capable of explaining this ability of reorganization.



For completeness, the equations are provided.

$$\frac{dW}{dt} = \frac{1}{\tau_w} g(h_w + T_{wa}A + T_{wy}Y) - d_w W \quad (1.58)$$

$$\frac{dY}{dt} = k_y L_1 - d_y Y + D_y \nabla^2 Y \quad (1.59)$$

$$\frac{dA}{dt} = a - (b + \beta)A + cA^2B - dYA + D_a \nabla^2 A \quad (1.60)$$

$$\frac{dB}{dt} = bA - cA^2B + D_b \nabla^2 B. \quad (1.61)$$

where $g(x)$ is the sigmoidal function

$$g(x) = \frac{1}{2} \left(1 + \frac{x}{\sqrt{1+x^2}} \right). \quad (1.62)$$

The parameter τ_i is the inverse maximal rate, and h_i sets the basal expression level. The T_{ij} parameters define the strength of the regulation (j regulating i). A positive T defines an activation, while a negative T leads to a repression. \square

Delay Differential Equation Models in Mathematical Biology

by

Jonathan Erwin Forde

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mathematics)
in The University of Michigan
2005

Doctoral Committee:

Assistant Professor Patrick W. Nelson, Chair
Professor Robert Krasny
Professor Jeffrey B. Rauch
Professor John W. Schiefelbein, Jr.
Professor Carl P. Simon

© Jonathan Erwin Forde 2006
All Rights Reserved

For my father, who pointed the way, and my mother, who helped me along it.

ACKNOWLEDGEMENTS

I would like to thank the many people who helped me reach this point. My advisor, Dr. Patrick Nelson, for taking me on as a student, introducing me to mathematical biology, and enduring my frequent tardiness. Dr. David Bortz, for always being available to help. Dr. Yang Kuang of Arizona State University, for his frequent insights into a difficult area and his faith in my abilities. The members of my committee, Professors Robert Krasny, Jeffrey Rauch, John Schiefelbein and Carl Simon, for their invaluable comments and patience. Thank you to Mom, Andrew, Katie and Grandma, my family: no matter where I have gone, home has always been with you. Without you all, this work would not have been possible. Finally, I cannot give enough thanks to my friend, colleague and office-mate Dr. Stanca Ciupe, for her friendship, supportiveness, all the conversation, and helping me find my way through graduate school.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vi
CHAPTER	
1. Preliminaries	1
1.1 Delay Differential Equations in Mathematical Biology	1
1.2 Basic Properties of Delay Differential Equations	2
1.3 Linear Delay Differential Equations with Constant Delays and Coefficients	4
1.4 The differential equation $\dot{z}(t) = az(t - \tau) - bz(t)$	5
1.5 A Comparison Lemma	7
1.6 Local Stability for Delay Differential Equations	8
1.6.1 The Pontriagin Criteria	8
1.6.2 Chebotarev's Theorem	9
1.6.3 Domain Subdivision	10
1.6.4 Frequency Methods	11
1.6.5 The Tsyppkin Criterion	12
2. Linear Stability Analysis via Sturm Sequences	14
2.1 General Method	14
2.1.1 Existence of Critical Delays	14
2.1.2 Nondegeneracy	18
2.2 Positive Real Roots and Sturm Sequences	20
2.3 Applications	23
2.4 General Order Two and Three Characteristic Equations	26
2.5 Conclusions	30
3. Single Species Models	32
3.1 A Fixed-Point Theorem from Nonlinear Functional Analysis	33
3.2 A General Single-Species Population Model with Delay	34
3.3 A Specific Single-Species Delay Model	40
3.3.1 Oscillatory Solutions	41
3.3.2 An Extension of Previously Known Results	45
3.4 Delay Dependent Parameters	48
3.5 Another General Model	50
3.6 Constant <i>per capita</i> Death Rates	53
3.7 Delay Dependent Parameters	59

4. Predator-Prey Interaction Models	64
4.1 The Lotka-Volterra Predator-Prey Interaction Model	64
4.2 A Delay Model of Predator-Prey Interaction	68
4.3 Preliminary Analysis	69
4.3.1 Positivity of Solutions	70
4.3.2 Uniform Boundedness of Solutions	70
4.3.3 Steady States	74
4.3.4 Linear Stability	75
4.4 Existence of Periodic Solution	78
4.4.1 The “Phase Plane”	80
4.4.2 Oscillation of Solutions	81
4.5 Future Work	86
5. Conclusion	88
BIBLIOGRAPHY	91

LIST OF FIGURES

Figure

3.1	The growth function, $b(x)x$, and the decay function, dx , intersecting at \bar{x}	42
3.2	The graph of $b^3e^{-1}e^{-be^{-1}}e^{-b^2e^{-1}}e^{-be^{-1}} - \ln(b)$ against b . When $b > e^2$ and this function is positive, we can prove the existence of periodic solutions to the delay differential equation (3.8)	48
3.3	The function $b(x)$, its tangent, and a line with slope greater than the tangent . . .	56
3.4	Solutions of the $\dot{x}(t) = (be^{-ax(t-\tau)} - d)x(t)$, with $a = 0.1$, $b = 10$, $d = 1$, with initial function $\bar{x} + 10t$ on $[-\tau, 0]$. $\tau_c = 0.6822$. The upper graph is for $\tau = 1$, and the second for $\tau = 0.5$	60
3.5	Solutions of the (3.35) with $a = 0.1$, $b = 10$, $d = 1$, $\mu = .7$, with initial function constantly 5 on $[-\tau, 0]$. The τ -region of instability determined in Theorem 3.20 is $[1.3520, 3.2894]$. The graphs are for $\tau = 0.7$, $\tau = 2$ and $\tau = 4$, respectively.	62
4.1	Periodic solutions of the Lotka-Volterra model with all parameters equal to 1 . . .	65
4.2	Solutions to the perturbed Lotka-Volterra model, $\varepsilon = .2$, $a = b = c = d = 1$	66
4.3	Global stability of $(1,0)$ in the absence of a nontrivial steady state	77
4.4	Global stability of (x^*, y^*) for small delays	79
4.5	Emergence of a stable limit cycle	80
4.6	Chaotic solutions in the phase plane	81
4.7	Time series for a chaotic solution	82
4.8	The Division of the phase planes in to the regions R_i	83

CHAPTER 1

Preliminaries

1.1 Delay Differential Equations in Mathematical Biology

The use of ordinary and partial differential equations to model biological systems has a long history, dating to Malthus, Verhulst, Lotka and Volterra. As these models are used in an attempt to better our understanding of more and more complicated phenomena, it is becoming clear that the simplest models cannot capture the rich variety of dynamics observed in natural systems. There are many possible approaches to dealing with these complexities. On one hand, one can construct larger systems of ordinary or partial differential equations, *i.e.*, systems with more differential equations. These systems can be quite good at approximating observed behavior, but they suffer from the downfall of containing many parameters, often signifying quantities which cannot be determined experimentally. Furthermore, obtaining an intuitive sense of which components are most important in determining a behavior regime can be quite difficult.

Another approach which is gaining prominence is the inclusion of time delay terms in the differential equations. The delays or lags can represent gestation times, incubation periods, transport delays, or can simply lump complicated biological processes together, accounting only for the time required for these processes to occur. Such

models have the advantage of combining a simple, intuitive derivation with a wide variety of possible behavior regimes for a single system. On the negative side, these models hide much of the detailed workings of complex biological systems, and it is sometimes precisely these details which are of interest. Delay models are becoming more common, appearing in many branches of biological modelling. They have been used for describing several aspects of infectious disease dynamics: primary infection [10], drug therapy [38] and immune response [11], to name a few. Delays have also appeared in the study of chemostat models [56], circadian rhythms [47], epidemiology [12], the respiratory system [51], tumor growth [52] and neural networks [7].

Statistical analysis of ecological data ([49], [50]) has shown that there is evidence of delay effects in the population dynamics of many species.

1.2 Basic Properties of Delay Differential Equations

While similar in appearance to ordinary differential equations, delay differential equations have several features which make their analysis more complicated. Let us examine an example of the form

$$(1.1) \quad \dot{x}(t) = f(x(t), x(t - \tau)).$$

To begin with, an initial value problem requires more information than an analogous problem for a system without delays. For an ordinary differential system, a unique solution is determined by an initial point in Euclidean space at an initial time t_0 . For a delay differential system, one requires information on the entire interval $[t_0 - \tau, t_0]$. Clearly, to know the rate of change at t_0 , one needs $x(t_0)$ and $x(t_0 - \tau)$, and for $\dot{x}(t_0 + \varepsilon)$, one needs to know $x(t_0 + \varepsilon)$ and $x(t_0 + \varepsilon - \tau)$. So, in order of the initial value problem to make sense, one needs to give an initial function or initial history,

the value of $x(t)$ for the interval $[-\tau, 0]$. Each such initial function determines a unique solution to the delay differential equation. If we require that initial functions be continuous, then the space of solutions has the same dimensionality as $\mathcal{C}([t_0 - \tau, t_0], \mathbb{R})$. In other words, it is infinite dimensional.

This infinite dimensional nature of delay differential equations is apparent in the study of linear systems. Just as for ordinary differential equations, one seeks exponential solutions, and computes a characteristic equation. Rather than a polynomial equation, one arrives at a transcendental equation of the form

$$P_0(\lambda) + P_1(\lambda)e^{-\lambda\tau} = 0,$$

where P_0 and P_1 are polynomial in λ . Generally, this equation has infinitely many solutions, corresponding to an infinite family of independent solutions to the linear differential equation [17]. The linear stability analysis is thus more difficult for these differential equations. Although standard methods for determining the location of roots of a polynomial (the Routh-Hurwitz criteria, see [16]) are not applicable here, there are methods available (see the next section and Chapter 2).

While as a general rule, the behavior of delay differential equations is “worse” than that of ordinary differential equations, this is not always the case. An excellent example is provided in [6]. It is well known that the solutions to $\dot{x}(t) = x(t)^2$ diverge to infinity in finite time. Solutions to the delay differential equation $\dot{x}(t) = x(t - \tau(t))^2$, however, are continuable for all time if $\tau(t)$ is positive for all t . In the case of a constant delay, the type with which we will be mostly concerned, this can be seen by the method of steps, that is, direct integration over intervals of length τ .

1.3 Linear Delay Differential Equations with Constant Delays and Coefficients

Next we explore the relationship between the location of the roots of the characteristic equation and the behavior of solutions of the linear system. In particular, we will see that equivalence between the stability of the zero solution and the location of all characteristic roots in the right half-plane holds for delay differential equations, just as for ordinary differential equations.

Consider a first order delay differential equation

$$(1.2) \quad \dot{x}(t) = \sum_{i=1}^m A_i x(t - \tau_i),$$

where A_i is a constant $n \times n$ matrix for all i , and $0 \leq \tau_i \leq \tau$ for all i and some fixed τ . As usual, any higher order linear system is equivalent to this by adding dummy variables. The characteristic equation of this system is

$$(1.3) \quad \det \left(\lambda I - \sum_{i=1}^m A_i e^{-\lambda \tau_i} \right) = 0.$$

We have the following two theorems, which can be found in [15].

Theorem 1.1. *Given any real number ρ , the characteristic equation (1.3) has at most a finite number of roots λ such that $\operatorname{Re}(\lambda) \geq \rho$.*

Essentially, the preceding theorem says that “most” of the roots of the equation (1.3) have negative real part. Furthermore, the roots cannot accumulate except about $\operatorname{Re}(\lambda) = -\infty$. In much of our future analysis, we will be interested in the space $\mathcal{C}([-r, 0], \mathbb{R})$, representing all initial functions. When endowed with the norm

$$\|\phi\| = \sup_{t \in [-r, 0]} \phi(t),$$

this is a Banach space.

Theorem 1.2. *If $\operatorname{Re}(\lambda) < \rho$ for every solution of the characteristic equation (1.3), then there exists a constant $M > 0$ such that, for each $\phi \in \mathcal{C}([t_0 - r, t_0], \mathbb{R})$, the solution to (1.2) satisfies*

$$\|y(t; \phi)\| \leq M \|\phi\| e^{\rho(t-t_0)}$$

So the behavior of linear delay differential equations is given an upper bound by the location of the eigenvalue with the largest real part. By combining these two results, we arrive at the following result, which forms the foundation of our linear stability analysis.

Corollary 1.3. *If $\operatorname{Re}(\lambda) < 0$ for every solution of the characteristic equation (1.3), then there exist constants $M, \gamma > 0$ such that, for each $\phi \in \mathcal{C}([t_0 - r, t_0], \mathbb{R})$, the solution to (1.2) satisfies*

$$\|y(t; \phi)\| \leq M \|\phi\| e^{-\gamma(t-t_0)}$$

In other words, if all of the eigenvalues have negative real part, then solutions to the linear delay differential equation decay exponentially to 0, exactly as is the case for ordinary differential equations.

1.4 The differential equation $\dot{z}(t) = az(t - \tau) - bz(t)$

We will often encounter the linear delay differential equation $\dot{z}(t) = az(t - \tau) - bz(t)$ when studying more complex equations. It is therefore useful to establish some of its basic properties at the outset.

Lemma 1.4. *If $|a| < b$, then all solutions of the differential equation $\dot{z}(t) = az(t - \tau) - bz(t)$ approach 0 as $t \rightarrow \infty$.*

Proof. Assuming a solution of the form $e^{\lambda t}$, we arrive at the characteristic equation for this equation,

$$(1.4) \quad \lambda = ae^{-\lambda\tau} - b.$$

We begin by showing that the real part of any solution to this differential equation is negative. Let $\lambda = \mu + i\sigma$. Then we have

$$\begin{aligned} \mu + i\sigma &= ae^{-\mu\tau}e^{-i\sigma\tau} - b \\ &= ae^{-\mu\tau}(\cos(\sigma\tau) - i\sin(\sigma\tau)) - b. \end{aligned}$$

Looking at the real part of this equation, we get

$$(1.5) \quad \mu + b = ae^{-\mu\tau} \cos(\sigma\tau).$$

If $\mu \geq 0$, then we get

$$b \leq \mu + b = ae^{-\mu\tau} \cos(\sigma\tau) \leq ae^{-\mu\tau} \leq a,$$

contradicting the assumption that $|a| < b$.

So all of the roots of this differential equation have negative real part. It is a simple application of Corollary 1.3 to see that then all solutions have a bound of the form

$$|z(t)| \leq Me^{-\gamma t}.$$

Thus, we see that solutions must approach 0 as $t \rightarrow \infty$. □

When the coefficients a and b are equal, solutions need not approach 0, but we can show that they do indeed approach some positive limit determined by the initial history ϕ . The proof of this lemma relies on the method of the Laplace transform. An excellent description of this theory in application to linear delay differential equations can be found in the textbook by Bellman and Cooke [2].

1.5 A Comparison Lemma

We will also be interested in a differential equation of the form

$$\dot{y}(t) = p(t)y(t - \tau) - dy(t),$$

where $p(t) \leq d$, $d > 0$. In practice, $p(t)$ will represent the nonlinearities of the model equation. To better understand the behavior of this system, we will try to compare its dynamics with those of the system

$$\dot{z}(t) = dz(t - \tau) - dz(t).$$

We begin with the following lemma.

Lemma 1.5. *If y and z are defined as above, and $y(t) = z(t) \geq 0$ for $t \in [a, a + \tau]$ for some a , then $y(t) \leq z(t), \forall t$.*

Proof. We define new variables $y_1(t) = e^{dt}y(t)$ and $z_1(t) = e^{dt}z(t)$. Then a simple calculation shows that

$$\dot{y}_1(t) = p(t)e^{d\tau}y_1(t - \tau)$$

$$\dot{z}_1(t) = de^{d\tau}z_1(t - \tau).$$

Also, for nonnegative initial data, $y_1(t)$ and $z_1(t)$ are nonnegative and nondecreasing for $t \geq a$. Now we examine the difference $w_1(t) = z_1(t) - y_1(t)$. This quantity is governed by the differential equation

$$\begin{aligned} \dot{w}_1(t) &= de^{d\tau}z_1(t - \tau) - p(t)e^{d\tau}y_1(t - \tau) \\ &\geq e^{d\tau}(dz_1(t - \tau) - dy_1(t - \tau)) \\ &= de^{d\tau}w_1(t - \tau) \end{aligned}$$

Suppose that $w_1(t) \geq 0$ for $t \in [a, T]$, $T \geq a + \tau$, then the inequality above means that $w_1(t)$ is nondecreasing for $t \in [T, T + \tau]$, and therefore $w_1(t) \geq 0$ on $[-\tau, T + \tau]$.

Now begin with the fact that $w_1(t) = 0$ for $t \in [a, a + \tau]$, and repeating the above argument shows that $w_1(t) \geq 0$ for $t \geq a$. It then follows immediately that $z(t) \geq y(t)$ for $t \geq a$. \square

1.6 Local Stability for Delay Differential Equations

For ordinary differential equations, the local stability of a steady state depends on the location of roots of the characteristic function, which is polynomial in form. The steady state is stable if and only if all of the roots have negative real part. The well-known Routh-Hurwitz criteria give precise conditions for this to occur for arbitrary polynomials. For delay differential equations, local stability is also determined by the location of the characteristic function, but in this case, this function takes the form of a so-called quasipolynomial, which is transcendental. Thus, there are infinitely many roots. Furthermore, the Routh-Hurwitz criteria are not applicable. Many approaches have been taken to determine the stability of steady states delay equations. Below, I present a brief survey of these methods, before moving to develop a new method available for certain delay systems.

1.6.1 The Pontriagin Criteria

When the delays in a system are commensurate, meaning that all are integer multiples of some fixed quantity, the characteristic function can be written in the form

$$(1.6) \quad D_1(z) = \sum_{\ell=0}^m \sum_{j=1}^r a_{\ell j} z^{\ell} e^{zj},$$

and if we set $z = i\sigma$, we can break this into real and imaginary parts as

$$D_1(i\sigma) = g(\sigma) + if(\sigma).$$

Pontriagin proved the following in [43], and a simplified proof can be found in [44].

Theorem 1.6. *If the roots of (1.6) all have negative real part, then all of the zeros of f and g are real, simple, and alternating, and*

$$\dot{g}(\sigma)f(\sigma) - g(\sigma)\dot{f}(\sigma) > 0, \quad \forall \sigma \in \mathbb{R}.$$

Furthermore, either of the following conditions is sufficient for stability.

1. *All zeros of f and g are real, alternating and simple, and the inequality above is fulfilled for at least one σ .*
2. *All zeros of g (or f) are real and simple, and for each zero, the inequality is satisfied.*

In practicality, these criteria suffer from several drawbacks. In the case of multiple delays, Theorem 1.6 holds only when the delays are commensurate, *i.e.*, when they are rational multiples of some common factor. In general, multiple delay systems are not equivalent to systems with commensurate delays. Even when there is only one delay, it is very difficult to determine the relationship between roots of the functions f and g , and the theorem provides no method for determining whether its hypotheses are satisfied or not.

1.6.2 Chebotarev's Theorem

Another approach has been to try to generalize the Routh-Hurwitz criteria directly [8]. To this end, we can take an expansion of the characteristic function as an infinite

series,

$$D_1(z) = a_0 + a_1z + a_2z^2 + \cdots.$$

Then we can again write $D_1(i\sigma) = u(\sigma) + iv(\sigma)$, and we will have

$$u(\sigma) = a_0 - a_2\sigma^2 + a_4\sigma^4 - + \cdots$$

$$v(\sigma) = a_1 - a_3\sigma^3 + a_5\sigma^5 - + \cdots$$

Then we can define determinants, as in the Routh-Hurwitz criteria,

$$\begin{aligned} Q_1 &= a_1 \\ Q_2 &= \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix} \\ &\vdots \\ Q_m &= \begin{vmatrix} a_1 & a_3 & a_5 & \cdots & a_{2m-1} \\ a_0 & a_2 & a_4 & \cdots & a_{2m-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_m \end{vmatrix}. \end{aligned}$$

We then have the following theorem.

Theorem 1.7. *Assume that $u(z)$ and $v(z)$ have no common zeros. Then the quasipolynomial D_1 is stable if and only if $Q_m > 0$ for all $m \in \mathbb{N}_0$.*

While similar in form to the Routh-Hurwitz criteria, this result is nearly impossible to apply, due to the infinite number of inequalities which must be verified.

1.6.3 Domain Subdivision

The method of domain subdivision or D-subdivision, uses some basic facts about the behavior of the roots of characteristic functions as a parameter changes to divide

parameter space into regions in which the number of roots with positive real parts is constant. The location of the roots depends continuously on the parameters of the model, and, as the parameters change, a new root can emerge in the right half-plane only if there is a set of parameters for which a purely imaginary root exists.

One may now subdivide the parameter space (domain) by hypersurfaces consisting of parameter regimes for which one or more purely imaginary roots exist. In the regions bounded by these hypersurfaces, the number of roots with positive real part is constant. Of course, the regions in which the number is zero and their complements are of most interest. This method is particularly easy to visualize when the system in question depends on two parameters, so that the domain is \mathbb{R}^2 and the hypersurfaces are curves.

1.6.4 Frequency Methods

A class of stability methods making use of the argument principle and a frequency response curve are particularly popular in control theory applications. The first of these is the Michailov criterion. If we consider an n -th order system with characteristic function $\Delta(z)$, then we have the following theorem.

Theorem 1.8 (Michailov Criterion). *A steady state with characteristic function Δ is asymptotically stable if and only if*

$$\arg \Delta(i\sigma)|_{\sigma=0}^{\sigma=\infty} = \frac{n\pi}{2}.$$

Unfortunately the graphical form of the curve $\Delta(i\sigma)$ in the complex plane is difficult to determine when a delay is included, especially when the length of the delay is varied.

A closely related criterion was developed by Nyquist. To begin with, one obtains the transfer function $W(s)$ from the Laplace transform of the linearized system, and

one then defines the frequency response to be $W(i\sigma)$.

Theorem 1.9 (Nyquist Criterion). *Suppose the open loop system is stable. Then the closed loop system is stable if and only if the frequency response of the open loop system does not enclose -1 .*

The complexity of the graphical form of the frequency response again makes the direct application of this criterion difficult. A variation on these themes can make the criteria easier to check, for example, with a computer computation, rather than graphical analysis. We begin by writing $\Delta(i\sigma) = U(\sigma) + iV(\sigma)$ and defining

$$R(\sigma) = \frac{U(\sigma)V'(\sigma) - U'(\sigma)V(\sigma)}{U^2(\sigma) + V^2(\sigma)}.$$

Theorem 1.10. *A steady state with characteristic function Δ and order n is asymptotically stable if and only if*

$$\int_0^\infty R(\sigma)d\sigma = \frac{n\pi}{2}.$$

1.6.5 The Tsyarkin Criterion

Finally we arrive at the method for analyzing linear stability which is most closely associated with the techniques we will develop in the next chapter. This criterion will provide necessary and sufficient conditions for the roots of the characteristic equation to remain in the left half plane for all lengths of delay. We look again at the transfer function, which, for a system with a single delay, τ , has the form

$$(1.7) \quad \frac{R(s)}{Q(s)}e^{s\tau},$$

where R and Q are polynomials of degrees $n - 1$ and n respectively. We then have

Theorem 1.11 (Tsyppkin Criterion). *Let Q be a stable polynomial, then the characteristic function Δ is stable for all delays τ if and only if*

$$|Q(i\sigma)| > |R(i\sigma)|$$

for all $\sigma \in \mathbb{R}$.

In Chapter 2, we will arrive at the same result by a different route on our way to finding more explicit conditions for the persistence of stability for all delays.

There is also a generalization of this criterion, due to El'sgol'ts [17], to the case of multiple delays τ_i , $i = 1, \dots, m$. In this case, the numerator of the transfer function (1.7) has the form

$$\sum_{i=1}^m R_i(s) e^{-s\tau_i}.$$

A necessary and sufficient condition for stability in this case is that Q be stable and

$$|Q(i\sigma)| > \sum_{i=1}^m |R(i\sigma)|.$$

CHAPTER 2

Linear Stability Analysis via Sturm Sequences

2.1 General Method

In this chapter, a new method for analyzing the stability of a steady state of a delay differential equation is introduced. As we have seen in our survey of methods for linear stability analysis, the introduction of a delay significantly increases the difficulty of locating the roots of the characteristic equation. Once a delay is included in a model, it is often of interest to determine whether or not varying the delay length can change the stability characteristics of a steady state. So, we will focus particularly on one approach: treating the length of the delay as a bifurcation parameter.

A stable steady state can become unstable if, by increasing the delay, a characteristic root changes from having a negative real part to having positive real part, and this occurs only if this root traverses the imaginary axis.

2.1.1 Existence of Critical Delays

At a steady state, the characteristic equation of the delayed differential equation will have the form

$$(2.1) \quad P(\lambda, \tau) \equiv P_1(\lambda) + P_2(\lambda)e^{-\lambda\tau} = 0,$$

where τ is the length of the discrete delay added, and P_1 and P_2 are polynomials.

We can rewrite (2.1) as

$$\sum_{j=0}^N a_j \lambda^j + e^{-\lambda\tau} \sum_{j=0}^M b_j \lambda^j = 0.$$

Assume that the steady state about which we have linearized is stable in the absence of a delay. Then for $\tau = 0$ all of the roots of the polynomial have negative real part. As τ varies, these roots change. We are interested in any critical values of τ at which a root of this equation transitions from having negative to having positive real parts. If this is to occur, there must be a boundary case, a critical value of τ , such that the characteristic equation has a purely imaginary root (see [17]). The following demonstrates how to determine whether or not such a τ exists, by reducing (2.1) to a polynomial problem and seeking particular types of roots, thus determining whether a bifurcation can occur as a result of the introduction of delay.

We begin by looking for a purely imaginary root, $i\sigma$, $\sigma \in \mathbb{R}$ of (2.1)

$$P_1(i\sigma) + P_2(i\sigma)e^{-i\sigma\tau} = 0.$$

We break the polynomial up into its real and imaginary parts, and write the exponential in terms of trigonometric functions to get

$$(2.2) \quad R_1(\sigma) + iQ_1(\sigma) + (R_2(\sigma) + iQ_2(\sigma))(\cos(\sigma\tau) - i\sin(\sigma\tau)) = 0.$$

In terms of the original polynomial coefficients, the new polynomials are

$$\begin{aligned} R_1(\sigma) &= \sum_j (-1)^{j+1} a_{2j} \sigma^{2j}, \\ Q_1(\sigma) &= \sum_j (-1)^j a_{2j+1} \sigma^{2j+1}, \\ R_2(\sigma) &= \sum_j (-1)^{j+1} b_{2j} \sigma^{2j}, \\ Q_2(\sigma) &= \sum_j (-1)^j b_{2j+1} \sigma^{2j+1}, \end{aligned}$$

Note that because $i\sigma$ is purely imaginary, R_1 and R_2 are even polynomials of σ , while Q_1 and Q_2 are odd polynomials.

In order for (2.2) to hold, both the real and imaginary parts must be 0, so we get the pair of equations

$$\begin{aligned} R_1(\sigma) + R_2(\sigma) \cos(\sigma\tau) + Q_2(\sigma) \sin(\sigma\tau) &= 0, \\ Q_1(\sigma) - R_2(\sigma) \sin(\sigma\tau) + Q_2(\sigma) \cos(\sigma\tau) &= 0, \end{aligned}$$

which we can rewrite as

$$\begin{aligned} (2.3) \quad -R_1(\sigma) &= R_2(\sigma) \cos(\sigma\tau) + Q_2(\sigma) \sin(\sigma\tau), \text{ and} \\ Q_1(\sigma) &= R_2(\sigma) \sin(\sigma\tau) - Q_2(\sigma) \cos(\sigma\tau). \end{aligned}$$

Squaring each equation and summing the results yields

$$(2.4) \quad R_1(\sigma)^2 + Q_1(\sigma)^2 = R_2(\sigma)^2 + Q_2(\sigma)^2.$$

We notice two things about this equation. First, this is a polynomial equation. The trigonometric terms disappear, and the delay, τ , has been eliminated. Secondly, it is an equality of *even* polynomials. This is because squaring an even or odd function always result in an even function, i.e., $f(-x)^2 = (\pm f(x))^2 = f(x)^2$.

Define a new variable $\mu = \sigma^2 \in \mathbb{R}$. Then equation (2.4) above can be written in terms of μ as

$$(2.5) \quad S(\mu) = 0,$$

where S is a polynomial. Note that we are only interested in $\sigma \in \mathbb{R}$, and thus if all of the real roots of S are negative, we will have shown that there can be no simultaneous solution σ^* of (2.3). Conversely, if there is a positive real root μ^* to S , there is a delay τ corresponding to $\sigma^* = \pm\sqrt{\mu^*}$ which solves both equations in (2.3).

To see this, suppose that we have found a σ^* such that $R_1(\sigma^*)^2 + Q_1(\sigma^*)^2 = R_2(\sigma^*)^2 + Q_2(\sigma^*)^2$. Let $C = \sqrt{R_2(\sigma^*)^2 + Q_2(\sigma^*)^2}$. The preceding equation then can be interpreted as stating that the point $(-R_1(\sigma^*), Q_1(\sigma^*))$ lies on the circle of radius C (the negative sign is for convenience later). Now let us return to the equations for the real and imaginary parts of the characteristic equation. These can now be written as:

$$\begin{aligned} -R_1(\sigma^*) &= C \left(\frac{R_2(\sigma^*)}{C} \cos(\sigma^* \tau) + \frac{Q_2(\sigma^*)}{C} \sin(\sigma^* \tau) \right), \text{ and} \\ Q_1(\sigma^*) &= C \left(\frac{R_2(\sigma^*)}{C} \sin(\sigma^* \tau) - \frac{Q_2(\sigma^*)}{C} \cos(\sigma^* \tau) \right). \end{aligned}$$

We can then write $\frac{R_2(\sigma^*)}{C} = \cos \alpha$ and $\frac{Q_2(\sigma^*)}{C} = \sin \alpha$, and then

$$\begin{aligned} -R_1(\sigma^*) &= C \cos(\sigma^* \tau - \alpha), \text{ and} \\ Q_1(\sigma^*) &= C \sin(\sigma^* \tau - \alpha). \end{aligned}$$

Since the point $(-R_1(\sigma^*), Q_1(\sigma^*))$ lies on the circle of radius C , it is then clear that there is a positive value $\tau = \tau^*$ that satisfies both equations simultaneously.

Should the polynomial (2.5) have more than one positive real root, we are interested in studying the one associated with the smallest delay, τ^* .

An alternate approach, more geometrical in nature, on finding the roots of the characteristic equation (2.1) is taken in [35] and [33]. In this case, for $\lambda = i\sigma$, we rewrite (2.1) as

$$(2.6) \quad -\frac{P_1(i\sigma)}{P_2(i\sigma)} = e^{-i\sigma\tau}.$$

As τ varies, plotting the right hand side in the complex plane traces out a unit circle, and the left hand side is a rational curve. The intersections of these curves represent the critical delays in which we are interested. Thus finding the roots of the

characteristic equation comes down to finding values of σ for which the left hand side of (2.6) has modulus 1. This reproduces equation (2.4), and the freedom to choose τ again ensures that the original characteristic polynomial (2.1) is satisfied for some τ^* .

2.1.2 Nondegeneracy

Having found a critical delay τ^* and the point $z = i\sigma^*$ at which a root of the characteristic equation hits the imaginary axis, it is necessary to confirm that the root continues into the positive half-plane as τ increases past τ^* . The criterion for this to occur is

$$\left. \frac{d}{d\tau} \operatorname{Re}(\lambda) \right|_{\lambda=i\sigma^*, \tau=\tau^*} > 0.$$

Equivalent in this case is

$$\left. \frac{d}{d\tau} \operatorname{Re}(\lambda) \right|_{\lambda=i\sigma^*, \tau=\tau^*} \neq 0,$$

since it is known for $\tau < \tau^*$ that all solutions λ to (2.1) have negative real part.

Lemma 2.1. *If $\lambda = i\sigma^*$ and $\tau = \tau^*$ satisfy the characteristic equation (2.1), then*

$$\left. \frac{d}{d\tau} \operatorname{Re}(\lambda) \right|_{\lambda=i\sigma^*, \tau=\tau^*} > 0$$

if and only if

$$(2.7) \quad R_1(\sigma^*)R_1'(\sigma^*) + Q_1(\sigma^*)Q_1'(\sigma^*) \neq R_2(\sigma^*)R_2'(\sigma^*) + Q_2(\sigma^*)Q_2'(\sigma^*).$$

Proof. Beginning with the characteristic equation (2.1), we can write

$$e^{-\lambda\tau} = -\frac{P_1(\lambda)}{P_2(\lambda)},$$

which implies,

$$-\lambda\tau = \log \left(-\frac{P_1(\lambda)}{P_2(\lambda)} \right).$$

Taking the derivative with respect to τ (treating λ as a function of τ , $\lambda = \lambda(\tau)$) gives

$$-\lambda - \tau \frac{d\lambda}{d\tau} = \frac{P'_1(\lambda)P_2(\lambda) - P_1(\lambda)P'_2(\lambda)}{P_1(\lambda)P_2(\lambda)} \cdot \frac{d\lambda}{d\tau},$$

where $' = \frac{d}{d\lambda}$. At $\lambda = i\sigma^*$ and $\tau = \tau^*$, the left hand side becomes $-i\sigma^* - \tau^* \frac{d\lambda}{d\tau}$. Since $i\sigma^*$ is purely imaginary, and τ^* is real, $\frac{d\lambda}{d\tau}$ is purely imaginary if and only if

$$\frac{P'_1(i\sigma^*)P_2(i\sigma^*) - P_1(i\sigma^*)P'_2(i\sigma^*)}{P_1(i\sigma^*)P_2(i\sigma^*)}$$

is real. This occurs only when the numerator and denominator are real multiples of one another. Now we can write

$$\frac{P'_1(i\sigma^*)P_2(i\sigma^*) - P_1(i\sigma^*)P'_2(i\sigma^*)}{P_1(i\sigma^*)P_2(i\sigma^*)} = \frac{(Q'_1 - iR'_1)(R_2 + iQ_2) - (Q'_2 - iR'_2)(R_1 + iQ_1)}{(R_1 + iQ_1)(R_2 + iQ_2)}.$$

Collecting real and imaginary parts, we find that

$$\left. \frac{d}{d\tau} \text{Re}(\lambda) \right|_{\lambda=i\sigma^*, \tau=\tau^*} = 0$$

if and only if

$$\frac{Q'_1 R_2 + R'_1 Q_2 - Q'_2 R_1 - R'_2 Q_1}{R_1 R_2 - Q_1 Q_2} = \frac{Q'_1 Q_2 - R'_1 R_2 + R_1 R'_2 - Q_1 Q'_2}{R_1 Q_2 + R_2 Q_1}.$$

Cross multiplying and cancelling like terms yields

$$R_1 R'_1 (R_2^2 + Q_2^2) + Q_1 Q'_1 (R_2^2 + Q_2^2) = R_2 R'_2 (R_1^2 + Q_1^2) + Q_2 Q'_2 (R_1^2 + Q_1^2).$$

But at $\sigma = \sigma^*$, $R_1^2 + Q_1^2 = R_2^2 + Q_2^2 \neq 0$. So this reduces to the condition

$$R_1 R'_1 + Q_1 Q'_1 = R_2 R'_2 + Q_2 Q'_2.$$

This is a necessary and sufficient condition for

$$\left. \frac{d}{d\tau} \text{Re}(\lambda) \right|_{\lambda=i\sigma^*, \tau=\tau^*} = 0.$$

Thus the derivative is not equal to 0 if (2.7) holds. \square

Practically, this condition can be checked by formally differentiating the equation (2.4) with respect to σ and verifying that equality does not hold for $\sigma = \sigma^*$.

In summary, we have reduced the question of whether the introduction of a delay can cause a bifurcation to a problem of determining if a polynomial has any positive real roots. If such roots can be found, then the argument above guarantees that there is a delay size τ^* such that one of the eigenvalues of the system crosses the imaginary axis, destabilizing its critical point. We have proven the following:

Lemma 2.2. *Given a system of differential equations $\dot{x}(t) = f(x(t), x(t - \tau))$ with a discrete delay τ , and a stable steady state for x_s for $\tau = 0$, and let*

$$\sum_{i=1}^N a_i \lambda^i + e^{-\lambda \tau} \sum_{i=1}^M b_i \lambda^i = 0$$

be the characteristic equation of the system about x_s . Then there exists a $\tau^ > 0$ for which x_s undergoes a nondegenerate change of stability if and only if the equation*

i) $S(\mu) = 0$ (as defined in equation (2.5)) has a positive real root $\mu^ = (\sigma^*)^2$, such that*

ii) $S'(\mu^) \neq 0$*

That is, when μ^ is a simple, positive real root of the equation*

$$\left[\sum (-1)^j a_{2j} \mu^j \right]^2 + \mu \left[\sum (-1)^j a_{2j+1} \mu^j \right]^2 = \left[\sum (-1)^j b_{2j} \mu^j \right]^2 + \mu \left[\sum (-1)^j b_{2j+1} \mu^j \right]^2.$$

2.2 Positive Real Roots and Sturm Sequences

Once the polynomial equation (2.5) has been obtained, one must determine whether it has any positive real roots. There are many approaches one might take. For degree 2 characteristic polynomials, there is always the quadratic formula. For third and

fourth degree polynomials, there are also explicit algorithms (see, for example, [29] or [35]).

One approach to showing that no bifurcation exists is to apply the Routh-Hurwitz condition. If these conditions are satisfied, then all of the roots of (2.5) have negative real part, and thus none are positive and real. This condition is not sharp, however, since there remains the possibility that the polynomial (2.5) has a conjugate pair of roots with positive real part and nonzero imaginary part. For example, consider the characteristic polynomial

$$(2.8) \quad \lambda^2 + 3\lambda + 5 + \lambda e^{-\lambda\tau} = 0.$$

In the absence of delay, this becomes,

$$\lambda^2 + 4\lambda + 5 = 0,$$

which clearly has only roots with negative real part, and thus the steady state is stable. Explicitly, the roots are $\lambda_{1,2} = -2 \pm i$. The polynomial (2.5) produced by the process we have described is

$$\mu^2 - 2\mu + 25 = 0,$$

whose roots are $1 \pm 2i\sqrt{6}$. This polynomial has no positive real solution, and yet fails the Routh-Hurwitz conditions.

In other words, the Routh-Hurwitz conditions can guarantee the absence of a bifurcation, but cannot give conditions under which a bifurcation *does* occur with increasing τ .

A simple approach to determining whether a positive real root exists is Descartes' Rule of Signs, whereby the number of sign changes in the coefficients is equal to the number of positive real roots, modulo 2. If the number of sign changes is odd, then

a solution is guaranteed. If, however, the number of sign changes is even, the rule cannot distinguish between, for example, 2 roots and 0 roots.

A more general approach to this problem is Sturm sequences. Suppose that a polynomial f has no repeated roots. Then f and f' are relatively prime. Let $f = f_0$ and $f' = f_1$. We obtain the following sequence of equations by the division algorithm

$$\begin{aligned} f_0 &= q_0 f_1 - f_2, \\ f_1 &= q_1 f_2 - f_3, \\ &\vdots \\ f_{s-2} &= q_{s-2} f_{s-1} - K, \end{aligned}$$

where K is some constant.

The sequence of *Sturm functions*, $f_0, f_1, f_2, \dots, f_{s-1}, f_s (= K)$ is called a *Sturm chain*. We may determine the number of real roots of the polynomial f in any interval in the following manner: Plug in each endpoint of the interval, and obtain a sequence of signs. The number of real roots in the interval is the difference between the number of sign changes in the sequence at each endpoint. For a complete proof of the method of Sturm sequences, see [45].

Example: $f(x) = x^2 - 1$. In this case, $f' = 2x$, so the division algorithm is:

$$x^2 - 1 = \frac{x}{2} \cdot (2x) - 1.$$

So the Sturm chain is simply $x^2 - 1, 2x, 1$. If we are interested in the interval $[0, \infty)$, then the chains of signs are

at 0 : -, 0, + , and

at ∞ : +, +, +.

There is one sign change in the first sequence and zero in the last, and we conclude that there is one positive real root to $f(x)$. Similarly, suppose we were interested in the interval $[-2, 2]$. Then the sign sequences are

at -2 : +, -, + , and

at 2 : +, +, +.

There are two sign changes in the first sequence and zero in the second, confirming that there are two roots in this interval.

Given a specified parameter set, this method gives a simple, implementable algorithm for determining whether a bifurcation occurs, without the need to run the full simulation of the system of equations for various delays.

2.3 Applications

In [39], we are faced with the characteristic equation

$$(2.9) \quad \lambda^3 + A\lambda^2 + (B - \delta c e^{-\lambda\tau})\lambda + \delta c\rho - \delta c(\rho - \psi')e^{-\lambda\tau} = 0,$$

where $A \equiv \delta + c + \rho$, $B \equiv \delta c + (\delta + c)\rho$, and $\psi' \equiv \rho - d_T > 0$, the notation being that of the paper. In the paper, it is shown that for $\tau \ll 1$ and $\tau \gg 1$ no change of stability occurs. We can extend this result to all $\tau > 0$.

In the notation we have been using, equation (2.9) yields

$$R_1(\sigma) = -A\sigma^2 + \delta c\rho,$$

$$Q_1(\sigma) = -\sigma^3 + B\sigma,$$

$$R_2(\sigma) = -\delta c d_T,$$

$$Q_2(\sigma) = -\delta c\sigma.$$

Using these specific polynomials, (2.4) becomes

$$(2.10) \quad \begin{aligned} &\sigma^6 + (A^2 - 2B)\sigma^4 + (B^2 - (\delta c)^2 - 2\delta c\rho A)\sigma^2 - (\delta c)^2(\psi'^2 - 2\rho\psi') = 0, \text{ or} \\ &\mu^3 + (A^2 - 2B)\mu^2 + (B^2 - (\delta c)^2 - 2\delta c\rho A)\mu - (\delta c)^2(\psi'^2 - 2\rho\psi') = 0. \end{aligned}$$

This can be simplified by substituting the known values of A , B , and ψ' . For the μ^2 coefficient, we have

$$\begin{aligned} A^2 - 2B &= (\delta + c + \rho)^2 - 2(\delta c + (\delta + c)\rho) \\ &= \delta^2 + c^2 + \rho^2 + 2\delta c + 2\rho c + 2\delta\rho - 2\delta c - 2(\delta + c)\rho \\ &= \delta^2 + c^2 + \rho^2. \end{aligned}$$

Further, for the μ coefficient, we have

$$\begin{aligned} B^2 - 2\delta c\rho A - (\delta c)^2 &= ((\delta c)^2 + (\delta\rho)^2 + (c\rho)^2 + 2\delta^2 c\rho + 2\delta\rho c^2 + 2\rho^2\delta c) \\ &\quad - 2\delta c\rho(\rho + c + \delta) - (\delta c)^2 \\ &= (\delta\rho)^2 + (c\rho)^2. \end{aligned}$$

And for the constant term we have

$$\psi^2 - 2\rho\psi' = \psi'(\rho - d_T - 2\rho) = -\psi'(\rho + d_T).$$

So we may write equation (2.10) as

$$\mu^3 + (\delta^2 + c^2 + \rho^2)\mu^2 + ((\delta\rho)^2 + (c\rho)^2)\mu + (\delta c)^2\psi'(\rho + d_T) = 0.$$

This is a polynomial with positive coefficients, and cannot have any positive real roots, therefore the introduction of a delay into the model in Nelson and Perelson [39] cannot lead to a bifurcation. This is an extension of the results presented in that paper, where it was proven by asymptotic methods that for very large and very small delays, the steady state was stable. The argument above shows that this is the case for all delay lengths.

In [38], the following characteristic equation is encountered for a system of delay differential equations

$$\lambda^2 + (\delta + c)\lambda + \delta c - \eta e^{-\lambda\tau} = 0,$$

where δ , c and η are positive constants. We have $P_1(\lambda) = \lambda^2 + (\delta + c)\lambda + \delta c$, and $P_2(\lambda) = -\eta$. Thus

$$R_1(\sigma) = -\sigma^2 + \delta c,$$

$$Q_1(\sigma) = (\delta + c)\sigma,$$

$$R_2(\sigma) = -\eta, \text{ and}$$

$$Q_2(\sigma) = 0.$$

By the method of the lemma, we arrive at

$$\begin{aligned} \eta^2 &= (\sigma^2 - \delta c)^2 + (\delta + c)^2 \sigma^2, \\ (2.11) \quad \eta^2 &= \sigma^4 - 2\delta c \sigma^2 + \delta^2 c^2 + (\delta^2 + 2\delta c + c^2) \sigma^2, \\ 0 &= \sigma^4 + (\delta^2 + c^2) \sigma^2 + \delta^2 c^2 - \eta^2. \end{aligned}$$

Let $\mu = \sigma^2$, then this becomes:

$$S(\mu) \equiv \mu^2 + (\delta^2 + c^2)\mu + \delta^2 c^2 - \eta^2 = 0.$$

Since the linear coefficient of S is positive, by Descartes' rule of signs, a positive real root can occur if and only if the constant coefficient is negative. So a change of stability occurs if and only if $0 > \delta^2 c^2 - \eta^2 = (\delta c + \eta)(\delta c - \eta)$, i.e., if and only if $\delta c < \eta$.

Checking nondegeneracy, we take the derivative of the last line of (2.11), and check that equality does not hold.

$$0 = 4(\sigma^*)^3 + 2(\delta^2 + c^2)\sigma^*, \text{ and}$$

$$0 = 4(\sigma^*)^2 + 2(\delta^2 + c^2),$$

which clearly has no roots. This shows, that a nondegenerate bifurcation does occur for $\delta c < \eta$. This reproduces the results in Nelson et al [38].

Culshaw and Ruan, in [14] applied this same method to conclude that no bifurcations occurred in a delay model with characteristic equation

$$(2.12) \quad \lambda^3 + a_1\lambda^2 + a_2\lambda + a_3e^{-\lambda\tau} + a_4\lambda e^{-\lambda\tau} + a_5 = 0.$$

In their paper, Culshaw and Ruan follow the method we have presented in Lemma 2, and arrive at the polynomial S if equation (2.5) in the form

$$z^3 + \alpha z^2 + \beta z + \gamma$$

Proposition 2 in [14] states that if $\gamma \geq 0$ and $\beta > 0$, then this polynomial has no positive real roots. The proof of this proposition also assumes that $\alpha > 0$. In this case all of the coefficients are positive, and there are certainly no positive roots. The condition $\alpha, \beta, \gamma > 0$ is sufficient, but it is not necessary for no roots to exist. In the next section we develop a criterion which will extend this result and give necessary and sufficient conditions for a characteristic equation of the form (2.12) to produce no bifurcations.

2.4 General Order Two and Three Characteristic Equations

Using Sturm sequences, we can derive some general results for low order characteristic equations. We begin with the general degree two equation, for which a general result is easy

$$(2.13) \quad \lambda^2 + a\lambda + b + (c\lambda + d)e^{-\lambda\tau} = 0.$$

A steady state with this characteristic is stable for $\tau = 0$ if all of the roots of

$$\lambda^2 + (a + c)\lambda + (b + d) = 0$$

have negative real part. By the Routh-Hurwitz conditions, this occurs if and only if $a + c > 0$ and $b + d > 0$.

Letting $\lambda = i\sigma$ and proceeding as in Lemma 2, we arrive at the following form of equation (2.5)

$$(2.14) \quad \mu^2 + (a^2 - c^2 - 2b)\mu + (b^2 - d^2) = 0.$$

Let $A \equiv a^2 - c^2 - 2b$ and $B \equiv b^2 - d^2$. Equation (2.14) has a positive real root in two circumstances. Since the lead coefficient is positive, if $B < 0$ then there is a single positive real root. If $B > 0$, the roots of (2.14) are

$$\frac{-A \pm \sqrt{A^2 - 4B}}{2},$$

and there is a simple positive root (in fact two simple positive real roots) if and only if $A < 0$ and $A^2 - 4B > 0$. Thus we can conclude

Proposition 2.3. *A steady state with characteristic equation (2.13) is stable in the absence of delay, and becomes unstable with increasing delay if and only if*

i. $a + c > 0$ and $b + d > 0$, and

ii. either $b^2 < d^2$, or $b^2 > d^2$, $a^2 < c^2 + 2b$ and $(a^2 - c^2 - 2b)^2 > 4(b^2 - d^2)$.

For similar results in the degree two case, and also for some more general results, see Kuang [32].

For the degree three problem, the situation is somewhat more complex. The general characteristic equation is

$$(2.15) \quad \lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 + (b_2\lambda^2 + b_1\lambda + b_0)e^{-\lambda\tau} = 0.$$

The steady state is stable in the absence of delay if the roots of

$$\lambda^3 + (a_2 + b_2)\lambda^2 + (a_1 + b_1)\lambda + (a_0 + b_0) = 0$$

have negative real part. This occurs if and only if $a_2 + b_2 > 0$, $a_0 + b_0 > 0$ and $(a_2 + b_2)(a_1 + b_1) - (a_0 + b_0) > 0$.

In this case the form of equation (2.5) is

$$(2.16) \quad \mu^3 + A\mu^2 + B\mu + C = 0,$$

where

$$(2.17) \quad A \equiv a_2^2 - b_2^2 - 2a_1, B \equiv a_1^2 - b_1^2 + 2b_2b_0 - 2a_2a_0 \text{ and } C \equiv a_0^2 - b_0^2.$$

As in the degree two case, since the lead coefficient is positive, there are two manners in which a positive real root can occur. The first and simplest is to have $C < 0$. Now suppose that $C > 0$. Since the polynomial is odd, we are guaranteed a negative real root. The only way to have a simple positive real root in this case is to have 2 positive real roots. In other words, all of the roots are real. Now suppose we take the Sturm chain of the polynomial (2.16), denoted f_0, f_1, f_2, f_3 . We evaluate the entire real line, i.e., from $-\infty$ and ∞ , and construct a table of the signs at these endpoints. $f_0 = \mu^3 + A\mu^2 + B\mu + C$ and $f_1 = 3\mu^2 + 2A\mu + B$, so we have

	$-\infty$	∞
f_0	-	+
f_1	+	+
f_2		
f_3		

We know that there must be three real roots. The difference in the number of sign changes at each endpoint must be three, but this is only possible if the Sturm sequence at one endpoint is always positive or always negative, and the sequence at

the other endpoint must alternate. So the completed table must have the form

	$-\infty$	∞
f_0	-	+
f_1	+	+
f_2	-	+
f_3	+	+

Notice that f_0 and f_2 are odd degree polynomials, and f_1 and f_3 are even degree polynomials, and the signs at $-\infty$ are the direct consequence of those at ∞ (the same for even polynomials, and the opposite for odd polynomials). Thus, the bifurcation occurs in the case $C > 0$ if and only if the lead coefficients f_2 and f_3 are positive. Carrying out the division algorithm, the lead coefficient of f_2 is

$$-(\frac{2}{3}B - \frac{2}{9}A^2),$$

which is positive if and only if $A^2 - 3B > 0$.

f_3 is the constant

$$-\frac{9}{4} \frac{4B^3 - A^2B^2 - 18ABC + 4CA^3 + 27C^2}{(A^2 - 3B)^2}.$$

After some algebraic manipulation, we can see that this is positive if and only if

$$(2.18) \quad 4(B^2 - 3AC)(A^2 - 3B) - (9C - AB)^2 > 0.$$

Now we have conditions to guarantee that there are three real roots. We must finally guarantee that one of these is positive. This occurs if (2.16) has a positive critical point. The derivative function is

$$f_1 = 3\mu^2 + 2A\mu + B,$$

whose roots are $\frac{-A \pm \sqrt{A^2 - 3B}}{3}$. One of these is positive if $A < 0$ or $A > 0$ and $B < 0$, so either A or B must be negative. So we have

Theorem 2.4. *A steady state with characteristic equation (2.15) is stable in the absence of delay, and becomes unstable with increasing delay if and only if A, B , and C are not all positive and*

- i. $a_2 + b_2 > 0$, $a_0 + b_0 > 0$, $(a_2 + b_2)(a_1 + b_1) - (a_0 + b_0) > 0$, and*
- ii. either $C < 0$, or $C > 0$, $A^2 - 3B > 0$ and the condition (2.18) is satisfied, where A, B and C are given by (2.17).*

2.5 Conclusions

So we have developed a method of reducing the question of the existence of a delay-induced loss of stability to the problem of finding real positive roots of a polynomial. Although this method has been utilized before, it is useful to see the form of the polynomials involved. These results are summarized in Lemma 2.2.

The method of this lemma can be used to verify and to extend the results in several cases from the literature. More generally, it is easy, using the technique, to arrive at general conditions on the coefficients of a characteristic equation of degree 2, such that it describes an asymptotically stable steady state which becomes unstable as the delay parameter is increased. This simple, practical test is given in Proposition 2.3, and is related to analysis done by Y. Kuang in Chapter 3 of his book [32].

The main result of this chapter, presented in Theorem 2.4, is for the degree three case, where Sturm sequences are used to develop an elementary (though perhaps algebraically complicated) test for bifurcation. It is hoped that this criterion will make the investigation of third order systems of delay differential equations simpler,

both analytically and numerically. It provides a general algorithm for determining stability that anyone modeling with delay differential equation models can use.

CHAPTER 3

Single Species Models

In the study of population dynamics, the use of differential equations to study single species populations is well established. Exponential and logistic growth models are the most common. We would like to study a class of differential equation models for a single species that involve a time delay. The goal is to determine whether the introduction of time delays might enrich the dynamics of these models, or whether their behavior is essentially the same as the ordinary differential equations models they modify. In particular, we are interested in determining the existence of periodic solutions for these models.

In this chapter, I will begin by stating the theorems from functional analysis which we will use to prove the existence of periodic solutions to the delay differential equations I will study. This section is followed by the exploration of a model of the form

$$(3.1) \quad \dot{x}(t) = b(x(t - \tau))x(t - \tau) - d(x(t))x(t),$$

with b nonincreasing and d nondecreasing, which represents the population dynamics of a single species with a delayed birth term. Basic properties of this model are determined, including the types of functions b and d which might lead to the existence of periodic solutions.

In the Section 3.3, we specify to the case $b(x) = be^{-ax}$ and $d(x)$ constant. In this case, I prove the existence of a class of solutions oscillating about the nontrivial steady state, and then go on to extend a result of Kuang [32], proving the existence of a periodic solution to this model in a wider parameter set than has previously been shown.

In Section 3.4, the final one dealing with this model, a delay-dependent term is added to the parameter b . The effects of this alteration are explored, and conditions are given for the existence and linear instability of the positive steady state.

Following this, I change the model to make the rate of change proportional to the current state of the variable, so the model takes the form

$$(3.2) \quad \dot{x}(t) = [b(x(t - \tau)) - d(x(t))] x(t).$$

The same general plan is followed as with the first model. I begin by exploring the basic properties of the model, and the forms of $b(x)$ and $d(x)$ which might give rise to periodic solutions.

In Section 3.6, the case of a constant per capita death rate is explored in detail, and it is shown that whenever the nontrivial steady state exists and is unstable, a periodic solutions exists. Finally, we introduce a delay dependence in the parameters of (3.2), and in the case $b(x) = be^{-ax}$, I derive the exact range of delays τ for which a positive periodic solution exists.

3.1 A Fixed-Point Theorem from Nonlinear Functional Analysis

The primary tool available for proving the existence of periodic solutions is the theorem below from nonlinear functional analysis. Before stating the theorem, we need to define what it means for a fixed point of a map to be ejective.

Definition 3.1. Let X be a Banach space, K a subset of X , and $x_0 \in K$. The point x_0 is said to be an *ejective point* of a map $A : X \setminus \{x_0\} \rightarrow X$ if there is an open neighborhood $G \subset X$ of x_0 such that, if $y \in G \cap K$, $y \neq x_0$, there is an integer $m = m(y) > 0$ such that $A^{(m)}(y) \notin G \cap K$.

Intuitively, a point is ejective if it is surrounded by a neighborhood of points, which the map will send outside the neighborhood eventually. We now state the theorem we apply in this chapter and Chapter 4.

Theorem 3.2. *If K is a closed, bounded, convex and infinite dimensional set in a Banach space X , and $A : K \setminus \{x_0\} \rightarrow K$ is completely continuous, and $x_0 \in K$ is ejective, then there is a fixed point of A in $K \setminus \{x_0\}$.*

A proof of this theorem is provided by Nussbaum [42]. The primary challenge in applying this result consists of constructing an appropriate map A . We will show that solutions of the system oscillate about the nontrivial steady state, and the “return map” acts on the space of initial functions. A fixed point of this return map corresponds to a periodic solution, since dictating the behavior of a solution on an interval of length τ determines all future behavior. Just as with ordinary differential equations, if an autonomous system returns to its initial condition (or initial function), it is periodic. This method is analogous to examining a Poincare map for an ordinary differential equation.

3.2 A General Single-Species Population Model with Delay

The first class of models we will examine will be of the form

$$(3.3) \quad \dot{x}(t) = b(x(t - \tau))x(t - \tau) - d(x(t))x(t).$$

We will consider that $b(x)$ is a continuous, positive, decreasing function, *i.e.*, that the per capita growth rate of the population decreases with increased population levels. This is an instance of density-limited growth, of which the logistic model is another example. The delay in this instance can represent a gestation or maturation period, so the number of individuals entering the population depends on the levels of the population at a previous instance of time.

The function $d(x)$ is nondecreasing and positive. This represents the per capita death rate, which may be increased by intraspecific competition.

Models of this type have been used extensively in the mathematical biology literature, especially when there is an interest in modelling oscillatory phenomena. In population biology, for example, [4] and [55] explore the model generally, while [48] is a specific application to housefly populations. Such models are also used in other branches of biology, such as physiology [36]. While oscillatory phenomena are noted, few analytic results about the existence of periodic solutions exist for such models. One such result is found in [32], Chapter 5, and I will refer to it often. More commonly, results proving the existence of positive periodic solutions rely on a non-autonomous periodic forcing term or periodic coefficients, with period greater than zero ([21], [22], [54]).

Now let us proceed with the analysis by proving the following basic fact.

Lemma 3.3. *Given positive initial data, solutions of equation (3.3), where b is a positive function, remain positive for all time.*

Proof. We can simply look at the rate of change by steps. By assumption, $x(t)$ is positive for $t \in [-\tau, 0]$, so for $t \in [0, \tau]$, it is easy to see that $\dot{x}(t) > -d(x(t))x(t)$. So if $T \in [0, \tau]$ is the first time at which $x(t) = 0$, then $\dot{x}(T) > 0$. This is clearly a

contradiction, so $x(t) > 0$ in this interval. Now simply apply the same analysis to $[\tau, 2\tau]$, and so on. So for all t , the solution remains positive. \square

The requirement that $b(x)$ be positive is necessary, in spite of the analogy to, for example, the logistic ordinary differential equation. If there is an \tilde{x} such that $b(\tilde{x}) < 0$, then there are positive initial histories which become negative. One could simply set the initial history to be \tilde{x} on $[-\tau, -\epsilon]$ for some small $\epsilon > 0$, and make it continuous on $[-\epsilon, 0]$ so that $x(0)$ is sufficiently small, say $x(0) = -b(\tilde{x})(\tau - \epsilon)\tilde{x}/2 > 0$. One sees that the solution will be driven negative in the interval $[0, \tau]$. If $x(t) \geq 0$ on $[0, \tau - \epsilon]$, then

$$\begin{aligned} x(\tau - \epsilon) &\leq x(0) + \int_0^{\tau - \epsilon} b(x(s - \tau))x(s - \tau)ds \\ &= x(0) + \int_0^{\tau - \epsilon} b(\tilde{x})\tilde{x}ds \\ &= -\frac{b(\tilde{x})}{2}(\tau - \epsilon)\tilde{x} + b(\tilde{x})(\tau - \epsilon)\tilde{x} \\ &= \frac{b(\tilde{x})}{2}(\tau - \epsilon)\tilde{x} < 0, \end{aligned}$$

contradicting the positivity of $x(t)$ on $[0, \tau - \epsilon]$.

I will now give three theorems which describe the most general division of possible behavior regimes for the differential equation (3.3). These results are slightly more general than the requirement that b be decreasing and d be increasing. Also, it is likely that these simple results have already been obtained elsewhere, but I have not seen them recorded. It is useful to see that the case I will consider in detail, that which will be covered by Theorem 3.4, is the only one with interesting long-term dynamics.

Theorem 3.4. *Consider the delay differential equation (3.3), if b is a positive function and $\sup b(x) < \inf d(x)$, then the zero steady state is globally asymptotically*

stable.

Proof. Let $B = \sup b(x)$ and $D = \inf d(x)$. We have, then, that $\dot{x}(t) < Bx(t - \tau) - Dx(t)$, but solutions of $\dot{y}(t) = By(t - \tau) - Dy(t)$ all approach 0 asymptotically as $t \rightarrow \infty$, according to Lemma 1.4, since $0 < B < D$. So all solutions of (3.3) approach 0 also. \square

Theorem 3.5. *Let b and d be positive functions. Suppose that there exists an \bar{x} such that $\text{sign}(b(x) - d(x)) = -\text{sign}(x - \bar{x})$, and $b'(\bar{x}) < d'(\bar{x})$. Then \bar{x} is a positive steady state, and the trivial steady state is unstable. If*

$$(3.4) \quad b'(\bar{x})\bar{x} > -2d(\bar{x}) - d'(\bar{x})\bar{x},$$

then \bar{x} is linearly stable for all τ . Otherwise, there exists a $\tau_c > 0$ such that \bar{x} is stable for $\tau < \tau_c$, and unstable for $\tau > \tau_c$.

Proof. To begin with, \bar{x} is a unique positive steady state, since $b(x) - d(x) = 0$ if and only if $x = \bar{x}$. It is the point at which $b(\bar{x}) = d(\bar{x})$. Linearizing about this steady state yields the equation

$$(3.5) \quad \dot{x}(t) = (d(\bar{x}) + b'(\bar{x})\bar{x})x(t - \tau) - (d(\bar{x}) + d'(\bar{x})\bar{x})x(t),$$

which has characteristic equation

$$\lambda = \alpha x(t - \tau) - \beta x(t),$$

where $\alpha = d(\bar{x}) + b'(\bar{x})\bar{x}$ and $\beta = d(\bar{x}) + d'(\bar{x})\bar{x}$. Since $b'(\bar{x}) < d'(\bar{x})$, $\alpha < \beta$. Furthermore, we know that for $|\alpha| < |\beta| = \beta$, all roots of the characteristic equation have negative real part. Since $\alpha < \beta$, this condition is satisfied if and only if $\alpha > -\beta$, but this is exactly the condition (3.4).

If this is not the case, then $\alpha < -\beta$. It is clear that for $\tau = 0$, the only characteristic root is $\lambda = \alpha - \beta < 0$. Thus, by the continuity of the location of roots, for small delays, the system is stable. The derived polynomial for the characteristic equation is $\sigma - \alpha^2 - \beta^2$, which clearly has a positive real root. Thus there is a τ_c for which the characteristic equation has a purely imaginary root. As τ increases past τ_c , a root enters the right half-plane. Since the derived polynomial has degree 1, our Sturm sequence analysis shows that this root can never exit. Thus for $\tau > \tau_c$, the steady state is unstable. \square

In [12] the authors prove that if $(b(x)x)' > 0$ for all x , then the steady state is asymptotically stable. A more general result about the linear stability of the model are also obtained in [12]. These results are contained in Theorem 3.5.

The only situation not covered by the theorems above is when $b(x) > d(x)$ for all x . In this case, there is no positive steady state, but the trivial steady state is unstable. This situation is covered by the following theorem.

Theorem 3.6. *If*

$$(3.6) \quad \lim_{x \rightarrow \infty} b(x) \geq \lim_{x \rightarrow \infty} d(x),$$

then all solutions of (3.3) with positive initial data are unbounded.

In particular, no positive periodic solutions are possible in this case. We will prove this theorem via a pair of lemmas.

Lemma 3.7. *Given the condition (3.6), a solution, $x(t)$, of equation (3.3) with positive initial data is bounded if and only if $\lim_{t \rightarrow \infty} x(t) = 0$.*

Proof. Since solutions are continuous, it is clear that if $x(t) \rightarrow 0$, then it is bounded. Now suppose that $x(t) < M$ for all t . In this case, define $N = b(M) - d(M)$. Since

b is decreasing and d is increasing, we have that $b(x(t)) - d(x(t)) \geq N$ for all t .

Integrating the differential equation (3.3) yields

$$\begin{aligned} x(t) &= x(0) + \int_0^t [b(x(s-\tau))x(s-\tau) - d(x(s))x(s)]ds \\ &= x(0) + \int_{-\tau}^0 b(x(s))x(s)ds + \int_0^{t-\tau} (b(x(s)) - d(x(s)))x(s)ds - \int_{t-\tau}^t d(x(s))x(s)ds. \end{aligned}$$

Define $A = x(0) + \int_{-\tau}^0 b(x(s))x(s)ds$, which is a constant determined by the initial history of x . Continuing from above, we can find a lower bound on $x(t)$ in the following manner

$$\begin{aligned} x(t) &= A + \int_0^{t-\tau} (b(x(s)) - d(x(s)))x(s)ds - \int_{t-\tau}^t d(x(s))x(s)ds \\ &\geq A + \int_0^{t-\tau} Nx(s)ds - \int_{t-\tau}^t d(M)Mds \\ (3.7) \quad &= A - d(M)M\tau + \int_0^{t-\tau} Nx(s)ds. \end{aligned}$$

Since $x(t) < M$, the lower bound given by (3.7) must be bounded for all t . In particular, the integral

$$\int_0^\infty Nx(s)ds$$

must be finite, which implies that $x(t) \rightarrow 0$ as $t \rightarrow \infty$, since $x(t)$ is always positive. □

Lemma 3.8. *The delay differential equation (3.3), under the conditions of Theorem 3.5 has no solutions which approach 0 as $t \rightarrow \infty$.*

Proof. Given an initial history, we again begin with

$$x(t) = x(0) + \int_{-\tau}^0 b(x(s))x(s)ds + \int_0^{t-\tau} (b(x(s)) - d(x(s)))x(s)ds - \int_{t-\tau}^t d(x(s))x(s)ds.$$

Notice that the first three terms of this expression are positive, and the final term is the only negative term. Define $B = \int_{-\tau}^0 b(x(s))x(s)ds$. If $x(t) \rightarrow 0$, then there exists

a $T > 0$ such that, for all $t > T$, $x(t) < \frac{B}{2d(M)\tau}$. Where M is an upper bound on $x(t)$.

Now for $t > T$,

$$\int_{t-\tau}^t d(x(s))x(s)ds \leq \int_{t-\tau}^t d(M)\frac{B}{2d(M)\tau}ds = \frac{B}{2}.$$

Thus, for $t > T$, $x(t) > \frac{B}{2}$, a contradiction. \square

Given Lemmas 3.7 and 3.8, it is now obvious that solutions with positive initial data must be unbounded, and thus Theorem 3.6 is proven.

3.3 A Specific Single-Species Delay Model

We will now look specifically at

$$(3.8) \quad \dot{x}(t) = bx(t - \tau)e^{-ax(t-\tau)} - dx(t),$$

which is a particular case of equation (3.3). We will assume that $b > d$, so that we are in the case of Theorem 3.5, where the nontrivial steady state exists.

This particular form of the more general model, with constant per capita death rate and exponentially decaying per capita birth rate has been used in many models, for example [4] and [24], especially those dealing with Nicholson's famous blowfly data ([40], [41]), which sparked much debate about the possibility of chaotic dynamics in natural populations.

Let us begin by looking at the particulars of this case. The nontrivial steady state occurs when $be^{-a\bar{x}} = d$, *i.e.*, $\bar{x} = \frac{1}{a} \ln \frac{b}{d}$. According to Theorem 3.5 \bar{x} is stable for all τ if and only if

$$\left. \frac{d}{dx} be^{-ax} \right|_{x=\bar{x}} > -2\frac{d}{\bar{x}}.$$

This is equivalent to the condition $b < de^2$.

Now suppose that $b > de^2$, and let $\alpha = \ln \frac{b}{d} - 1$. Then $\alpha > 1$ and the characteristic equation is

$$\lambda = -d\alpha e^{-\lambda\tau} - d.$$

When $\tau = 0$, this is $\lambda = -d\alpha - d$. Suppose $\tau > 0$ and that $\lambda = i\sigma$, $\sigma > 0$ is a purely imaginary root. Then the real and imaginary parts of the characteristic equation are

$$d = -d\alpha \cos(\sigma\tau),$$

$$\sigma = d\alpha \sin \sigma\tau.$$

Squaring these and summing, we get $\sigma^2 + d^2 = d^2\alpha^2$, i.e. $\sigma = d(\alpha^2 - 1)^{\frac{1}{2}}$.

Rewriting the real and imaginary parts of the characteristic equation, we see,

$$\begin{aligned} \cos \sigma\tau &= -\frac{1}{\alpha} < 0, \\ \sin \sigma\tau &= \frac{(\alpha^2 - 1)^{\frac{1}{2}}}{\alpha} > 0. \end{aligned}$$

So for τ_c , the critical delay at which an eigenvalue crosses into the right half-plane, $\sigma\tau_c \in (\frac{\pi}{2}, \pi)$, and the critical delay is

$$(3.9) \quad \tau_c = \frac{1}{d(\alpha^2 - 1)^{\frac{1}{2}}} \cos^{-1} \left(-\frac{1}{\alpha} \right).$$

For $\tau > \tau_c$ the steady state is unstable. From now on, we will assume that $b > de^2$.

3.3.1 Oscillatory Solutions

Now let us take an initial function in the set

$$K = \{\phi \in \mathcal{C}([-\tau, 0], \mathbb{R}^+) : \phi(-\tau) = \bar{x}, \phi(t) > \bar{x}, \forall t \in (-\tau, 0]\}.$$

So long as $x(t) > \bar{x}$, a solution to (3.8) with an initial history in K will be decreasing, since the entire graph of bxe^{-ax} lies below that of dx when $x > \bar{x}$ (see Figure 3.1).

Let us also define the value $x_m < \bar{x}$ so that $x_mb(x_m) = d\bar{x}$. In the region (x_m, \bar{x}) , the

entire graph of $xb(x)$ lies above dx , and if a solution remains in this region, then it must be increasing. We now show that any solution with initial history in $K \setminus \{\bar{x}\}$ must oscillate about \bar{x} infinitely often.

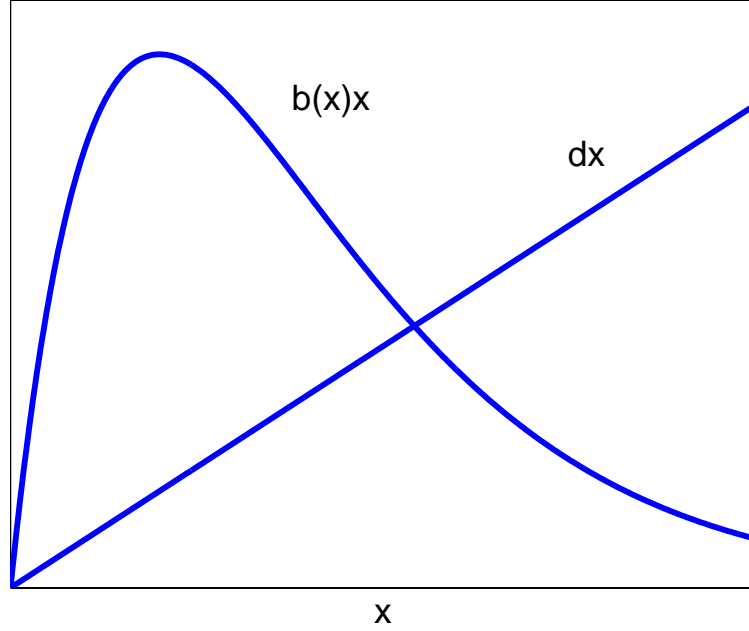


Figure 3.1: The growth function, $b(x)x$, and the decay function, dx , intersecting at \bar{x}

Lemma 3.9. *If $\phi \in K$, then there exist times $0 < t_1 < t_2$ such that if $x(t)$ is a solution to (3.8) with initial function ϕ , then $x(t_1) = x(t_2) = \bar{x}$, $\dot{x}(t_1) < 0$ and $\dot{x}(t_2) > 0$ and $x(t) \neq \bar{x}$ for any other $t \in (0, t_2)$*

Proof. Suppose that $x(t) > \bar{x}$ for all t , then x is monotone decreasing and bounded below. Thus, $x(t)$ has a limit, and since \dot{x} must approach 0 as x approaches this limit, it is clear from the differential equation that $x(t) \rightarrow \bar{x}$.

In order to prove that solutions with initial data in the class K cannot remain above \bar{x} and have \bar{x} as a limit, we must now look more carefully at the critical delay length τ_c . We know that the nontrivial steady state is unstable if and only if $\tau > \tau_c$,

and we have seen that $\sigma\tau_c \in (\frac{\pi}{2}, \pi)$. From the imaginary part of the characteristic equation when $\tau = \tau_c$, recall that $\sigma = d\alpha \sin(\sigma\tau)$. We get the following chain of inequalities, given that the nontrivial steady state is unstable

$$\begin{aligned}\sigma\tau &> \sigma\tau_c > \frac{\pi}{2} \\ \tau &> \frac{\pi}{2} \frac{1}{d\alpha \sin(\sigma\tau)} \\ &> \frac{\pi}{2} \frac{1}{d\alpha} > \frac{1}{d\alpha}.\end{aligned}$$

The form of this inequality we will use is

$$-d\alpha < -\frac{1}{\tau}.$$

Now consider the function $B(x) = xb(x)$. Taking the derivative at the point $x = \bar{x}$, we get $B'(\bar{x}) = -d\alpha < 0$. Note, in particular, that B is decreasing in a neighborhood of \bar{x} . For any slope $s \in (B'(\bar{x}), 0)$, there exists a $\delta > 0$ such that for $0 < x - \bar{x} \leq \delta$, $B(x) - B(\bar{x}) < s(x - \bar{x})$. In particular, we now take $s = -\frac{1}{\tau}$.

Let $T > \tau$ be a time such that $x(T) = \bar{x} + \delta$. Then for $t \in [T, T + \tau]$ we have

$$\begin{aligned}\dot{x}(t) &= B(x(t - \tau)) - d(x(t)) \\ &< B(x(t - \tau)) - d\bar{x} \\ &< B(x(T)) - B(\bar{x})\end{aligned}$$

since $x(t)$ is decreasing for $t > 0$ and B is decreasing in a neighborhood of \bar{x} . Also, $B(\bar{x}) = d\bar{x}$. Continuing,

$$\dot{x}(t) < -\frac{1}{\tau}(x(T) - \bar{x}) = -\frac{\delta}{\tau}$$

But if $\dot{x}(t) < -\frac{\delta}{\tau}$ on the interval $[T, T + \tau]$, then $x(T + \tau) < x(T) - \tau\frac{\delta}{\tau} = \bar{x}$, contradicting the assumption that $x(t)$ remains above \bar{x} . We are lead to the conclusion

that there exists a time t_1 such that $x(t_1) = \bar{x}$, $x(t) > \bar{x}$ for $t \in (0, t_1)$, and $\dot{x}(t_1) < 0$, as desired.

For $t \in (t_1, t_1 + \tau)$, $x(t) \leq \bar{x}$. To see this, suppose that $x(t) = \bar{x}$, then $\dot{x}(t) = x(t - \tau)b(x(t - \tau)) - d\bar{x} \geq 0$. This implies, $x(t - \tau)b(x(t - \tau)) \geq d\bar{x}$, but this is not possible, since at time $t - \tau$, $xb(x)$ is less than $d\bar{x}$, as is apparent in the figure 3.1. Now suppose that $x(t) < \bar{x}$ for all $t > t_1$. Integrating (3.8), one arrives at

$$(3.10) \quad x(t) - \bar{x} = \int_{t_1 - \tau}^{t_1} f(x(s))x(s)ds + \int_{t_1}^{t - \tau} (f(x(s)) - d)x(s)ds - \int_{t - \tau}^t dx(s)ds$$

$$(3.11) \quad \geq \int_{t_1}^{t - \tau} (f(x(s)) - d)x(s)ds + A - d\tau\bar{x},$$

where A is defined to be $\int_{t_1 - \tau}^{t_1} f(x(s))x(s)ds$, and is fixed by the value of the solution before entering the region $x < \bar{x}$. If the integral $\int_{t_1}^{t - \tau} (f(x(s)) - d)x(s)ds$ fails to converge, then $x(t) \rightarrow \infty$, since the integrand is positive. As this contradicts the assumption that $x(t) < \bar{x}$, we must assume that the integral converges. In particular, the integrand must approach zero. This can occur if and only if x approaches 0 or \bar{x} . We can rule out the case of $x(t) \rightarrow 0$ using equation (3.10). As $x \rightarrow 0$, the final term on the right hand side becomes arbitrarily small, and thus $x(t) - \bar{x} > 0$. Which contradicts the assumption that $x \rightarrow 0$.

We conclude that if $x(t) < \bar{x}$ then $x(t) \rightarrow \bar{x}$. If this is the case, then there exists a time T so that for $x(t) > x_m$ for all $t > T$, and for these times $x(t)$ is increasing.

The proof that a time t_2 exists such that the solution $x(t)$ must increase across the level \bar{x} at time t_2 is analogous to the proof of the existence of t_1 , above, and is omitted. □

We are easily led to the following, much more general, result.

Corollary 3.10. *Any solution of the delay differential equation (3.8) with positive initial data is equal to \bar{x} infinitely often.*

Proof. If we assume that the solution $x(t)$ satisfies $x(t) > \bar{x}$ for all $t > T$, then the analysis in the proof of the previous theorem derives a contradiction. Similarly, if $x(t) < \bar{x}$ for $t > T$, the previous proof arrives at a contradiction. \square

3.3.2 An Extension of Previously Known Results

In [32], the author proves the existence of periodic solutions for certain equations of the form

$$\dot{x}(t) = B(x(t - \tau)) - D(x(t)).$$

An essential component of this proof, required to guarantee certain properties of the solution map, was the existence of a value $\underline{x} \in (x_M, \bar{x})$ such that $B(D^{-1}(B(\underline{x}))) > D(\underline{x})$. In this section, I provide a broader condition, which not only encompasses a larger set in the space of parameters, but is also directly verifiable without the need to find \underline{x} . The proof of the existence of periodic solutions from [32] will again apply to this broader case, extending the previous results.

Let $B(x) = xb(x)$, $D(x) = xd(x)$, and let x_M be the point at which B achieves its maximum. Also define $x_m \in (0, x_M)$ such that $B(x_m) = B(\bar{x})$. If

$$(3.12) \quad D^{-1}(B(D^{-1}(B(x_M)))) > x_m,$$

then the solution operator maps K into K .

Suppose that the initial function $\phi \in K$. Then so long as $x(t)$ remains above \bar{x} , the solution $x(t)$ is decreasing. As we have seen, the form of the equation dictates that the solution must cross \bar{x} at some point t_1 . For the next τ time units, the value of $B(x(t - \tau))$ increases, since $x(t - \tau)$ decreases, and B is decreasing for $x > \bar{x}$.

Claim: $x(t) \neq \bar{x}$ for $t \in (t_1, t_1 + \tau)$.

Proof. If $x(\tilde{t}) = \bar{x}$ for some $\tilde{t} \in (t_1, t_1 + \tau)$, and that \tilde{t} is the smallest such time. Then

$D(x(\tilde{t})) = D(\bar{x}) = B(\bar{x}) > B(x(\tilde{t} - \tau))$, and thus $\dot{x}(\tilde{t}) < 0$, contradicting the fact that $x(t) < \bar{x}$ for $t \in (t_1, \tilde{t})$. \square

So for the interval $(t_1, t_1 + \tau)$, the solution x is below \bar{x} . We now show for these times x is above x_m . Let us deal with this in two cases: x achieves its minimum at $t_1 + \tau$, and it achieves its minimum at some time in $(t_1, t_1 + \tau)$. The first case is impossible, since $\dot{x}(t_1 + \tau) = B(\bar{x}) - d(x(t)) > B(\bar{x}) - D(\bar{x}) = 0$. So the minimum must occur in the interval $(t_1, t_1 + \tau)$. At the minimum,

$$\begin{aligned} 0 &= \dot{x}(t) = B(x(t - \tau)) - D(x(t)) \\ D(x(t)) &= B(x(t - \tau)) \geq B(D^{-1}(B(x_M))) \\ x(t) &\geq D^{-1}(B(D^{-1}(B(x_M)))) > x_m. \end{aligned}$$

Thus, in the interval $(t_1, t_1 + \tau)$, the solution $x(t)$ remains in the region (x_m, \bar{x}) . In this region, $B(y) > D(x)$ for all x and y . It follows that x is increasing for $t \geq t_1 + \tau$ for as long as it remains below \bar{x} . By the same argument as before, the solution must cross \bar{x} at some time $t_2 > t_1 + \tau$. Arguing analogously to the above, since x stays above x_m in the interval $(t_2 - \tau, t_2)$, the maximum of x on the interval $(t_2, t_2 + \tau)$ is less than $F(x_M)$.

Thus, K is mapped into K by the solution operator. Now the arguments from Kuang apply to show that periodic solutions exist whenever the steady state is linearly unstable.

For what parameter regimes does the condition (3.12) hold? To begin with, recall that in our case $B(x) = bxe^{-ax}$ and $D(x) = dx$. For our functions B and D , the value of x_M can be determined by simply checking where $B'(x) = 0$. One finds that $x_M = \frac{1}{a}$. It is much more difficult to determine the value of x_m . Rather, we can find another condition, equivalent to (3.12), which does not require knowledge of the

actual value of x_m . One has

$$(3.13) \quad B(D^{-1}(B(D^{-1}(B(x_M)))))) > B(x_m),$$

since $D^{-1}(B(D^{-1}(B(x_M)))) \in (0, \bar{x})$, and in this region, $x > x_m$ is equivalent to $B(x) > B(\bar{x}) = d\bar{x}$. To apply this condition, one only needs knowledge of $B(x_m) = B(\bar{x}) = \frac{d}{a} \ln(\frac{b}{d})$.

Now, insert $x_M = \frac{1}{a}$ into (3.13).

$$\begin{aligned} b\left(\frac{1}{a}\right) &= \frac{b}{a}e^{-1} \\ D^{-1}\left(B\left(\frac{1}{a}\right)\right) &= \frac{b}{ad}e^{-1} \\ B\left(D^{-1}\left(B\left(\frac{1}{a}\right)\right)\right) &= \frac{b^2}{ad}e^{-1}e^{-\frac{b}{d}e^{-1}} \\ D^{-1}\left(B\left(D^{-1}\left(B\left(\frac{1}{a}\right)\right)\right)\right) &= \frac{b^2}{ad^2}e^{-1}e^{-\frac{b}{d}e^{-1}} \\ B\left(D^{-1}\left(B\left(D^{-1}\left(B\left(\frac{1}{a}\right)\right)\right)\right)\right) &= \frac{b^3}{ad^2}e^{-1}e^{-\frac{b}{d}e^{-1}}e^{-\frac{b^2}{d^2}e^{-1}e^{-\frac{b}{d}e^{-1}}} \end{aligned}$$

For the condition to hold, we need the expression above to be greater than $B(\bar{x}) = \frac{d}{a} \ln(\frac{b}{d})$. It is clear then that the only truly independent parameter is $\frac{b}{d}$. In fact, by rescaling the differential equation, we can assume that the parameter d is equal to 1. We have then

$$\begin{aligned} \frac{b^3}{a}e^{-1}e^{-be^{-1}}e^{-b^2e^{-1}e^{-be^{-1}}} &> \frac{1}{a} \ln(b) \\ b^3e^{-1}e^{-be^{-1}}e^{-b^2e^{-1}e^{-be^{-1}}} &> \ln(b) \end{aligned}$$

This condition is by no means easy on the eye. We can plot the difference of the left and right hand sides (see Figure 3.2), and see when the function is positive, in order to get an idea of the range of the parameter b for which the condition is satisfied. Recall that we are only interested in $b > e^2$, which is approximately 7.3891.

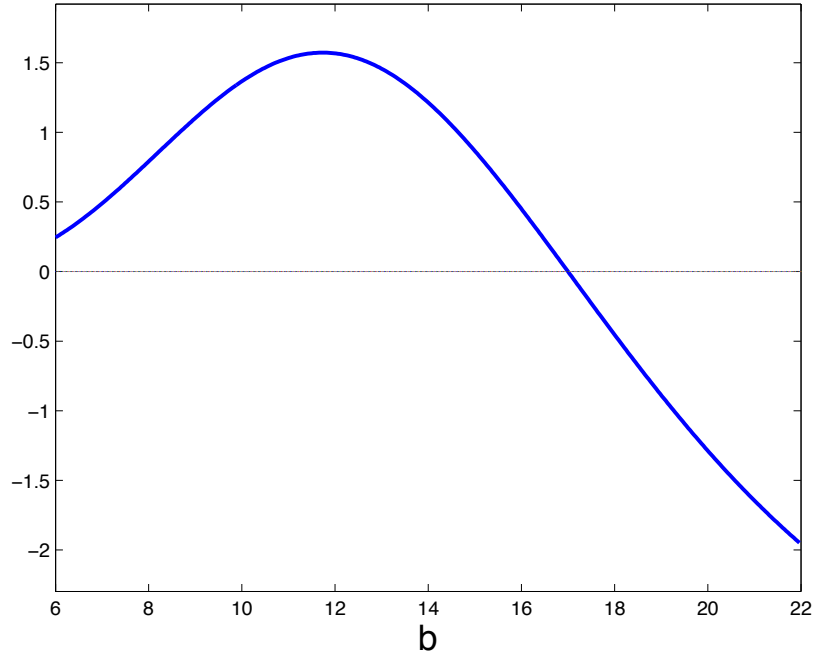


Figure 3.2: The graph of $b^3 e^{-1} e^{-be^{-1}} e^{-b^2 e^{-1} e^{-be^{-1}}} - \ln(b)$ against b . When $b > e^2$ and this function is positive, we can prove the existence of periodic solutions to the delay differential equation (3.8)

3.4 Delay Dependent Parameters

Staying with the same model as in the previous section, let us examine the effect of allowing one of the parameters to depend on the length of the delay τ . Specifically, consider

$$(3.14) \quad \dot{x}(t) = be^{-\mu\tau}x(t-\tau)e^{-ax(t-\tau)} - dx(t).$$

Since the first term in this equation represents recruitment or birth rate, the modification of this parameter could represent the decreased survivorship over a longer incubation or maturation time. I will examine the effect of this delay dependence on the existence and stability of the nontrivial steady state.

The mathematical difficulty imposed by this alteration is twofold. First of all, the location of the steady state will now vary with the length of the delay. Secondly,

the form of the characteristic equation will change due to the direct inclusion of the delay in the parameters, and the indirect changes resulting from the varying location of the steady state.

Let us begin by locating the steady states of the model (3.14). The zero steady state still exists, and a nontrivial steady state is given by

$$be^{-\mu\tau}e^{-a\bar{x}} = d$$

which leads to

$$\bar{x} = \frac{1}{a} \ln \frac{b}{de^{\mu\tau}}$$

In particular, if $\tau > \frac{1}{\mu} \ln \frac{b}{d}$, there is no positive steady state. In this case, given positive initial data, we have

$$\dot{x}(t) \leq be^{-\mu\tau}y(t - \tau) - dy(t),$$

with $be^{-\mu\tau} < d$, so the solution goes to 0, and the trivial steady state is globally stable.

Now we examine the characteristic equation for the positive steady state, given a particular delay $\tau < \frac{1}{\mu} \ln \frac{b}{d}$. We linearize the equation (3.14) as usual, and assume an exponential solution to get the new characteristic equation

$$(3.15) \quad \lambda = -d\alpha(\tau)e^{-\lambda\tau} - d,$$

where $\alpha(\tau) = 1 - \ln \frac{b}{de^{\mu\tau}}$.

This characteristic equation is essentially the same as that for the delay-independent case; only $\alpha(\tau)$ is affected. In the case of delay-independent parameters, we found a critical time delay τ_c , given in equation (3.9), such that the characteristic equation

$$\lambda = -d\alpha e^{-\lambda\tau} - d$$

has a root with positive real part if and only if $\tau > \tau_c$. We will now use this result to get the condition for instability of (3.14).

Theorem 3.11. *The nontrivial steady state of the delay differential equation (3.14) is unstable if and only if*

$$(3.16) \quad \tau > \frac{1}{d(\alpha(\tau)^2 - 1)^{\frac{1}{2}}} \cos^{-1} \left(-\frac{1}{\alpha(\tau)} \right).$$

Notice in particular that this condition includes the requirement that $\alpha(\tau)^2 - 1 > 0$, which is equivalent to $\ln \frac{b}{de^{\mu\tau}} (\ln \frac{b}{de^{\mu\tau}} - 2) > 0$. This is equivalent to the condition that $be^{-\mu\tau} > de^2$, similar to the condition $b > de^2$, which needed to be satisfied in order for a change of stability to occur in the delay-independent case. So we have

Theorem 3.12. *The positive steady state of (3.14) exists and is unstable if and only if $\tau < \frac{1}{\mu} \ln \frac{b}{d}$, and inequality (3.16) is satisfied. In this case, all solutions with positive initial data oscillate about the steady state.*

3.5 Another General Model

Now let us turn our attention to a slightly different model formulation.

$$(3.17) \quad \dot{x}(t) = (b(x(t - \tau)) - d(x(t)))x(t),$$

where b and d are again decreasing and increasing, respectively. As opposed to the model in equation 3.3, in this model, only the nonlinear components of the birth term are delayed. This could be thought of to correspond to a delayed density dependence in the per capita birth rate. The delayed logistic models is a particular example of (3.17). Dynamics of this form often form part of predator-prey and food chain models, for example [37].

The conditions for the existence of a positive steady state are the same as before, but the linearizations are different. As before, we have the following two results, which are included for completeness, in spite of their simplicity.

Theorem 3.13. *If $b(0) < d(0)$, then the delay differential equation (3.17) has no positive steady state, and the trivial steady state is globally asymptotically stable.*

Proof. It is clear that $\dot{x}(t) \leq (b(0) - d(0))x(t)$, and so solutions to the full delay differential equation are bounded by $x(0)e^{(b(0)-d(0))t}$, which approaches 0 as $t \rightarrow \infty$. □

Theorem 3.14. *If*

$$\lim_{x \rightarrow \infty} b(x) > \lim_{x \rightarrow \infty} d(x),$$

in equation (3.17), then any solution with positive initial history approaches ∞ as $t \rightarrow \infty$

Proof. It is clear in this case that the graph of $\max_{x \geq 0} d(x) < \min_{x \geq 0} b(x)$, so $\dot{x}(t)$ is positive for all t . If such an increasing solution is bounded, then it has a limit $L > 0$, but this would imply $0 = \lim_{t \rightarrow \infty} \dot{x}(t) = (b(L) - d(L))L$, which is clearly impossible. □

The most interesting case of this model is, however, when the graphs of b and d intersect, so that there is a nontrivial steady state. In contrast to the model (3.3), in this case the nontrivial steady state does not always change stability. Let $x(t) \equiv \bar{x}$ be the unique positive steady state of this delay differential equation, i.e. $b(\bar{x}) = d(\bar{x})$. Then the linearization of the equation about this steady state is

$$(3.18) \quad \dot{x}(t) = b'(\bar{x})\bar{x}x(t - \tau) - d'(\bar{x})\bar{x}x(t),$$

and the characteristic equation is

$$(3.19) \quad \lambda = -ae^{-\lambda\tau} - b,$$

where we define

$$a = -b'(\bar{x})\bar{x} > 0, \text{ and}$$

$$b = d'(\bar{x})\bar{x} > 0.$$

When the delay τ is sufficiently small, this characteristic equation has only roots with negative real part, and the steady state is stable. For some parameter regimes, however, longer delays result in an unstable steady state. These results are summarized in the following theorem.

Theorem 3.15. *If $d'(\bar{x}) > -b'(\bar{x})$, then the nontrivial steady state \bar{x} is linearly stable for all τ . For $d'(\bar{x}) < -b'(\bar{x})$, there exists a τ_c such that for $\tau < \tau_c$, the steady state is stable, and for $\tau > \tau_c$, it is unstable.*

Proof. We have the characteristic equation (3.19). Write $\lambda = \mu + i\sigma$, and we can separate this equation into its real and imaginary parts, yielding

$$(3.20) \quad \mu + b = -ae^{-\mu\tau} \cos(\sigma\tau)$$

$$(3.21) \quad \sigma = ae^{-\mu\tau} \sin(\sigma\tau).$$

If $b > a$ and $\mu \geq 0$, then the magnitude of the left hand side of the real part (3.20) is always strictly greater than the magnitude of the right hand side. Thus only roots with negative real part exist, for all τ . This proves the first part of the theorem.

Now suppose $a > b$. It is clear that when $\tau = 0$, the steady state is stable ($\lambda = -a - b < 0$). We use the method described in Chapter 2. The derived polynomial equation in this case is $\sigma^2 + b^2 - a^2 = 0$. This has a solution if and only if

$a > b$. Since there is only one possible imaginary root, once a root passes to the right half plane, further increases in τ cannot remove it, so the steady state is unstable for all $\tau > \tau_c$. This completes the proof of the second part. \square

3.6 Constant *per capita* Death Rates

Now let us specify to the case of $d(x) = d$, a constant, so that we have the differential equation

$$(3.22) \quad \dot{x}(t) = (b(x(t - \tau)) - d)x(t).$$

We will focus on the interesting case, where $b(0) > d$, b is decreasing and $b(\bar{x}) = d$ for some unique \bar{x} . For this case, we prove that this system has periodic orbits when the nontrivial steady state is unstable.

Let us begin with the linear stability analysis. The nontrivial steady state, \bar{x} exists, and the linearization at this point is

$$(3.23) \quad \dot{x}(t) = b'(\bar{x})\bar{x}x(t - \tau).$$

This leads to the characteristic equation

$$(3.24) \quad \lambda = -\beta e^{-\lambda\tau},$$

where $\beta = -b'(\bar{x})\bar{x} > 0$. Note that when $\tau = 0$, the steady state is stable, as the characteristic equation has exactly one root, which is negative. If we separate the components of the eigenvalue as $\lambda = \mu + i\sigma$, then the real and imaginary parts of the characteristic equation are

$$(3.25) \quad \mu = -\beta e^{-\mu\tau} \cos(\sigma\tau),$$

$$(3.26) \quad \sigma = \beta e^{-\mu\tau} \sin(\sigma\tau).$$

Now suppose that (3.24) has a purely imaginary root, $\lambda = i\sigma$. The equation becomes,

$$(3.27) \quad 0 = -\beta \cos(\sigma\tau),$$

$$(3.28) \quad \sigma = \beta \sin(\sigma\tau).$$

We are looking for the smallest positive value of τ such that there is a solution $\sigma > 0$. From the real part (3.27), we see that $\sigma\tau = \frac{\pi}{2}$ is the smallest possible value for this product. Using this information in the imaginary part (3.28) we see that $\sigma = \beta = -b'(\bar{x})\bar{x}$. So we see that the critical delay τ_c at which the first eigenvalue with positive real part emerges is $\tau_c = \frac{\pi}{2\sigma}$, *i.e.*,

$$(3.29) \quad \tau_c = \frac{-\pi}{2b'(\bar{x})\bar{x}},$$

and for $\tau > \tau_c$, the nontrivial steady state \bar{x} is unstable.

Any characteristic root of (3.24) with positive real part is also simple. If not, then we must have

$$(3.30) \quad \lambda = -\beta e^{-\lambda\tau},$$

$$(3.31) \quad 1 = \beta\tau e^{-\lambda\tau}.$$

Substituting the first formula in the second gives

$$(3.32) \quad 1 = -\tau\lambda,$$

and it is clear that is $\text{Re}(\lambda) > 0$, then equation (3.32) cannot be.

Let us take the time now to record a couple of facts which we will refer to in proving later results. If we choose the delay τ such that the steady state is unstable, then $b'(\bar{x}) < \frac{-\pi}{2\bar{x}\tau}$. Furthermore, when $\mu > 0$, $\cos(\sigma\tau) < 0$ (from equation (3.25)) and $\sin(\sigma\tau) > 0$, when we consider the complex root with nonnegative imaginary part. So $\sigma\tau \in (\frac{\pi}{2}, \pi)$.

Lemma 3.16. *Suppose that $x(t)$ is a solution of equation (3.22), $x(t_0) = \bar{x}$, and $x(t) < \bar{x}$ for all $t \in [t_0 - \tau, t_0]$. Then for all $t > t_0$, $x(t) < \bar{x}e^{(b(0)-d)\tau} = M$.*

Proof. The function $x(t)$ is increasing for $t \in [t_0, t_0 + \tau]$, since $b(x(t - \tau)) > d$ for these times. Since $b(x)$ is a decreasing function, $\dot{x}(t) \leq (b(0) - d)x(t)$, so it is clear that $x(t_0 + \tau) \leq M$. For $t \in [t_0 + \tau, t_0 + 2\tau]$, $b(x(t - \tau)) < d$, so $x(t)$ is decreasing. If $x(t)$ remains above \bar{x} for all $t \geq t_0$, then it is always decreasing, and $x(t) < M, \forall t$. Otherwise, there is a time, t_1 such that $x(t_1) = \bar{x}$. In this case, $x(t)$ decreases on the interval $[t_1, t_1 + \tau]$. If $x(t)$ now remains below \bar{x} for $t > t_1$, then we are done. Otherwise, there is a time t_2 such that $x(t_2) = \bar{x}$. We have returned to the situation of the lemma. So we have proven that such solutions either oscillate about \bar{x} with $x(t) < M$, or else are eventually monotone (in which case $x(t) \rightarrow \bar{x}$). \square

The final preparatory definition we require is of a subset, $K \subset \mathcal{C}([-\tau, 0], \mathbb{R}^+)$ of the Banach space of initial functions.

$$K = \{\phi \in \mathcal{C}([-\tau, 0], \mathbb{R}^+) : \phi(-\tau) = \bar{x}, \phi \text{ nondecreasing, and } \phi(0) \leq M\}.$$

We will show that for any solution $x(t)$ with initial function $\phi \in K_1 = K \setminus \{\bar{x}\}$, there is a time $\tilde{t} = \tilde{t}(\phi)$ such that $x(\tilde{t} + s)$ is in K_1 .

Theorem 3.17. *Suppose that $\phi \in K_1$, and that $x(t)$ is the solution to the differential equation (3.22) with initial function ϕ . Then there exists a time t_1 such that $x(t_1) = \bar{x}$, and $\dot{x}(t_1) < 0$. Further, there exists a time $t_2 > t_1 + \tau$ such that $x(t_2) = \bar{x}$ and $\dot{x}(t_2) > 0$. If $\tilde{t} = t_2 + \tau$, then the function defined by $x(\tilde{t} + s)$ for $-\tau \leq s \leq 0$ is in K_1 .*

Proof. Suppose that t_1 does not exist, then for $t > 0$, $x(t)$ is decreasing and bounded below by \bar{x} . It follows that $x(t)$ approaches a limit L as $t \rightarrow \infty$. This is only possible

if $L = \bar{x}$. Since $b'(\bar{x}) < \frac{-\pi}{2\bar{x}\tau}$, for any $\alpha < \frac{\pi}{2}$, it is true that

$$b(x) - b(\bar{x}) \leq -\frac{\alpha}{\tau\bar{x}}(x - \bar{x}),$$

for x such that $|x - \bar{x}| < \delta' = \delta'(\alpha)$. In particular, this is true for $\alpha = 1$. See Figure (3.3) for an illustration of this fact. Choose $\delta < \min\{x(\tau), \delta'(1)\}$. For $x - \bar{x} > \delta$,

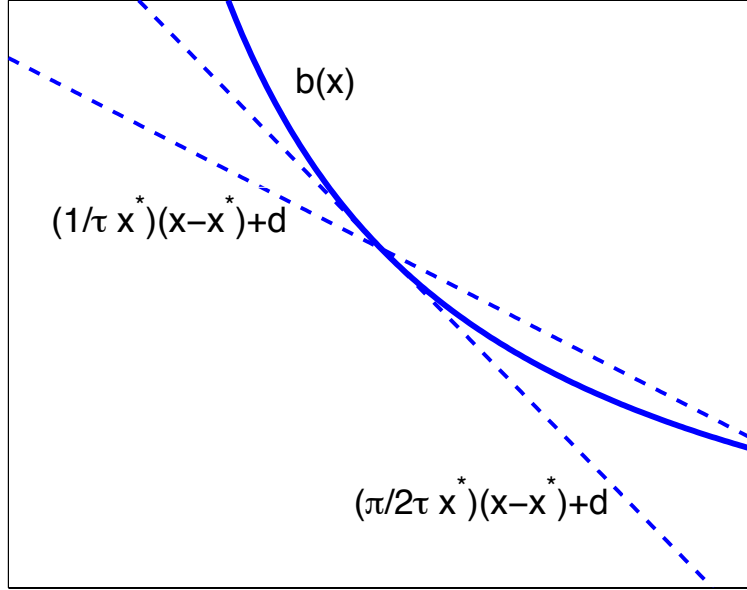


Figure 3.3: The function $b(x)$, its tangent, and a line with slope greater than the tangent

$$b(x) - d < b(\bar{x} + \delta) - d < -\frac{1}{\tau\bar{x}}\delta,$$

since $b(\bar{x}) = d$.

Now let T be a time such that $x(T) = \bar{x} + \delta$. Due to the definition of δ , $T > \tau$, and $x(t) > \bar{x} + \delta$ for $t \in [T - \tau, T]$. Then for $t \in [T, T + \tau]$, we have

$$\begin{aligned} \dot{x}(t) &= (b(x(t - \tau)) - d)x(t) \\ &\leq (b(x(t - \tau)) - d)\bar{x} \\ &\leq -\frac{1}{\tau\bar{x}}\delta\bar{x} = -\frac{\delta}{\tau}. \end{aligned}$$

Now $x(T + \tau) < x(T) - \frac{\delta}{\tau}\tau = x(T) - \delta = \bar{x}$, which is a contradiction.

We have shown that any solution with initial history in K_1 must cross the nontrivial steady state, at a time which we call t_1 . From this crossing time, the solution continues to decrease for exactly τ units of time, and then begins to increase. We now show that the solution must reach the nontrivial steady state again. Essentially the same analysis works as before, now we have

$$b(x) - b(\bar{x}) \geq -\frac{1}{\tau\bar{x}}(x - \bar{x}) = \frac{1}{\tau\bar{x}}(\bar{x} - x).$$

From this point on, the work is analogous, with the directions of the inequalities reversed. \square

The next order of business is to show that the steady state \bar{x} is an ejective fixed point to the return map. To do this we follow a method described in Kuang [32] (Section 2.9) and proven by Chow and Hale [9]. If we consider the linearized equation

$$\dot{x}(t) = -\beta x(t - \tau),$$

then for any eigenvalue λ , there is a decomposition of the space of initial functions $\mathcal{C}([-\tau, 0], \mathbb{R}^+) = P_\lambda \oplus Q_\lambda$ into subspaces invariant under the solution operator, and P_λ is the generalized eigenspace of eigenvalue λ . Let π_λ be the projection onto P_λ . Rather than proving it directly from the definition, we will use the following theorem to show that the steady state \bar{x} is ejective.

Theorem 3.18. *Suppose that the following conditions are satisfied:*

1. *There is a characteristic root λ with $\operatorname{Re}(\lambda) > 0$.*
2. *There is a closed convex set K , $\bar{x} \in K$ and $\delta > 0$ so that*

$$\inf\{\|\pi_\lambda(\phi)\| : \phi \in K, \|\phi\| = \delta\} > 0,$$

and

3. There is a completely continuous function $\tau : K \setminus \{\bar{x}\} \rightarrow [\alpha, \infty)$, $\alpha \geq 0$ such that the map defined by

$$A\phi = x_{\tau(\phi)}(\phi), \quad \phi \in K \setminus \{\bar{x}\},$$

takes $K \setminus \{\bar{x}\}$ into K and is completely continuous.

Then \bar{x} is ejective.

Since the eigenvalue λ is simple, P_λ is a one dimensional space. We define

$$\phi_1(\theta) = \frac{1}{1 + \lambda\tau} e^{\lambda\theta} = \gamma e^{\lambda\theta}, \text{ for } \theta \in [-\tau, 0]$$

$$\psi(s) = e^{-\lambda s}, \text{ for } s \in [0, \tau],$$

$$\Phi_1 = (\phi_1, \bar{\phi}_1),$$

$$\Psi = (\psi, \bar{\psi}).$$

For the linear operator L in (3.23) and $\phi \in K_1$ we define a measure $\eta(\theta)$, by

$$L(f) = -\beta\phi(-\tau) = \int_{-\tau}^0 d\eta(\theta)\phi(\theta)$$

$$\eta(-\tau) = 0, \eta(\theta) = -\beta, \text{ for } \theta \in (-\tau, 0]$$

We now compute the bilinear form

$$\begin{aligned} (\psi, \phi_1) &= \psi(0)\phi_1(0) - \int_{-\tau}^0 \int_0^\theta \psi(\xi - \theta)\phi_1(\xi)d\xi d\eta(\theta) \\ &= \gamma + \int_{-\tau}^0 \int_{-\tau}^\xi \psi(\xi - \theta)\phi_1(\xi)d\eta(\theta)d\xi \\ &= \gamma - \int_{-\tau}^0 \beta\psi(\xi + \tau)\phi_1(\xi)d\xi \\ &= \gamma - \gamma\beta \int_{-\tau}^0 e^{-(\xi+\tau)\lambda} e^{\lambda\xi} d\xi \\ &= \gamma(1 - \beta\tau e^{-\lambda\tau}) = \gamma(1 + \lambda\tau) = 1. \end{aligned}$$

Also, we have

$$\begin{aligned}
\frac{1}{\gamma}(\bar{\psi}, \phi_1) &= 1 - \beta \int_{-\tau}^0 e^{-\bar{\lambda}(\xi+\tau)} e^{\lambda\xi} d\xi \\
&= 1 - \beta e^{-\bar{\lambda}\tau} \left[\frac{1}{\lambda - \bar{\lambda}} e^{(\lambda - \bar{\lambda})\xi} \right]_{-\tau}^0 \\
&= 1 - \beta \frac{1}{\lambda - \bar{\lambda}} (e^{-\bar{\lambda}\tau} - e^{-\lambda\tau}) \\
&= \frac{1}{\bar{\lambda} - \lambda} (\lambda + \beta e^{-\lambda\tau} - (\bar{\lambda} + \beta e^{-\bar{\lambda}\tau})) = 0.
\end{aligned}$$

From these two computations, it follows readily that $(\psi, \bar{\phi}_1) = 0$ and $(\bar{\psi}, \phi_1) = 1$. So, (Ψ, Φ_1) is the identity, so for any $\phi \in \mathcal{C}([-\tau, 0], \mathbb{R}^+)$, $\pi_\lambda \phi = \Phi_1(\Psi, \phi)$. So we need to show that

$$\inf\{||(\psi, \phi - \bar{x})|| : \phi \in K_1, ||\phi - \bar{x}|| = \delta\} > 0.$$

Let $\lambda = \mu_i \sigma$, and recall that $\mu > 0$, $\sigma \tau \in (\frac{\pi}{2}, \pi)$. We can compute the coefficient $(\psi, \phi - \bar{x})$, and split it into its real and imaginary parts, yielding

$$(3.33) \quad \text{Real part: } \phi(0) - \bar{x} - \beta \int_{-\tau}^0 e^{-\mu(\xi+\tau)} (\phi(\xi) - \bar{x}) \cos(\xi + \tau) \sigma d\xi$$

$$(3.34) \quad \text{Imaginary part: } \beta \int_{-\tau}^0 e^{-\mu(\xi+\tau)} (\phi(\xi) - \bar{x}) \sin(\xi + \tau) \sigma d\xi$$

If the infimum is 0, then there is a sequence $\phi_n \in K_1$ with $||\phi_n - \bar{x}|| = \delta$, and both the real and imaginary parts above go to zero. For the given range of σ and ξ , $\sin(\xi + \tau)\sigma > 0$ and bounded away from 0 when ξ is near 0. Further, $\phi_n - \bar{x}$ is increasing, so the integral in (3.34) can only go to zero only if $||\phi_n - \bar{x}|| \rightarrow 0$, which is a contradiction. Thus the fixed point \bar{x} is ejective, and we can apply the Theorem (3.2). This system has periodic solutions when the steady state is unstable.

3.7 Delay Dependent Parameters

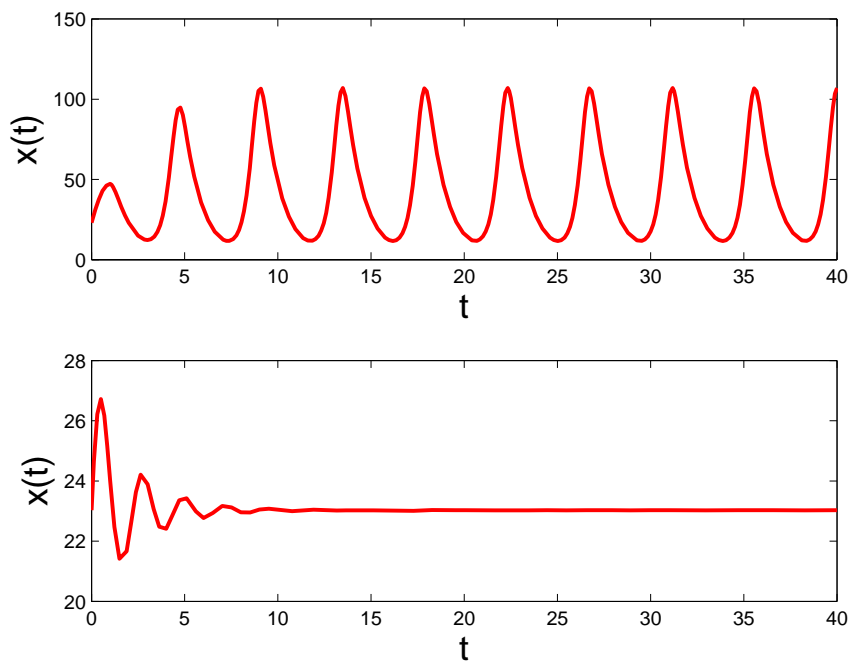


Figure 3.4: Solutions of the $\dot{x}(t) = (be^{-ax(t-\tau)} - d)x(t)$, with $a = 0.1$, $b = 10$, $d = 1$, with initial function $\bar{x} + 10t$ on $[-\tau, 0]$. $\tau_c = 0.6822$. The upper graph is for $\tau = 1$, and the second for $\tau = 0.5$.

As in section 3.4, we will now examine the effects of allowing a parameter in the equation (3.22) depend on the delay, τ . We will use the same type of dependence, so that we are interested in

$$(3.35) \quad \dot{x}(t) = (e^{-\mu\tau}b(x(t-\tau)) - d)x(t).$$

This form of the delay model allows us to obtain much more explicit results than were possible in Section 3.4. The location of the nontrivial steady state is now the value \bar{x} , for which

$$b(\bar{x}) = de^{\mu\tau},$$

and since b is decreasing, the \bar{x} is no longer biologically meaningful if $b(0) < de^{\mu\tau}$.

Thus as τ increases, the nontrivial steady state will disappear.

The characteristic equation for (3.35) is

$$\lambda = e^{-\mu\tau} b'(\bar{x}) \bar{x} e^{-\lambda\tau},$$

which is similar in form to the characteristic equation (3.24) for the delay-independent case. We can use the analysis use in the previous section to prove the following result.

Theorem 3.19. *If*

$$(3.36) \quad \frac{\pi e^{\mu\tau}}{-2b'(\bar{x})\bar{x}} < \tau < \frac{1}{\mu} \log \frac{b(0)}{d},$$

then the nontrivial steady state of (3.35) exists and is unstable. Furthermore, there exist positive, periodic solutions of this differential equation.

It must be remembered that \bar{x} is a decreasing function of τ . The first inequality in (3.36) is the condition for instability, obtained from our calculations of the critical delay, τ_c , in the delay-independent case. The second inequality is the condition for the positivity of the nontrivial steady state.

If we specify to the case where $b(x) = be^{-ax}$, as we have considered previously, then the picture becomes remarkably clear. In this case, $b'(\bar{x}) = -ab(\bar{x}) = -ade^{\mu\tau}$, $b(0) = b$, and $\bar{x} = \frac{1}{a} \ln \frac{b}{de^{\mu\tau}}$. Thus the condition for the instability of the steady state (3.36) becomes

$$\begin{aligned} \tau &> \frac{\pi}{-2ad\bar{x}} \\ &= \frac{\pi}{2d \ln \frac{b}{de^{-\mu\tau}}} \\ &= \frac{\pi}{2d \ln \frac{b}{d} - \mu\tau}. \end{aligned}$$

This becomes the quadratic equation in τ ,

$$\mu\tau^2 - \tau \ln \frac{b}{d} + \frac{\pi}{2d} < 0,$$

which is satisfied if and only if

$$(3.37) \quad \frac{1}{2\mu} \left(\ln \frac{b}{d} - \sqrt{\left(\ln \frac{b}{d} \right)^2 - \frac{2\pi\mu}{d}} \right) < \tau < \frac{1}{2\mu} \left(\ln \frac{b}{d} + \sqrt{\left(\ln \frac{b}{d} \right)^2 - \frac{2\pi\mu}{d}} \right).$$

If $\left(\ln \frac{b}{d} \right)^2 < \frac{2\pi\mu}{d}$, then no change of stability occurs.

Next we apply the second inequality from (3.36), which guarantees the existence of a positive steady state. We get $\tau < \frac{1}{\mu} \ln \frac{b}{d}$. Note that this bound lies within the bounds provided in (3.37). In fact, this is exactly the midpoint of the left and right bounds. Putting these facts together, we arrive at the following theorem.

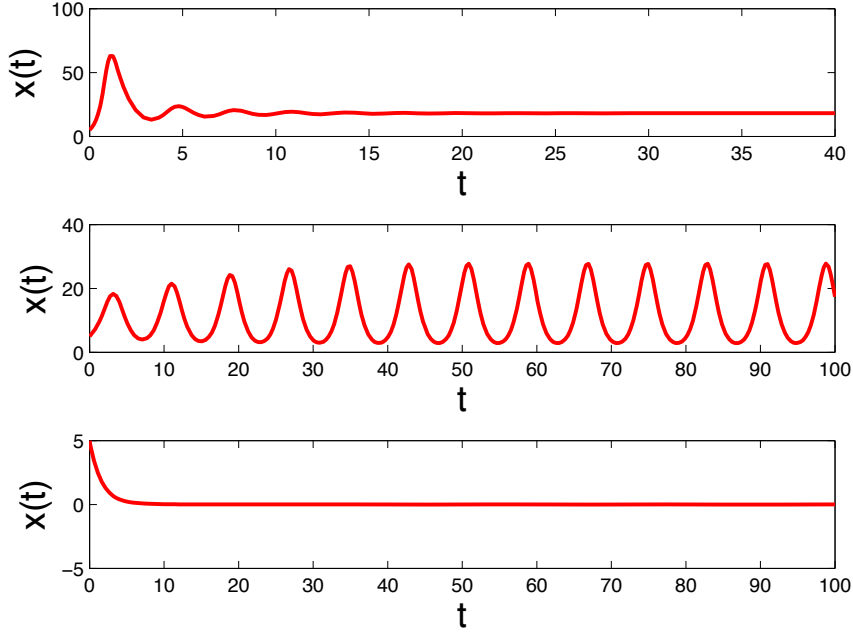


Figure 3.5: Solutions of the (3.35) with $a = 0.1$, $b = 10$, $d = 1$, $\mu = .7$, with initial function constantly 5 on $[-\tau, 0]$. The τ -region of instability determined in Theorem 3.20 is $[1.3520, 3.2894]$. The graphs are for $\tau = 0.7$, $\tau = 2$ and $\tau = 4$, respectively.

Theorem 3.20. *Consider the delay differential equation*

$$(3.38) \quad \dot{x}(t) = (be^{-\mu\tau}e^{-ax(t-\tau)} - d)x(t),$$

with $b > d$. If

$$\left(\ln \frac{b}{d}\right)^2 < \frac{2\pi\mu}{d},$$

then the nontrivial steady state is stable for all delays τ for which it exists. Otherwise,

for

$$\frac{1}{2\mu} \left(\ln \frac{b}{d} - \sqrt{\left(\ln \frac{b}{d}\right)^2 - \frac{2\pi\mu}{d}} \right) < \tau < \frac{1}{\mu} \ln \frac{b}{d},$$

the nontrivial steady state is unstable, and positive periodic solutions exist. For smaller τ , the nontrivial steady state is stable, and for larger τ , it is no longer positive, and the zero steady state is globally stable.

CHAPTER 4

Predator-Prey Interaction Models

4.1 The Lotka-Volterra Predator-Prey Interaction Model

One of the most universally recognized models in mathematics is the classic model for the interaction of a single predator species and a single prey species developed by Alfred Lotka [34] and Vito Volterra [53]. If we let x represent the prey species, and we let y represent the predator species, then the model has the form,

$$\begin{aligned} \dot{x}(t) &= ax - bxy \\ \dot{y}(t) &= cxy - dy, \end{aligned} \tag{4.1}$$

where a, b, c and d are positive constants. We see that this model includes an exponential growth term for prey in the absence of predation, and an exponential decay for predators in the absence of prey. The interaction of the two species is represented by a mass action term, which implicitly assumes that the two species encounter each other at a rate proportional to each population level, and that the effect of predation on each is in turn proportional to the number of encounters.

This system of two ordinary differential equations has two steady state solutions, $(0, 0)$ and $(\frac{d}{c}, \frac{a}{b})$. It is well known that the trivial steady state is a saddle, while the nontrivial steady state is a center, and solutions in the phase plane form an infinite family of periodic orbits (Figure 4.1).

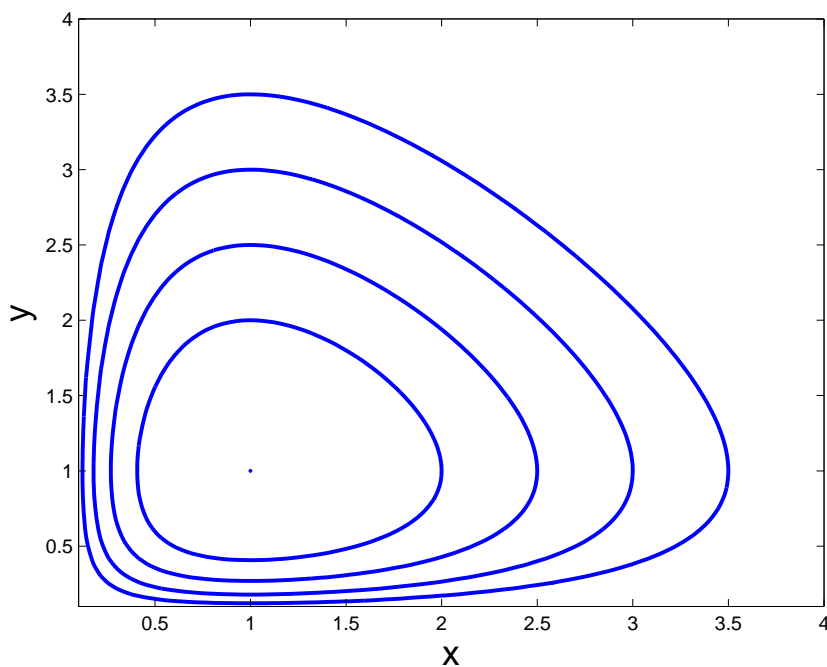


Figure 4.1: Periodic solutions of the Lotka-Volterra model with all parameters equal to 1

Periodic solutions are certainly a desirable feature of a model of predator-prey interaction, as near-periodic behaviors are often observed in nature ([18], [19], [46], [31]), although it is likely that predation is not the only factor contributing to long phase cyclic dynamics. Unfortunately, the basic Lotka-Volterra model (4.1) is not mathematically sound. It is *structurally unstable*, that is, an arbitrarily small change in the nature of the model fundamentally changes the qualitative behavior of the solutions.

For example, we could change the system in the following way

$$\begin{aligned}\dot{x}(t) &= ax - bxy - \varepsilon x^2 \\ \dot{y}(t) &= cxy - dy,\end{aligned}$$

$\varepsilon > 0$. This alteration corresponds to changing the growth of the prey in the absence of predation to logistic growth with a very large carrying capacity ($\frac{a}{\varepsilon}$). This small

change in the nature of the model completely alters the nature of the phase portrait of the models. The infinite family of periodic orbits is lost and replaced by solutions which all approach the nontrivial steady state (Figure 4.2).

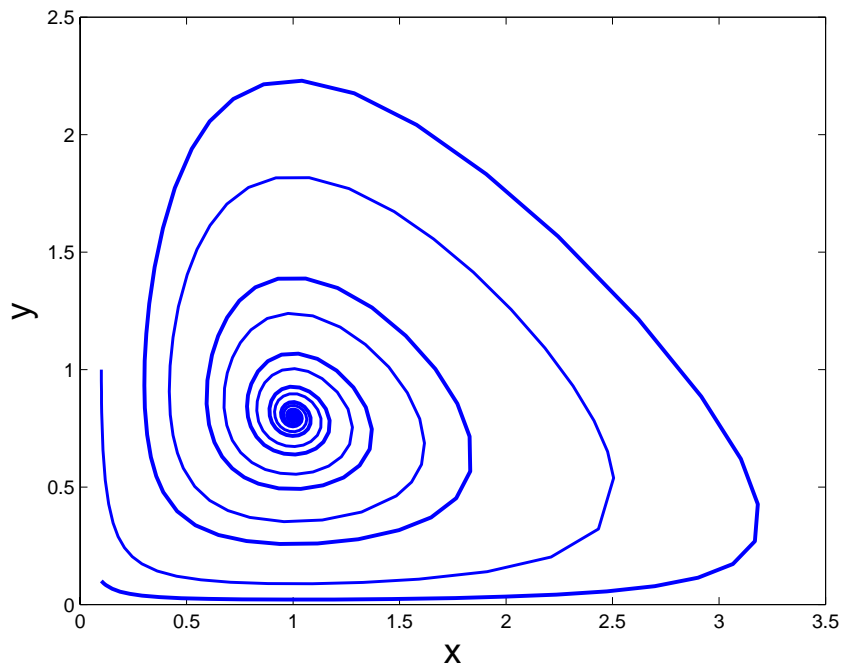


Figure 4.2: Solutions to the perturbed Lotka-Volterra model, $\varepsilon = .2$, $a = b = c = d = 1$

There are several possible ways of making the Lotka-Volterra system more palatable mathematically and biologically, each leading to interesting modelling questions and mathematical results. To begin with, we will retain the logistic growth term for the prey population in the absence of predation. Ideally, we will develop a model which corresponds well with biologically observed behavior regimes, including some kind of periodic behavior or sustained oscillation, and which is mathematically robust.

One option is to include stochastic effects in the model. This can often lead to sustained oscillations due to the constant perturbation of the system. While this is an intriguing option, it is beyond the scope of my current research.

Another option is to choose more robust nonlinearities in the predation term. While mass action is reasonable, it is not the only possibility. If we write the predation term as $p(x)y$, $p(x)$ is known as the *functional response*, and is a quantification of the relative responsiveness of the predation rate to changes in prey density at various population levels of prey. Kot [30] and Begon [1] describe four categories of functional response encountered in the ecological literature ([25], [26], [27], [28], [13]). Type I is the standard mass action or linear response

$$p(x) = cx.$$

Type II is the so-called Monod response

$$p(x) = \frac{cx}{a + x},$$

which is hyperbolic, with a saturation level (c) due to the time it takes to handle prey. Type III is a sigmoidal response

$$p(x) = \frac{cx^2}{a^2 + x^2},$$

which includes the feature that predators are inefficient when prey levels are low. These three types of functional response are all increasing functions of the prey population x . A Type IV response includes a decrease at large population levels, corresponding to prey group defenses or toxicity to predators. In the following, we will consider functional responses of Types I-III.

Thirdly, one may alter the Lotka-Volterra model by including a delay. A delay takes into account the non-instantaneous nature of biological processes. Statistical evidence has been reported ([49], [50]) of delayed effects in the density dependence of the growth rate of several insect and plant species. Another possibility for the inclusion of delays is in the interaction term $p(x)y$. This would represent the time

necessary to convert prey biomass into predator biomass, for instance due to gestation periods or time required for maturation. Some ecologists have also suggested that the inclusion of a delay could help to explain certain phenomena observed in long population cycles [5]. The inclusion of delays make the analysis of these models more difficult, but also broadens the spectrum of possible behavior regimes.

4.2 A Delay Model of Predator-Prey Interaction

We will look at a system with three populations, x is the prey population, y represents mature predators, and y_j is the juvenile predator population, which does not hunt.

$$(4.2) \quad \frac{dx}{dt} = rx\left(1 - \frac{x}{K}\right) - yp(x)$$

$$(4.3) \quad \frac{dy}{dt} = be^{-d_j\tau}y(t-\tau)p(x(t-\tau)) - dy(t)$$

$$(4.4) \quad \frac{dy_j}{dt} = by(t)p(x(t)) - be^{-d_j\tau}y(t-\tau)p(x(t-\tau)) - d_jy_j(t)$$

Let us look at the third equation in more detail. Consumed prey are converted to juvenile (immature) predator instantly with a conversion rate b . They remain in this stage of development for τ units of time, decaying exponentially at rate d_j . After this time, the survivors are removed to the class of mature predators y . It is easy to see that the third equation can be decoupled from the others, as the quantity y_j does not appear in either of the first two equations. This gives a system of two equations, and a change of variables simplifies things as well, so that we are left with

$$(4.5) \quad \begin{aligned} \dot{x}(t) &= x(1-x) - yp(x) \\ \dot{y}(t) &= be^{-d_j\tau}y(t-\tau)p(x(t-\tau)) - dy. \end{aligned}$$

The function $p(x)$ represents the adult predators functional response to prey, and we make the following assumptions

- $p(0) = 0$, *i.e.*, no predation occurs in the absence of prey,
- p is increasing,
- $p(x)/x$ is bounded and not 0 at $x = 0$.

These requirements include function responses of types I, II and III, but not IV, as the latter violate the second requirement.

The most important feature of the model is the term

$$p(x(t - \tau))y(t - \tau),$$

with the delay in both state variables. Due to this, the $\frac{dy}{dt}$ is no longer proportional to $y(t)$, the current state of the system. If the delay were omitted from $y(t)$, the behavior of this system would be much simpler to understand. Biologically, however, this type of nonlinear inclusion of the delay is entirely natural, and more logical than including the delay only in the x term. In fact, this type of term is common in delayed infection disease models [14], [39].

4.3 Preliminary Analysis

We begin by establishing some basic properties of solutions to the system (4.5).

- Given positive initial data, solutions remain positive for all time.
- Solutions are bounded (in fact, eventually uniformly bounded regardless of initial data).
- Thirdly, we need to determine steady states and their stability.
 - The non-trivial steady state becomes unstable for larger delays.
 - Periodic solutions exist.

4.3.1 Positivity of Solutions

It is relatively easy to establish that solutions remain positive for all times, given a bounded positive initial history on an interval $[a - \tau, a]$. For y , on the interval $[a, a + \tau]$, we have $\dot{y}(t) \geq -dy(t)$. Thus it is clear that y must remain positive. Further, y remains finite on this interval. If M is the bound on $x(t)$ on $[a - \tau, a]$, then $\dot{y}(t) \leq be^{-d_j\tau}y(t - \tau)p(M) - dy(t)$, which implies

$$\begin{aligned}\dot{y}(t) - dy(t) &= be^{-d_j\tau}y(t - \tau)p(x(t - \tau)) \\ \frac{d}{dt}(e^{-dt}y(t)) &= be^{-dt}e^{-d_j\tau}y(t - \tau)p(x(t - \tau))\end{aligned}$$

Integrating both sides from a to $a + \varepsilon$, $\varepsilon \in [0, \tau]$, yields

$$\begin{aligned}e^{-d(\varepsilon+a)}y(a + \varepsilon) &= e^{-da}y(a) + \int_a^{a+\varepsilon} be^{-ds}e^{-d_j\tau}y(s - \tau)p(x(s - \tau))ds \\ y(a + \varepsilon) &= e^{d\varepsilon}y(a) + \int_a^{a+\varepsilon} be^{d(\varepsilon+a-s)}e^{-d_j\tau}y(s - \tau)p(x(s - \tau))ds.\end{aligned}$$

The right hand side is finite for $\varepsilon \in [0, \tau]$, since the integrand is also bounded.

For the prey population, the rate of change is essentially proportional to x

$$\dot{x} = x(1 - x - y\frac{p(x)}{x}).$$

The state variable x can only become negative if $1 - x - y\frac{p(x)}{x}$ becomes infinite and negative as $x \rightarrow 0$, but the function $\frac{p(x)}{x}$ is bounded, and y is bounded for $t \in [a, a + \tau]$. It follows that x cannot become negative on this interval. We may iterate this argument to show that x and y are positive and finite for all $t \geq a - \tau$.

4.3.2 Uniform Boundedness of Solutions

Next we show that all solutions of (4.5) are eventually in a fixed region.

Theorem 4.1. *There exists an $M > 0$ such that for any solution $(x(t), y(t))$ of the system (4.5) with positive initial data,*

$$\max \left\{ \limsup_{t \rightarrow \infty} x(t), \limsup_{t \rightarrow \infty} y(t) \right\} \leq M.$$

Proof. Since p is a positive function for $x > 0$, and solutions remain positive for all t , we have $\dot{x}(t) \leq x(t)(1 - x(t))$. Prey solutions of (4.5) with positive initial data are thus given an upper bound by solutions of $\dot{z}(t) = z(1 - z)$ with positive initial conditions. All such solutions converge to 1, so we can conclude that $\limsup_{t \rightarrow \infty} x(t)$ is given an upper bound by 1, regardless of initial data.

Now consider the second equation of (4.5). Suppose that $be^{-d_j\tau}p(1) - d < 0$ (we shall see that this is the condition of nonexistence of a nontrivial steady state). There exists an $\varepsilon > 0$ such that $be^{-d_j\tau}p(1 + \varepsilon) - d < 0$, due to the continuity of p . Since $\limsup_{t \rightarrow \infty} x(t) \leq 1$, there exists a T_1 such that $x(t) < 1 + \varepsilon$ for all $t > T_1 - \tau$. This T_1 will depend on the particular solution (*i.e.*, initial data), but the bound provided for $\limsup_{t \rightarrow \infty} y(t)$ will not depend on T_1 . For $t > T_1$, we have

$$\begin{aligned} \dot{y}(t) &= be^{-d_j\tau}p(x(t - \tau))y(t - \tau) - d(y) \\ &\leq be^{-d_j\tau}p(1 + \varepsilon)y(t - \tau) - dy(t) \\ &= ay(t - \tau) - dy(t), \end{aligned}$$

where we define $a = be^{-d_j\tau}p(1 + \varepsilon) < d$. We have seen in Lemma 1.4 that solutions of

$$\dot{z}(t) = az(t - \tau) - dz(t)$$

approach 0 as $t \rightarrow \infty$. Further, the comparison lemma 1.5 now tells us that $y(t)$ is bounded by z , and thus goes to 0 as well. Clearly, then $\limsup_{t \rightarrow \infty} y(t) = 0$ in this case.

We are left with the case $be^{-d_j\tau} \geq d$. Since 1 is a bound on $\limsup_{t \rightarrow \infty} x(t)$, for a particular solution, there exists a time T_2 such that $x(t) < 2$ for all $t \geq T_2$. Thus

$$(4.6) \quad \dot{y}(t) \leq be^{-d_j\tau} p(2)y(t - \tau) - dy(t).$$

Looking at the equation for \dot{y} again, we have $\dot{y}(t) \geq -dy(t)$. From this, one easily concludes that for $t_2 > t_1$,

$$y(t_2) \geq y(t_1)e^{d(t_2-t_1)}.$$

In particular, let $t_2 = t > \tau$ and $t_1 = t - \tau$, and one obtains $y(t - \tau) \leq y(t)e^{d\tau}$.

Combining this information with equation (4.6) yields

$$\begin{aligned} \dot{y}(t) &\leq (be^{-d_j\tau} p(2)e^{d\tau} - d)y(t) \\ &= \Delta y(t), \end{aligned}$$

defining Δ by this second equality. Now for $t_2 > t_1$, $y(t_2) < y(t_1)e^{\Delta(t_2-t_1)}$, and this implies

$$(4.7) \quad t_2 - t_1 \geq \frac{1}{\Delta} \ln \frac{y(t_2)}{y(t_1)}.$$

Define $p_1(x)$ by $p(x) = xp_1(x)$. By our assumptions about p , we know that p_1 is bounded, positive and bounded away from 0 for $x \geq 0$. Suppose that there exists a time T_3 such that $p_1(x(t))y(t) > 1$ for all $t \geq T_3$. Then for $t \geq T_3$,

$$\dot{x}(t) = x(t)(1 - x(t) - p_1(x(t))y(t)) \leq -x(t)^2.$$

Solutions to the differential $\dot{x} = -x^2$ tend uniformly to zero, so for any $z_0 > 0$, there exists a time $T_4 > \tau$ such that $x(t) < z_0$ for all $t > T_2 + T_3 + T_4$. In particular, we shall consider the case of z_0 such that $be^{-d_j\tau} p(z_0) < de^{-d\tau} < d$. This yields the estimate of the rate of change of y

$$\dot{y} \leq ay(t - \tau) - dy(t),$$

with $a < d$ for $t \geq T_2 + T_3 + T_4$. This implies that $y(t) \rightarrow 0$, contradicting the assumption that $p_1(x(t))y(t) > 1$ for $t \geq T_2$. So $p_1(x)y$ does not remain above 1.

From this we will conclude that there is some number that $y(t)$ does not remain above. Since p_1 is positive and bounded away from 0, there exists an $m > 0$ such that $p_1(x) > m$ for all $x \geq 0$. Suppose that $y(t) > \frac{1}{m}$ for all $t > T_5$. Then $p_1(x(t))y(t) > m\frac{1}{m} = 1$ for all $t > T_5$, contradicting the result previously obtained.

Define

$$M = \max\left\{2, \frac{1}{m}e^{\Delta(T_3+T_4)}\right\}.$$

As $\limsup_{t \rightarrow \infty} x(t) \leq 1$, it is clear that $\limsup_{t \rightarrow \infty} x(t) \leq M$. It remains to show that $\limsup_{t \rightarrow \infty} y(t) \leq \frac{1}{m}e^{\Delta(T_3+T_4)}$. Suppose not. Since $y(t)$ cannot remain above $\frac{1}{m}$, there must be arbitrarily large times $\bar{t}_2 > \bar{t}_1 > 0$ such that

$$(4.8) \quad y(\bar{t}_1) = \frac{1}{m}$$

$$(4.9) \quad y(\bar{t}_2) = \frac{1}{m}e^{\Delta(T_3+T_4)}, \text{ and}$$

$$(4.10) \quad \dot{y}(\bar{t}_2) > 0.$$

One can chose $\bar{t}_1 > T_2$, where the value T_2 depends on the particular solution. Now apply the estimate (4.7), and find

$$\bar{t}_2 - \bar{t}_1 \geq \frac{1}{\Delta} \frac{\Delta(T_3 + T_4) \ln \frac{1}{m}}{\ln \frac{1}{m}} = T_3 + T_4.$$

Thus $\bar{t}_1 + T_3 + T_4 \leq \bar{t}_2$.

But $t > T_2 + T_3 + T_4$, $\dot{y}(t) < 0$. For such times, $\dot{y}(t) < be^{-d_j\tau}p(z_0)y(t-\tau) - dy(t) < de^{-d\tau}y(t-\tau) - dy(t) < de^{-d\tau}e^{-d\tau}y(t) - dy(t) = 0$. This contradicts the assumption (4.10). Thus $\limsup_{t \rightarrow \infty} y(t) < M$, and the theorem is proven. \square

From the proof of this theorem, the following result emerges. We shall refer to it when we study the steady states of the model (4.5).

Corollary 4.2. *When $be^{-d_j\tau}p(1) - d < 0$, solutions to (4.5) with positive initial data satisfy*

$$\lim_{t \rightarrow \infty} (x(t), y(t)) = (1, 0).$$

Proof. As we have seen in the previous proof, when $be^{-d_j\tau}p(1) - d < 0$, $\limsup_{t \rightarrow \infty} y(t) = 0$. Due to the positivity of solutions, this is equivalent to $\lim_{t \rightarrow \infty} y(t) = 0$. Recall also that $\limsup_{t \rightarrow \infty} x(t) \leq 1$. Thus, for $\varepsilon > 0$ there exists a time T such that for $t \geq T$, $x(t) < 1 + \varepsilon$, and, possibly by increasing T , one can assume that $p(x(t))y(t) < p(1 + \varepsilon)y(t) < \varepsilon$. Now for $t > T$, if $x(t) > 1$, then

$$\dot{x}(t) = x(t)(1 - x(t)) - p(x(t))y(t) < (1 + \varepsilon)(1 - x(t)) < 0.$$

So x is decreasing. On the other hand, if $x(t) < 1$, then

$$\dot{x}(t) = x(t)(1 - x(t)) - p(x(t))y(t) > x(t)(1 - (1 - \varepsilon)) - \varepsilon = -\varepsilon(x(t) - 1) > 0$$

for $t > T$. So, in this case, x is increasing. It follows immediately that for $t > T$, $x(t)$ cannot cross $x = 1$, and is monotone. A limit must therefore exist, and $x(t) \rightarrow 1$ is the only possibility. \square

4.3.3 Steady States

To determine the steady states of the system (4.5), we simply assume that a constant (\bar{x}, \bar{y}) is a solution and determine what these constant values must be. The equations for determining steady states are

$$(4.11) \quad 0 = \bar{x}(1 - \bar{x} - \frac{\bar{y}p(\bar{x})}{\bar{x}})$$

$$(4.12) \quad 0 = be^{-d_j\tau}p(\bar{x})\bar{y} - d\bar{y}.$$

If $\bar{y} = 0$, then the second equation is satisfied, and the first gives $(0, 0)$ and $(1, 0)$ as steady states.

If $\bar{y} \neq 0$, then the steady state equations become

$$(4.13) \quad 0 = 1 - \bar{x} - \frac{\bar{y}p(\bar{x})}{\bar{x}}$$

$$(4.14) \quad d = be^{-d_j\tau}p(\bar{x}).$$

For the equation (4.13), we must clearly have $\bar{x} \in (0, 1)$. Since p is an increasing function, it is clear that the second equation has a solution if and only if

$$(4.15) \quad p(1) > \frac{d}{be^{-d_j\tau}}.$$

So, if the condition (4.15) is satisfied, the system (4.5) has three steady state solutions: $(0, 0)$, $(1, 0)$, and a nontrivial steady state (x^*, y^*) . If (4.15) is not satisfied, then only the first two steady states exist. Note, in particular, that as the length, τ , of the delay is increased, this condition will eventually fail, due to the rational function on the left hand side of (4.15).

4.3.4 Linear Stability

The linearization of the delayed Lotka-Volterra system (4.5) about the steady state $(0, 0)$ is

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_\tau \\ y_\tau \end{pmatrix},$$

where $x_\tau = x(t - \tau)$, and similarly for y . This linear system clearly has eigenvalues 1 and $-d$, and is thus a saddle.

The linearization about the steady state $(1, 0)$ is

$$\begin{pmatrix} -1 & -p(1) \\ 0 & -d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & be^{-d_j\tau}p(1) \end{pmatrix} \begin{pmatrix} x_\tau \\ y_\tau \end{pmatrix}.$$

The characteristic equation is

$$(4.16) \quad (\lambda + 1)(\lambda + d - be^{-d_j\tau}p(1)e^{-\lambda\tau}).$$

We will now see that the stability or instability of the steady state $(1, 0)$ corresponds exactly to the nonexistence or existence of the nontrivial steady state. Clearly, $\lambda = -1$ is an eigenvalue, but has no bearing on linear stability. The stability of this steady state therefore depends on the location of the roots of

$$(4.17) \quad \lambda + d - be^{-d_j\tau}p(1)e^{-\lambda\tau} = 0.$$

Recall that the condition for the existence of a nontrivial steady state is $d < be^{-d_j\tau}p(1)$. In this case, if we rewrite the characteristic equations as

$$\lambda = be^{-d_j\tau}p(1)e^{-\lambda\tau} - d,$$

then the left hand side is 0 when $\lambda = 0$ and increases to infinity, and the left hand side is positive when $\lambda = 0$, and decreases to 0. Therefore, we see that there is always a positive real eigenvalue when the nontrivial steady state exists.

When the nontrivial steady state does not exist (*i.e.*, $d \geq be^{-d_j\tau}p(1)$) we can show that there are no eigenvalues with positive real part. Setting $\lambda = \mu + i\sigma$, with $\mu > 0$, the real part of the characteristic equation is

$$\begin{aligned} 0 &= \mu + d - be^{-d_j\tau}p(1)e^{-\mu\tau} \cos(\sigma\tau) \\ &\geq \mu + d - be^{-d_j\tau}p(1) > 0 \end{aligned}$$

So the steady state $(1, 0)$ is stable in the absence of the nontrivial steady state. In fact, it we have already shown in Corollary 4.2 that in this case, $(1, 0)$ is globally stable, as demonstrated in Figure 4.3

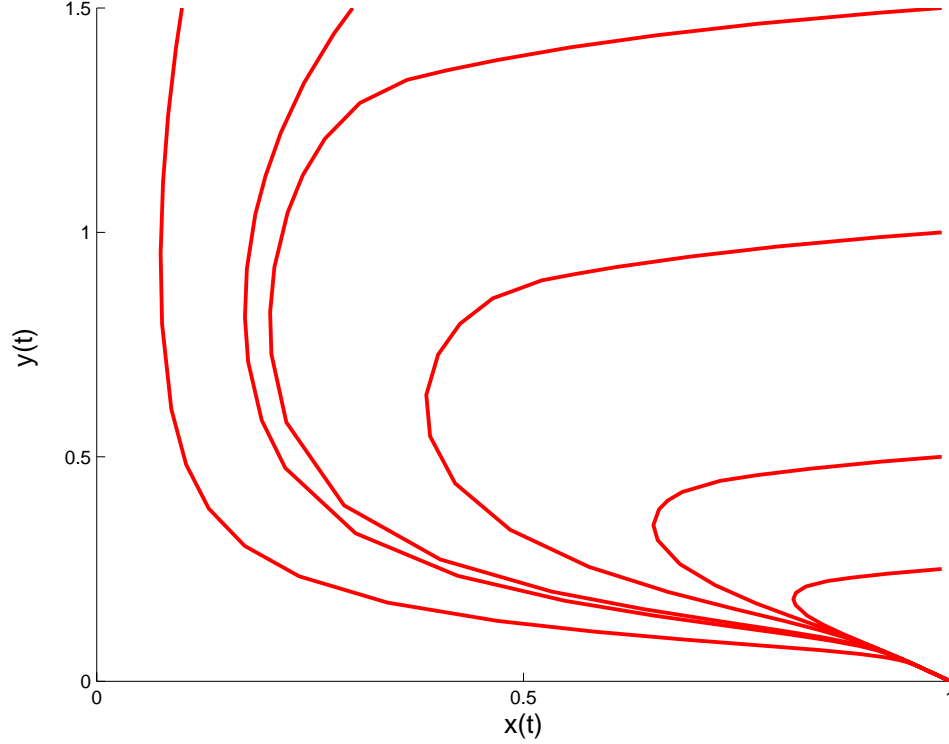


Figure 4.3: Global stability of $(1,0)$ in the absence of a nontrivial steady state

When the nontrivial steady state does exist, *i.e.*, when $d < be^{-d_j\tau}p(1)$, $(1,0)$ is always unstable. In fact, in this case the characteristic equation always has a real, positive root. To see this consider, as before,

$$\lambda = be^{-d_j\tau}p(1)e^{-\lambda\tau} - d.$$

When $\lambda = 0$, the left hand side is zero, while the right hand side is positive. As λ increases along the real line, the left hand side increases to infinity, while the right hand side decreases to $-d$. Since the functions on the left and right sides are continuous, they must intersect, proving the existence of a positive real eigenvalue.

The linear stability picture for the nontrivial steady state, (x^*, y^*) is more complicated. If we take $p(x) = px$, then we can show that for small delays, the steady state is stable.

$$P(\lambda, \tau) + Q(\lambda, \tau)e^{-(\lambda+d_j)\tau},$$

where

$$\begin{aligned}
P(\lambda, \tau) &= \lambda^2 + (2x^* + y^*p'(x^*) - 1 + d)\lambda + d(2x^* + y^*p'(x^*) - 1) \\
&= (\lambda + 2x^* + y^*p'(x^*) - 1)(\lambda + d) \\
Q(\lambda, \tau) &= p(x^*)\lambda - bp(x^*)(2x^* - 1) \\
&= p(x^*)(\lambda - b(2x^* - 1))
\end{aligned}$$

We treat the length of delay, τ , as a bifurcation parameter. One should note, in particular, that the coefficients of these polynomials depend on the location of the steady state (x^*, y^*) , which, in turn, depends on τ . When the parameters of the model are independent of delay, *i.e.*, $d_j = 0$, the location of this steady state is fixed, we may refer to the general criteria for determining whether delay induced instability occurs, which were developed earlier (Chapter 2, also [20]).

When parameters depend on delay, no such criteria exist. Using methods which depend in an essential manner on numerical estimations [3], Gourley and Kuang [23] determined that there is a range of delays for which the nontrivial steady state exists and is unstable. In this case, all steady states are unstable, and all solution are eventually trapped in a fixed region. One is naturally led to consider the possibility of periodic solutions.

4.4 Existence of Periodic Solution

The goal of my work on this two dimensional system has been to make progress toward a proof of the following conjecture.

Conjecture 4.3. *For the system*

$$\begin{aligned}\dot{x}(t) &= x(1-x) - yp(x), \\ \dot{y}(t) &= be^{-d_j\tau}y(t-\tau)p(x(t-\tau)) - dy,\end{aligned}$$

if the non-trivial steady state exists and is unstable, then a positive, nonconstant periodic solution exists.

Numerical simulations give some hope that this result might hold. If we arrange the parameters so that the nontrivial steady state exists in the absence of delay, then for small delays, this steady state is globally stable (Figure 4.4).

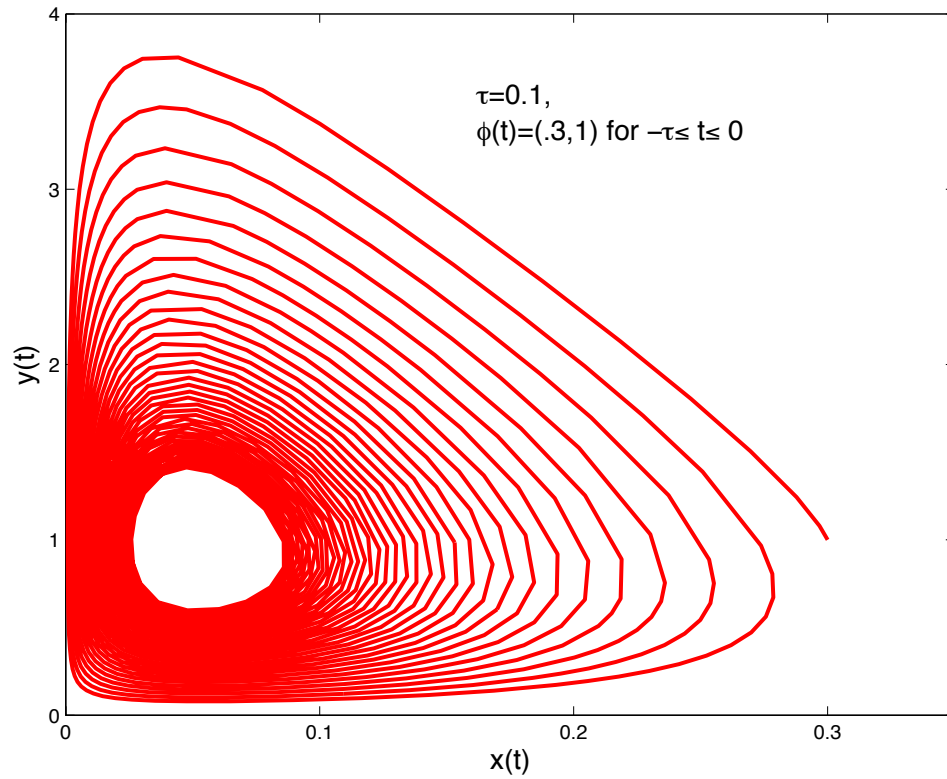


Figure 4.4: Global stability of (x^*, y^*) for small delays

As the delay is increased, a stable limit cycle appears to emerge (Figure 4.5). For certain parameter regimes, however, the behavior of solutions appears chaotic (Figures 4.6,4.7).

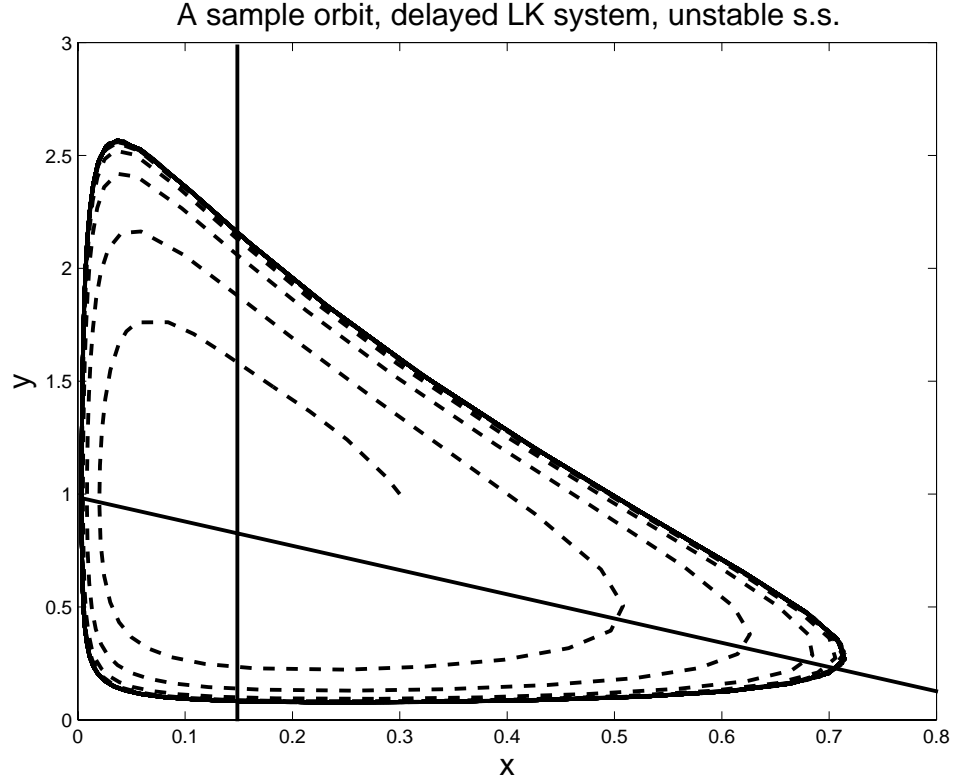


Figure 4.5: Emergence of a stable limit cycle

4.4.1 The “Phase Plane”

If we plot y against x , then we get the “phase plane”, where it is easier to see the interaction of the two population levels. In particular, it is useful to divide the x - y plane into the following regions,

$$R_1 = \{(x, y) : x \leq 0, f(x, y) \geq 0\}$$

$$R_2 = \{(x, y) : x \leq 0, f(x, y) \leq 0\}$$

$$R_3 = \{(x, y) : x \geq 0, f(x, y) \leq 0\}$$

$$R_4 = \{(x, y) : x \geq 0, f(x, y) \geq 0\},$$

where $f(x, y)$ is defined by $\dot{x} = -p(x)f(x, y)$, *i.e.*, $f(x, y) = y - \frac{x(1-x)}{p(x)}$.

This division of the phase plane is depicted in Figure 4.8. It should be noted that only the curve Γ is a true nullcline (in this case for x). When solutions are above

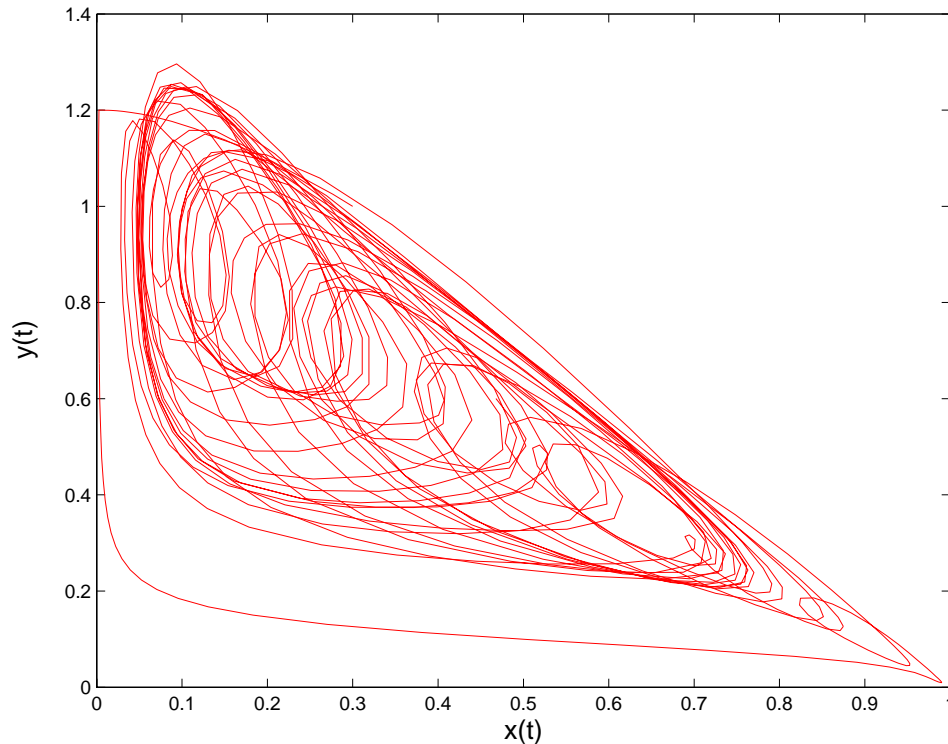


Figure 4.6: Chaotic solutions in the phase plane

this curve, x is decreasing, and when below, x is increasing. The vertical line $x = x^*$ is included only for reference. Due to the delays involved in the rate of change of y , no meaningful nullcline can be drawn.

Furthermore, this is not a phase plane in the traditional sense; solutions can cross each other, or even themselves. This possibility is demonstrated in Figure 4.6. Due to this complication, we cannot apply such geometrically-based results as Poincare-Bendixson and Bendixson-Dulac to prove the existence or otherwise of periodic solutions. We expect from the phase plane depicted in Figure 4.8 that solutions will oscillate in a counterclockwise direction, but this behavior is much trickier to prove than in the case of ordinary differential equations.

4.4.2 Oscillation of Solutions

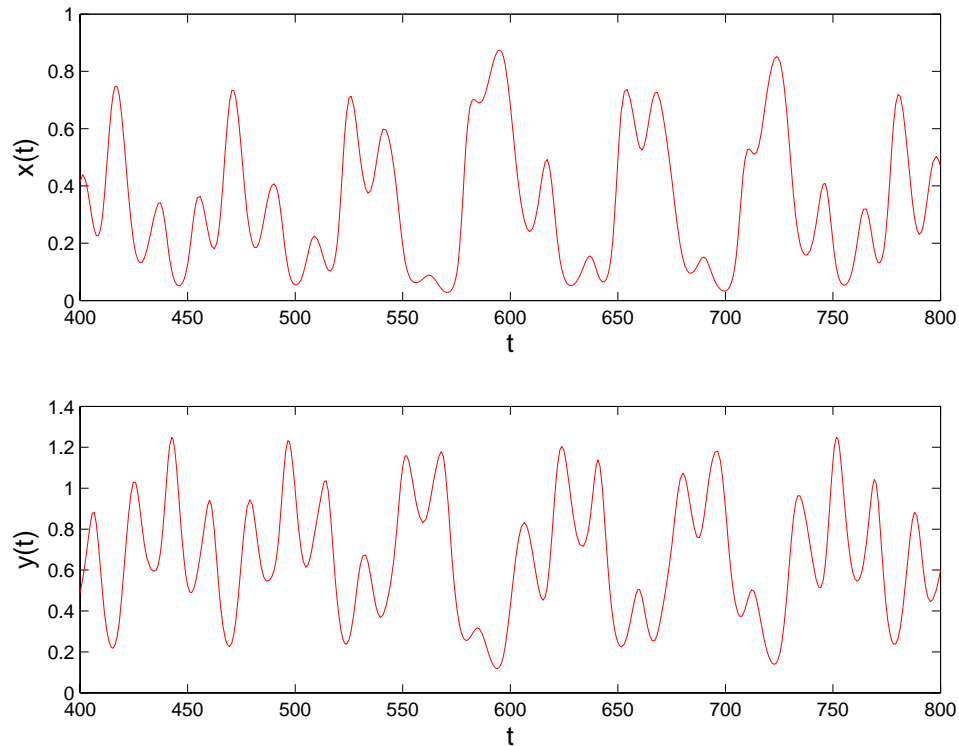


Figure 4.7: Time series for a chaotic solution

As a first step in showing that solutions do indeed oscillate about the steady state when it is unstable, we show that if the x component of solutions remain eventually above or below $x = x^*$, they must approach the steady state. This result is contained in the following theorems

Theorem 4.4. *If there exists a T such that $x(t) < x^*$ for all $t > T$, then $(x(t), y(t)) \rightarrow (x^*, y^*)$ as $t \rightarrow \infty$.*

Proof. We begin with the differential equation for $y(t)$

$$\dot{y}(t) = be^{-d_j\tau}y(t-\tau)p(x(t-\tau)) - dy(t).$$

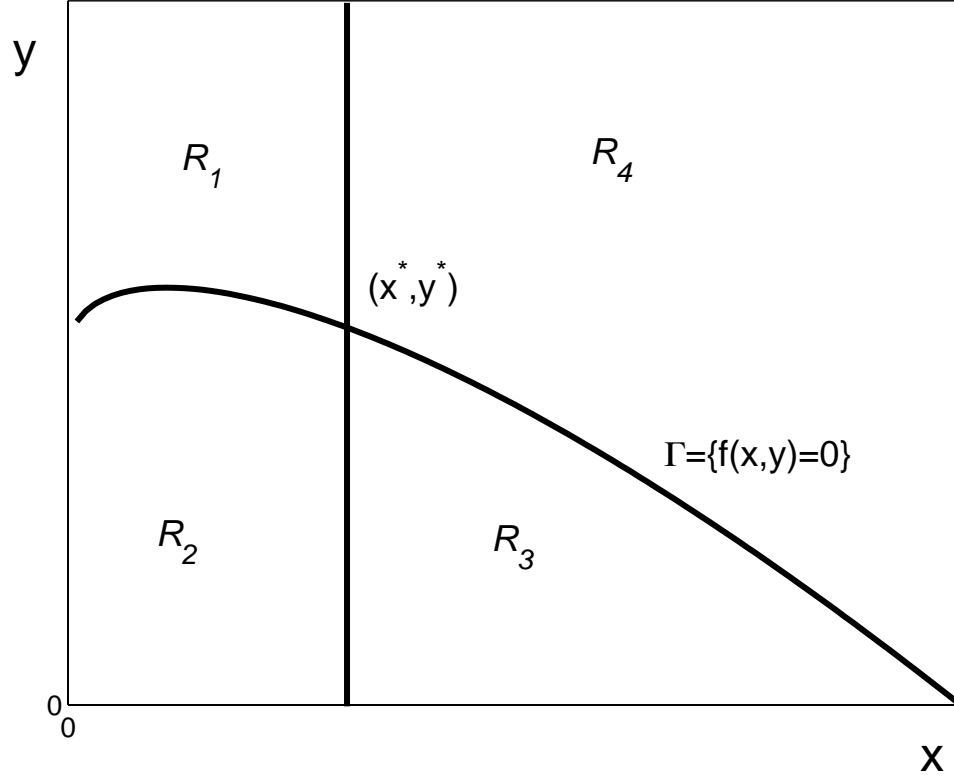


Figure 4.8: The Division of the phase planes in to the regions R_i

Now integrate both sides from T to t , to get

$$\begin{aligned}
 y(t) - y(T) &= \int_T^t [be^{-d_j\tau}y(s-\tau)p(x(s-\tau)) - dy(s)]ds \\
 &= \int_{T-\tau}^{t-\tau} be^{-d_j\tau}y(s)p(x(s))ds - \int_T^t dy(s)ds \\
 &= \int_{T-\tau}^T be^{-d_j\tau}y(s)p(x(s))ds + \int_T^{t-\tau} be^{-d_j\tau}y(s)p(x(s)) - \int_T^t dy(s)ds.
 \end{aligned}$$

Now define the constant A by

$$A = y(T) + \int_{T-\tau}^T be^{-d_j\tau}y(s)p(x(s))ds.$$

Note that A is completely determined by the initial history of the delay differential equation on the time interval $[T - \tau, T]$.

From the above equation, we can derive two inequalities. First, we have

$$(4.18) \quad y(t) \leq A + \int_T^t be^{-d_j\tau} y(s)p(x(s)) - \int_T^t dy(s)ds$$

$$(4.19) \quad = A - \int_T^t (d - be^{-d_j\tau} p(x(s)))y(s)ds.$$

We shall now use this bound on y to see that $x(t) \rightarrow x^*$ under the hypothesis of this theorem.

We begin with the case $x(t) < x^*$, *i.e.*, $be^{-d_j\tau} p(x(t)) < d$, and consider the inequality (4.19). The integrand is positive, so the integral is increasing with t . Since $y(t)$ is known to be positive, we must have

$$\int_T^\infty (d - be^{-d_j\tau} p(x(s)))y(s)ds < \infty,$$

and the continuity of the integrand then allows us to conclude that

$$(d - be^{-d_j\tau} p(x(t)))y(t) \rightarrow 0,$$

as $t \rightarrow \infty$. One may not immediately conclude that either of the terms of this product approaches 0, but we will show that indeed, $d - be^{-d_j\tau} p(x(t))$ must approach 0, which is to say, that $x \rightarrow x^*$.

To see this, consider the times t_1, t_2, \dots at which $x(t)$ has a relative minimum. It is obvious that these times can only occur when the solution crosses the curve Γ . In the region where $x < x^*$, the y values of the curve Γ are bounded below by some non-zero m . Thus $(d - be^{-d_j\tau} p(x(t_i)))y(t_i) \geq (d - be^{-d_j\tau} p(x(t_i)))m \geq 0$. Since the left-hand side goes to 0, the right hand side must do so as well. But this is only possible if $be^{-d_j\tau} p(x(t_i)) \rightarrow d$, *i.e.*, $x(t_i) \rightarrow x^*$, and if the relative minima approach x^* , then it is simple to see that $x(t) \rightarrow x^*$.

If $x \rightarrow x^*$, then $\dot{x} \rightarrow 0$, and we can see from the differential equation for x that $y(t) \rightarrow y^*$. This proves the theorem for the first case. \square

We can prove the same result in the case that $x(t) > x^*$. Before doing so, we need to establish the following lemma.

Lemma 4.5. *If $x(t) > x^*$ for $t > T$, and the initial history of x and y are positive, then y is bounded away from 0 for $t > T$.*

Proof. For positive initial data, it has already been shown in [23] we have already seen that solutions are positive. We deal with two cases: y has finite number of relative minima, and y has an infinite number relative minima.

In the first case, if $y(t)$ is not bounded away from 0, then $y(t) \rightarrow 0$, and there exists a $T_2 > T$ such that $\dot{y}(t) < 0$ for all $t > T_2$. So for $t > T_2 + \tau$

$$\begin{aligned} 0 &> be^{-d_j\tau}p(x(t-\tau))y(t-\tau) - dy(t) \\ y(t) &> \frac{be^{-d_j\tau}p(x(t-\tau))}{d}y(t-\tau) \geq y(t-\tau) \end{aligned}$$

which contradicts the assumption that $y(t)$ is decreasing.

For the second case, consider the times $t_1 < t_2 < t_3 < \dots$ at which $y(t)$ has a relative minimum. At such times we have $\dot{y}(t_i) = 0$, i.e.

$$y(t_i) = \frac{be^{-d_j\tau}p(x(t_i-\tau))}{d}y(t_i-\tau) \geq y(t_i-\tau) \geq y(t_j)$$

for some $j < i$. We can continue thus until we arrive at $y(t)$ for some $t \in [T - \tau, T]$, and thus

$$\ell = \min_{t \in [T-\tau, T]} y(t) > 0$$

is a positive lower bound of $y(t)$ with $t > T$. □

Theorem 4.6. *If there exists a T such that $x(t) > x^*$ for $t > T$, then $(x(t), y(t)) \rightarrow (x^*, y^*)$ as $t \rightarrow \infty$.*

Proof. Let M be an upper bound on $y(t)$. We begin as before with

$$(4.20) \quad y(t) = A + \int_T^{t-\tau} be^{-d_j\tau} y(s)p(x(s))ds - \int_T^t dy(s)ds$$

$$(4.21) \quad = A + \int_T^{t-\tau} (be^{-d_j\tau} p(x(s)) - d)y(s)ds - d \int_{t-\tau}^t y(s)ds$$

$$(4.22) \quad \geq A + \int_T^{t-\tau} (be^{-d_j\tau} p(x(s)) - d)y(s)ds - dM\tau.$$

The function $y(t)$ is bounded above, so the lower bound given by (4.22) must remain finite as $t \rightarrow \infty$. As in the proof of the previous theorem, since the integrand is positive, we must have $(be^{d_j\tau} p(x(t)) - d)y(t) \rightarrow 0$, but Lemma 4.5 proves that y is bounded away from 0 under the hypotheses of the theorem. It follows that, $be^{-d_j\tau} p(x(t)) - d \rightarrow 0$, and as in the previous theorem, this implies that $(x(t), y(t)) \rightarrow (x^*, y^*)$. \square

Now, when we choose τ large enough that the nontrivial steady state (x^*, y^*) is unstable, it remains to derive a contradiction from this limiting behavior. Given such a contradiction, we conclude that $x(t)$ is not less than x^* for all t , and the solution curve must leave the region $R_1 \cup R_2$. The only possibility for this to occur is for the curve to pass from region R_2 to region R_3 at a point with $y < y^*$. This is clear since x is decreasing when the solution is above the curve Γ .

We have shown the following,

Theorem 4.7. *If there exists a T such that $x(t) < x^*$ or $x(t) > x^*$ for all $t > T$, then*

$$(x(t), y(t)) \rightarrow (x^*, y^*)$$

as $t \rightarrow \infty$.

4.5 Future Work

Although much has been done to better our understanding of this system, much work remains. To begin with, a contradiction must be derived to the possibility of a solution approaching the linearly unstable nontrivial steady state as $t \rightarrow \infty$. Barring this, some other argument must be made to guarantee that solutions cross into the region R_3 . Once this is accomplished a similar argument will provide the desired return map.

More generally, there are qualitative questions to answer about the nature of the solution space for this model. For example, are multiple periodic solutions possible? Also, when nontrivial periodic solutions do exist, what are their stability properties? Numerical evidence (for example Figures 4.6 and 4.7) suggests the existence of chaotic solution regimes. What conditions lead to this behavior for solutions?

Finally, how are these dynamics changed when the system is expanded to include more equations? Such systems can be used as models for food chains. Even in the case of ordinary differential equations, food chain systems based on the same principles as Lotka-Volterra predator prey systems can display a wide variety of dynamics. Understanding the delay models could provide more insight into the nature of such systems, or demonstrate that such models are inappropriate for modeling such biological situations.

CHAPTER 5

Conclusion

The use of delay differential equations in the modeling of biological phenomena has become more prevalent in recent years. Analytic results about the behavior of such models is still largely lacking. While numerical simulations provide a basic understanding of these systems, and allow, for example, the use of parameter fitting, even when analytic results are unavailable. To be sure, increased computation capacity and speed make the use of such simulations easier. A better analytic understanding of these models, however, would make the use of numerics even more useful, and help in the selection of appropriate models in the first place.

The methods of Chapter 2 provide a straightforward and easily applicable method for analyzing the linear stability of the steady states of such models. The later chapters focused on showing the existence of periodic solutions. The methods for approaching such questions remain quite cumbersome. Ideally, a better understanding of the functional analytic theorems at work here would lead to easier determination of the existence or otherwise of periodic solutions, at least in the case of a system of only two differential equations. For ordinary differential equations, one has theorems such as Poincare-Bendixson which allow one to draw conclusions based solely on global properties (the existence of a trapping region) and linear instability. I

hope that continued study of the question of periodicity will lead to steps in the direction of such theorems for delay models. At the very least, a simpler method of determining the ejectivity of a fixed point would be quite welcome.

I have spent much time in this thesis attempting to determine the properties of delay differential equations models. I have mentioned that understanding these properties would make it easier to determine the appropriateness of these models for biological phenomena. Much work remains to be done on this question. Although it seems intuitively clear that delays occur in nature, and that they might therefore play a significant role in the dynamics of a given system, the models I have studied are only first approximations. All of the models studied incorporate a discrete delay. In other words, the dynamics depend on the current state of the system and the state of the system *exactly* τ time units ago. This way of including the delay requires much refinement.

Consider the example of human pregnancy. The gestation period is generally stated to be nine months, but this is hardly exact. If such a reproductive delay is significant in the dynamics of some model, then surely the variation about the mean delay time will also be significant. Discrete delays are only an approximation. These systems ought to be studied, since the chance of obtaining concrete results is greater for discrete delays than for their distributed cousins, and knowledge of their behavior provides insight into more complete, distributed models. One suspects that the behavior of the discrete model should correspond to the expected behavior, for example, of a stochastic model, where the length of delay is determined by a probability distribution function. If discrete delay models are to serve as approximations, however, it will be important to determine the extent to which their behavior is an artifact of the essentially discontinuous inclusion of past data.

As biologists turn to mathematics to provide a framework for understanding more and more complicated phenomena, it is important to have as many modeling techniques as possible available for use. While the inclusion of delays is but one approach among many, the theory behind it should continue to be developed, with an eye especially toward practical results and the ability to draw applicable conclusions.

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] M. Begon and M. Mortimer. *Population Ecology*. Blackwell Scientific Publications, Oxford, 1981.
- [2] R. Bellman and K. L. Cooke. *Differential-Difference Equations*. Academic Press, New York, 1963.
- [3] E. Beretta and Y. Kuang. Geometric stability switch criteria in delay differential systems with delay dependent parameters. *SIAM J. Math. Anal.*, 33(5):1144–1165, 2002.
- [4] S.P. Blythe. Instability and complex dynamic behaviour in population models with long time delays. *Theor. Pop. Biol.*, 22:147–176, 1982.
- [5] R. Boonstra, C.J. Krebs, and N.C. Stenseth. Population cycles in small mammals: The problem of explaining the low phase. *Ecology*, 79:1479–1488, 1998.
- [6] T.A. Burton. *Stability and Periodic Solutions of Ordinary and Functional Differential Equations*. Academic Press, New York, 1985.
- [7] S. A. Campbell, R. Edwards, and P. van den Driessche. Delayed coupling between two neural network loops. *SIAM J. Appl. Math.*, 65(1):316–335, 2004.
- [8] N. G. Chebotarev and N. N. Meiman. The Routh-Hurwitz problem for polynomials and entire functions. *Trudy Mat. Inst. Steklov.*, 26, 1949.
- [9] S.-N. Chow and J. K. Hale. Periodic solutions of autonomous equations. *J. Math. Anal. Appl.*, 66:495–506, 1978.
- [10] S. M. Ciupe, B. L. de Bivort, D. M. Bortz, and P. W. Nelson. Estimates of kinetic parameters from HIV patient data during primary infection through the eyes of three different models. *Math. Biosci.* in press.
- [11] K. Cooke, Y. Kuang, and B. Li. Analyses of an antiviral immune response model with time delays. *Canad. Appl. Math. Quart.*, 6(4):321–354, 1998.
- [12] K. L. Cooke, P. van den Driessche, and X. Zou. Interaction of maturation delay and nonlinear birth in population and epidemic models. *J. Math. Biol.*, 39:332–352, 1999.
- [13] M. J. Crawley. *Natural Enemies: The Population Biology of Predators, Parasites and Disease*. Blackwell Scientific Publications, Oxford, 92.
- [14] R. V. Culshaw and S. Ruan. A delay-differential equation model of HIV infection of CD4+ T-cells. *Math. Biosci.*, 165:27–39, 2000.
- [15] R. D. Driver. *Ordinary and Delay Differential Equations*. Springer-Verlag, New York, 1977.
- [16] L. Edelstein-Keshet. *Mathematical Models in Biology*. McGraw-Hill, New York, 1988.
- [17] L.E. El'sgol'ts and S.B. Norkin. *An Introduction to the Theory and Application of Differential Equations with Deviating Arguments*. Academic Press, New York, 1973.

- [18] J.P. Finerty. *The Population Ecology of Cycles in Small Mammals*. Yale University Press, New Haven, 1980.
- [19] J.R. Flowerdew. *Mammals: Their Reproductive Biology and Population Ecology*. Edward Arnold, London, 1987.
- [20] J. Forde and P. W. Nelson. Applications of Sturm sequences to bifurcation analysis of delay differential equation models. *J. Math. Anal. Appl.*, 300:273–284, 2004.
- [21] H. I. Freedman and J. H. Wu. Periodic solutions of single species models with periodic delay. *SIAM J. Math. Anal.*, 23:689–701, 1992.
- [22] H. I. Freedman and H. X. Xia. Periodic solutions of single species models with delay. *Differential Equations, Dynamical Systems and Control Science*, pages 55–74, 1994.
- [23] S. A. Gourley and Y. Kuang. A stage structured predator-prey model and its dependence on maturation delay and death rate. *J. Math. Biol.*, 49:188–200, 2004.
- [24] W. S. C. Gurney, S. P. Blythe, and R. M. Nisbet. Nicholson’s blowfly revisited. *Nature (London)*, 287:17–21, 1980.
- [25] C. S. Holling. The characteristics of simple types of predation and parasitism. *Can. Entomol.*, 91:385–398, 1959.
- [26] C. S. Holling. The components of predation as revealed by the study of small mammal predation of the European pine sawfly. *Can. Entomol.*, 91:293–320, 1959.
- [27] C. S. Holling. The functional response of predators to prey density and its role in mimicry and population regulation. *Mem. Entomol. Soc. Can.*, 45:1–60, 1965.
- [28] C. S. Holling. The functional response of invertebrate predators to prey density. *Mem. Entomol. Soc. Can.*, 47:2–86, 1966.
- [29] E.I. Jury and M. Mansour. Positivity and nonnegativity conditions of a quartic equation and related problems. *IEEE, Transactions on Automatic Control*, 26:444–451, 1981.
- [30] M. Kot. *Elements of Mathematical Ecology*. Cambridge University Press, Cambridge, 2001.
- [31] C.J. Krebs, S. Boutin, R. Boonstra, A.R.E. Sinclair, J.N.M. Smith, M.R.T. Dale, K. Martin, and R. Turkington. Impact of food and predation on the snowshoe hare cycle. *Science*, 269:1112–1115, 1995.
- [32] Y. Kuang. *Delay Differential Equations with Applications to Population Biology*. Academic Press, New York, 1993.
- [33] M. S. Lee and C.S. Hsu. On the τ -decomposition method of stability analysis for retarded dynamical systems. *SIAM J. of Control*, 7:242–59, 1969.
- [34] A. Lotka. *Elements of Physical Biology*. Williams and Wilkins, Baltimore, 1925.
- [35] N. MacDonald. *Biological Delay Systems: Linear Stability Theory*. Cambridge University Press, Cambridge, 1989.
- [36] M. C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197:287–289, 1977.
- [37] R.M. May. *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, 1974.
- [38] P. W. Nelson, J. D. Murray, and A. S. Perelson. A model of HIV-1 pathogenesis that includes an intracellular delay. *Math. Biosci.*, 163:201–215, 2000.

- [39] P. W. Nelson and A. S. Perelson. Mathematical analysis of delay differential equation models of HIV-1 infection. *Math. Biosci.*, 179:73–94, 2002.
- [40] A.J. Nicholson. An outline of the dynamics of animal populations. *Aust. J. Zool.*, 2:9–65, 1954.
- [41] A.J. Nicholson. The self adjustment of populations of change. *Cold Spring Harb. Symp. quant. Biol.*, 22:153–173, 1957.
- [42] R. D. Nussbaum. Periodic solutions to some nonlinear autonomous functional differential equations. *Ann. Mat. Pura Appl. (4)*, 101:263–306, 1974.
- [43] L. S. Pontriagin. On the zeros of some elementary transcendental functions. *Izv. Acad. Nauk SSSR*, 6(3):115–134, 1942.
- [44] M. M. Postnikov. Stable polynomials. *Nauka*, 1982.
- [45] A. Prestel and C. N. Delzell. *Positive Polynomials: from Hilbert’s 17th problem to real algebra*. Springer-Verlag, Berlin, 2001.
- [46] A.R.E. Sinclair, D. Chitty, C.I. Stefan, and C.J. Krebs. Mammal population cycles: evidence for intrinsic differences during snowshoe hare cycles. *Can. J. Zool./Rev. Can. Zool.*, 81:216–220, 2003.
- [47] P. Smolen, D. Baxter, and J. Byrne. A reduced model clarifies the role of feedback loops and time delays in the *Drosophila* circadian oscillator. *Biophys. J.*, 83:2349–2359, 2002.
- [48] C. E. Taylor and R. R. Sokal. Oscillation in housefly populations due to time lag. *Ecology*, 57:1060–1067, 1976.
- [49] P. Turchin. Rarity of density dependence or population regulation with lags. *Nature*, 344:660–663, 1990.
- [50] P. Turchin and A. D. Taylor. Complex dynamics in ecological time series. *Ecology*, 73:289–305, 1992.
- [51] B. Vielle and G. Chauvet. Delay equation analysis of human respiratory stability. *Math. Biosci.*, 152(2):105–122, 1998.
- [52] M. Villasana and A. Radunskaya. A delay differential equation model for tumor growth. *J. Math. Biol.*, 47(3):270–294, 2003.
- [53] V. Volterra. Varizioni e fluttuazioni del numero d’individui in specie animali conviventi. *Mem. R. Acad. Naz. dei Lincei (ser. 6)*, 2:31–113, 1926.
- [54] W. Wang, P. Fergola, and C. Tenneriello. Global attractivity of periodic solutions of population models. *J. Math. Anal. Appl.*, 211:498–511, 1997.
- [55] P.J. Wangersky and W. J. Cunningham. On time lags in equations of growth. *Proc. Nat. Acad. Sci. USA*, 42:699–702, 1956.
- [56] T. Zhao. Global periodic solutions for a differential delay system modeling a microbial population in the chemostat. *J. Math. Anal. Appl.*, 193:329–352, 1995.

ABSTRACT

Delay Differential Equation Models in Mathematical Biology

by

Jonathan Erwin Forde

Chair: Patrick W. Nelson

In this dissertation, delay differential equation models from mathematical biology are studied, focusing on population ecology. In order to even begin a study of such models, one must be able to determine the linear stability of their steady states, a task made more difficult by their infinite dimensional nature. In Chapter 2, I have developed a method of reducing such questions to the problem of determining the existence or otherwise of positive real roots of a real polynomial. The method of Sturm sequences is then used to make this determination. In particular, I developed general necessary and sufficient conditions for the existence of delay-induced instability in systems of two or three first order delay differential equations. These conditions depend only on the parameters of the system, and can be easily checked, avoiding the necessity of simulations in these cases.

With this tool in hand, I begin studying delay differential equations for single species, extending previously obtained results about the existence of periodic solu-

tions, and developing a proof for a previously unproven case. Due to the infinite dimensional nature of these equations, it is quite difficult to prove the existence of periodic solutions. Nonetheless, knowledge of their existence is essential if one is to make decisions about the suitability of such models to biological situations. Furthermore, I explore the effect of delay-dependent parameters in these models, a feature whose use is becoming more common in the mathematical biology literature.

Finally, I look at a delayed predator-prey model with delay dependent parameters. Although I was unable to obtain a complete proof for the existence of periodic solutions, significant progress has been made in understanding the nature of this system, and it is hoped that future work will continue to clarify this picture. This model seems to display chaotic behavior for certain parameter regimes, and thus the existence of periodic solutions may be precluded in the most general case.

5-2009

Parameter Synthesis in Nonlinear Dynamical Systems: Application to Systems Biology

Alexandre Donze
Carnegie Mellon University

Gilles Clermont
University of Pittsburgh

Axel Legay
Carnegie Mellon University

Christopher J. Langmead
Carnegie Mellon University

Follow this and additional works at: <http://repository.cmu.edu/compsci>

Published In

S. Batzoglou (Ed.): RECOMB 2009, LNCS 5541, 155-169.

This Conference Proceeding is brought to you for free and open access by the School of Computer Science at Research Showcase @ CMU. It has been accepted for inclusion in Computer Science Department by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

Parameter Synthesis in Nonlinear Dynamical Systems: Application to Systems Biology

Alexandre Donzé¹, Gilles Clermont²,
Axel Legay¹, and Christopher J. Langmead^{1,3*}

¹ Computer Science Department, Carnegie Mellon University, Pittsburgh, PA

² Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA

³ Lane Center for Computational Biology, Carnegie Mellon University, Pittsburgh, PA

Abstract. The dynamics of biological processes are often modeled as systems of nonlinear ordinary differential equations (ODE). An important feature of nonlinear ODEs is that seemingly minor changes in initial conditions or parameters can lead to radically different behaviors. This is problematic because in general it is never possible to know/measure the precise state of any biological system due to measurement errors. The parameter synthesis problem is to identify sets of parameters (including initial conditions) for which a given system of nonlinear ODEs does not reach a given set of undesirable states. We present an efficient algorithm for solving this problem that combines sensitivity analysis with an efficient search over initial conditions. It scales to high-dimensional models and is exact if the given model is affine. We demonstrate our method on a model of the acute inflammatory response to bacterial infection, and identify initial conditions consistent with 3 biologically relevant outcomes.

Key words: Verification, Nonlinear Dynamical Systems, Uncertainty, Systems Biology, Acute Illness

1 Introduction

The fields of Systems Biology, Synthetic Biology, and Medicine produce and use a variety of formalisms for modeling the dynamics of biological systems. Regardless of its mathematical form, a model is an invaluable tool for thoroughly examining how the behavior of a system changes when the initial conditions are altered. Such studies can be used to generate verifiable predictions, and/or to address the uncertainty associated with experimental measurements obtained from real systems.

In this paper, we consider the *parameter synthesis problem* which is to identify sets of parameters for which the system does (or does not) reach a given set of states. Here, the term “parameter” refers to both the initial conditions of the model (e.g., bacterial load at time $t = 0$) and dynamical parameters (e.g., the bacterium’s doubling rate). For example, in the context of medicine, we might be interested in partitioning the parameter space into two regions — those that, without medical intervention, deterministically lead to the patient’s recovery, and those that lead to the patient’s death. The parameter synthesis problem is relatively easy to solve when the system has linear dynamics, and there are a variety of methods for doing so (e.g., [6–8]). Our algorithm, in contrast, solves the parameter synthesis problem for *nonlinear dynamical systems*. That is, for systems of nonlinear ordinary differential equations (ODEs). Moreover, our approach can also be extended to nonlinear hybrid systems (i.e., those containing mixtures of discrete and continuous variables, see [11] for details). Nonlinear ODE and hybrid models are very common in the Systems Biology, Synthetic Biology, and in Medical literature but there are very few techniques for solving the parameter synthesis problem in such systems. This paper’s primary contribution is a practical algorithm that can handle systems of this complexity.

Our algorithm combines sensitivity analysis with an efficient search over parameters. The method is exact if the model has affine dynamics. For nonlinear dynamical systems, we can guarantee an arbitrarily

* Corresponding Author: cjl@cs.cmu.edu

high degree of accuracy with respect to identifying the boundary delineating reachable and non-reachable sets. Moreover, our method runs in minutes, even on high-dimensional models. We demonstrate the method by examining two models of the inflammatory response to bacterial infection [20, 26]. In each case, we identify sets of initial conditions that lead to each of 3 biologically relevant outcomes.

The contributions of this paper are as follows:

- An algorithm for computing parameter synthesis in nonlinear dynamical systems. This work builds on and extends formal verification techniques that were first introduced in the context of continuous and hybrid nonlinear dynamical systems [13].
- The results of two studies on two different models of the inflammatory response to bacterial infection. The first model is a 4-equation model, the second is a 17-equation model.

This paper is organized as follows: We outline previous work in reachability for biological systems in Sec. 2. Next, we present our algorithm in Sec. 3. We demonstrate our method on two models of acute inflammation in Sec. 4. We finish by discussing our results and ideas for future work in Sec. 5.

2 Background

Our work falls under the category of *formal verification*, a large area of research which focus on techniques for computing provable guarantees that a system satisfies a given property. Formal verification methods can be characterized by the kind of system they consider (e.g., discrete-time vs continuous-time, finite-state vs continuous-state, linear vs non-linear dynamics, etc), and by the kind of properties they can verify (e.g., reachability – the system can be in a given state, liveness – the system will be in a given set of state infinitely often, etc). The algorithm presented in this paper is intended for verifying reachability properties under parameter uncertainty in nonlinear hybrid systems. The most closely related work in this area uses *symbolic* methods for restricted class of models (e.g., timed automata [4], linear hybrid systems [1, 19, 17]). Symbolic methods for hybrid systems have the advantage that they are exhaustive, but in general only scale to systems of small size (< 10 continuous state variables). Another class of techniques invokes abstractions of the model [2]. Such methods have been applied to biological systems whose dynamics can be described by multi-affine functions. Here, examples include applications to genetic regulatory networks (e.g., [6–8]). Batt and co-workers proposed an approach to verify reachability and liveness properties written in the linear temporal logic (LTL) [24] (LTL can be used to check assumptions about the future such as equilibrium points) of genetic regulatory networks under parameter uncertainty. In that work, the authors show that one can reduce the verification of qualitative properties of genetic regulatory networks to the application of Model Checking techniques [10] on a conservative discrete abstraction. Our method is more general in the sense that we can handle arbitrary nonlinear systems but a limitation is that we cannot handle liveness properties. However, we believe that our algorithm can be extended to handle liveness properties by combining it with a recent technique proposed by Fainekos [15]. We note that there is also work in the area of analyzing piecewise (stochastic) hybrid systems (e.g., [14, 16, 18, 9]). Our method does not handle stochastic models at the present time.

Several techniques relying on numerical computations of the reachable set apply to systems with general nonlinear dynamics ([5, 28, 22]). In [5], the authors presents an hybridization technique, which consists in approximating the system with a piecewise-affine approximation to take advantage of the wider family of methods existing for this class of systems. In [21], the authors reduce the reachability problem to a partial differential equation which they solve numerically. As far as we know, none of these techniques have been

applied successfully to nonlinear systems of more than a few variables. By contrast, our method builds on techniques proposed in [12, 13] which can be applied to significantly larger models.

A more “traditional” tool used for the analysis of nonlinear ODEs is bifurcation analysis, which was applied to the biological models used in our experiments ([26, 20]). Our approach deviates from bifurcation analysis in several ways. First, it is simpler to apply since it only relies on the capacity to compute numerical simulations for the system, avoiding the need of computing equilibrium points or limit cycles. Second, it provides the capacity of analyzing transient behaviors. Finally, when it encounters an ambiguous behavior (e.g., bi-stability) for a given parameter set, it reports that the parameter has uncertain dynamics and can refine the result to make such uncertain sets as small as desired.

3 Algorithm

In this section, we give a mathematical description of the main algorithm used in this work.

3.1 Preliminaries

The set \mathbb{R}^n and the set of $n \times n$ matrices are equipped with the infinite norm, noted $\|\cdot\|$. We define the diameter of a compact set \mathcal{R} to be $\|\mathcal{R}\| = \sup_{(x,x') \in \mathcal{R}^2} \|x - x'\|$. The distance from x to \mathcal{R} is $d(x, \mathcal{R}) = \inf_{y \in \mathcal{R}} \|x - y\|$. The Hausdorff distance between two sets \mathcal{R}_1 and \mathcal{R}_2 is:

$$d_H(\mathcal{R}_1, \mathcal{R}_2) = \max\left(\sup_{x_1 \in \mathcal{R}_1} d(x_1, \mathcal{R}_2), \sup_{x_2 \in \mathcal{R}_2} d(x_2, \mathcal{R}_1)\right).$$

Given a matrix S and a set \mathcal{P} , $S\mathcal{P}$ represents the set $\{Sp, p \in \mathcal{P}\}$. Given two sets \mathcal{R}_1 and \mathcal{R}_2 , $\mathcal{R}_1 \oplus \mathcal{R}_2$ is the Minkowski sum of \mathcal{R}_1 and \mathcal{R}_2 , i.e., $\mathcal{R}_1 \oplus \mathcal{R}_2 = \{x_1 + x_2, x_1 \in \mathcal{R}_1, x_2 \in \mathcal{R}_2\}$.

3.2 Simulation and Sensitivity Analysis

We consider a dynamical system $\mathcal{S}ys = (f, \mathcal{P})$ of the form:

$$\dot{x} = f(t, x, p), \quad p \in \mathcal{P}, \quad (1)$$

where $x \in \mathbb{R}^n$, p is a *parameter vector* and \mathcal{P} is a compact subset of \mathbb{R}^{n_p} . We assume that f is continuously differentiable. Let $\mathcal{T} \subset \mathbb{R}^+$ be a time set. For a given p , a *trajectory* ξ_p is a function of \mathcal{T} which satisfies the ODE (Eq. 1), i.e., for all t in \mathcal{T} , $\dot{\xi}_p(t) = f(t, \xi_p(t), p)$. For convenience, we include the initial state in the parameter vector by assuming that if $p = (p_1, p_2, \dots, p_{n_p})$ then $\xi_p(0) = (p_1(0), p_2(0), \dots, p_n(0))$. Under these conditions, we know by the Cauchy-Lipshitz theorem that the trajectory ξ_p is uniquely defined.

The purpose of sensitivity analysis techniques is to predict the influence on a trajectory of a perturbation of its parameter vector. A first order approximation of this influence can be obtained by a Taylor expansion of $\xi_p(t)$ around p . Let $\delta p \in \mathbb{R}^{n_p}$. We have:

$$\xi_{p+\delta p}(t) = \xi_p(t) + \frac{\partial \xi_p}{\partial p}(t) \delta p + \mathcal{O}(\|\delta p\|^2). \quad (2)$$

The second term in the right hand side of Eq. (2) is the derivative of the trajectory with respect to p . Since p is a vector, this derivative is a matrix, which is called the *sensitivity matrix*. We denote it as: $S_p(t) = \frac{\partial \xi_p}{\partial p}(t)$

The sensitivity matrix can be computed as the solution of a system of ODEs. Let $\mathbf{s}_i = \frac{\partial \xi_p}{\partial p_i}(t)$ be the i^{th} column of S_p . If we apply the chain rule to its time derivative, we get:

$$\begin{cases} \dot{\mathbf{s}}_i(t) = \frac{\partial f}{\partial x}(t, x(t), p) \mathbf{s}_i(t) + \frac{\partial f}{\partial p_i}(t, x(t), p), \\ \mathbf{s}_i(0) = \frac{\partial x(0)}{\partial p_i}. \end{cases} \quad (3)$$

Here $\frac{\partial f}{\partial x}(t, x(t), p)$ is the Jacobian matrix of f at time t . The equation above is thus an affine, time-varying ODE. In the core of our implementation, we compute ξ_p and the sensitivity matrix S_p using the CVODES numerical solver [27], which is designed to solve efficiently and accurately ODEs (like Eq. 1) and sensitivity equations (like Eq. 3).

3.3 Reachable Set Estimation Using Sensitivity

The reachability problem is the problem of computing the set of all the states visited by the trajectories starting from all the possible initial parameters in \mathcal{P} at a given time t .

Definition 1 (Reachable Set). *The reachable set induced by the set of parameters \mathcal{P} at time t is:*

$$\mathcal{R}_t(\mathcal{P}) = \bigcup_{p \in \mathcal{P}} \xi_p(t).$$

The set $\mathcal{R}_t(\mathcal{P})$ can be approximated by using sensitivity analysis. Assume that for a given parameter p in \mathcal{P} we computed a trajectory ξ_p and the sensitivity matrix S_p associated with it. Given another parameter vector p' in \mathcal{P} , we can use this matrix to get an estimate $\hat{\xi}_{p'}^p(t)$ of $\xi_{p'}(t)$. This is done by dropping higher order terms in the Taylor expansion given in Equation 2. We have:

$$\hat{\xi}_{p'}^p(t) = \xi_p(t) + S_p(t)(p' - p). \quad (4)$$

If we extend this estimation to all parameters p' in \mathcal{P} , we get the following estimate of the reachable set $\mathcal{R}_t(\mathcal{P})$:

$$\hat{\mathcal{R}}_t^p(\mathcal{P}) = \bigcup_{p' \in \mathcal{P}} \hat{\xi}_{p'}^p(t) = \{\xi_p(t) - S_p(t)p\} \oplus S_p(t)\mathcal{P}. \quad (5)$$

Thus $\hat{\mathcal{R}}_t^p$ is an affine mapping of the initial set \mathcal{P} into \mathbb{R}^n (see Figure 1).

It can be shown that if the dynamics are affine, i.e., if $f(t, x, p) = A(t, p)x + b(t, p)$, then the estimation is exact. However, in the general case, $\hat{\mathcal{R}}_t^p(\mathcal{P})$ is different from $\mathcal{R}_t(\mathcal{P})$. Since the estimation is based on a first order approximation around parameter p , it is local in the parameter space and its quality depends on how “big” \mathcal{P} is. In order to improve the estimation, we can partition \mathcal{P} into smaller subsets $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_l$ and compute trajectories using new initial parameters p_1, p_2, \dots, p_l to get more precise local estimates. As a practical matter, we need to be able to estimate the benefit of such a refinement. To do so, we compare $\hat{\mathcal{R}}_t^p(\mathcal{P}_j)$ — the estimate we get when using the “global” center, p ; to $\hat{\mathcal{R}}_t^{p_j}(\mathcal{P}_j)$ — the estimate we get when using the “local” center, p_j , and $p'_i \in \mathcal{P}_j$. We do this for each \mathcal{P}_j . Figure 1 illustrates the essential features of the algorithm.

Proposition 1. *We have*

$$d_H(\hat{\mathcal{R}}_t^p(\mathcal{P}_j), \hat{\mathcal{R}}_t^{p_j}(\mathcal{P}_j)) \leq \text{Err}(\mathcal{P}, \mathcal{P}_j), \quad (6)$$

where

$$\text{Err}(\mathcal{P}, \mathcal{P}_j) = \|\xi_{p_j}(t) - \hat{\xi}_{p_j}^p(t)\| + \|S_{p_j}(t) - S_p(t)\| \|\mathcal{P}_j\|. \quad (7)$$

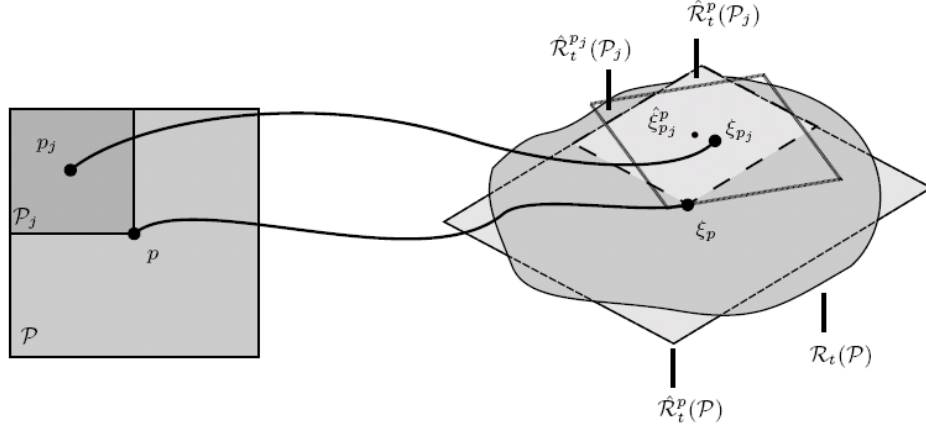


Fig. 1. Comparison between a “global” and a “local” estimate of the reachable set. The large square on the left hand side represent a region of parameter space, \mathcal{P} . The oval-shaped region on the right hand side corresponds to the true reachable set, $\mathcal{R}_t(\mathcal{P})$, induced by parameters \mathcal{P} at time t . The large parallelogram on the right hand side corresponds to the *estimated* reachable set, $\hat{\mathcal{R}}_t^p(\mathcal{P})$, using a sensitivity analysis based on trajectory labeled ξ_p which starts at point $p \in \mathcal{P}$. The point labeled $\hat{\xi}_{p_j}^p$, for example, is an estimate of where a trajectory starting at point p_j would reach at time t . If we partition \mathcal{P} and consider some particular partition, \mathcal{P}_j , we can then compare the estimated reachable sets $\hat{\mathcal{R}}_t^p(\mathcal{P}_j)$ and $\hat{\mathcal{R}}_t^{p_j}(\mathcal{P}_j)$, which correspond to the small light-gray and small dark gray parallelograms, respectively. We continue to refine until the distance between $\hat{\mathcal{R}}_t^p(\mathcal{P}_j)$ and $\hat{\mathcal{R}}_t^{p_j}(\mathcal{P}_j)$ (Eq. 7) falls below some user-specified tolerance.

In other words, the difference between the global and the local estimate can be decomposed into the error introduced in the estimation $\hat{\xi}_{p_j}^p(t)$ of the state reached at time t using p_j (first term on RHS of Eq. 7), and another term involving the difference between the local and the global sensitivity matrices and the distance from local center (second term on RHS of Eq. 7).

Proof. let y be in $\hat{\mathcal{R}}_t^p(\mathcal{P}_j)$. There exists p_y in \mathcal{P}_j such that $y = \hat{\xi}_{p_y}^p(t)$. We need to compare

$$\hat{\xi}_{p_y}^p(t) = \xi_p(t) + S_p(t)(p_y - p) \quad (8)$$

with

$$\hat{\xi}_{p_y}^{p_j}(t) = \xi_{p_j}(t) + S_{p_j}(t)(p_y - p_j). \quad (9)$$

By introducing

$$\hat{\xi}_{p_j}^p(t) = \xi_p(t) + S_p(t)(p_j - p) \quad (10)$$

and after some algebraic manipulations of (8), (9), and (10), we get

$$\begin{aligned} \hat{\xi}_{p_y}^p(t) - \hat{\xi}_{p_y}^{p_j}(t) &= \xi_{p_j}(t) - \hat{\xi}_{p_j}^p(t) + (S_{p_j}(t) - S_p(t))(p_y - p_j) \\ &\leq \|\xi_{p_j}(t) - \hat{\xi}_{p_j}^p(t)\| + \|S_{p_j}(t) - S_p(t)\| \|\mathcal{P}_j\| = Err(\mathcal{P}, \mathcal{P}_j). \end{aligned} \quad (11)$$

Let $x = \hat{\xi}_{p_y}^{p_j}(t)$ which is in $\hat{\mathcal{R}}_t^{p_j}(\mathcal{P}_j)$, then it can be shown that $\|y - x\| \leq Err(\mathcal{P}, \mathcal{P}_j)$ and so $d(y, \hat{\mathcal{R}}_t^{p_j}(\mathcal{P}_j)) \leq Err(\mathcal{P}, \mathcal{P}_j)$. This is true for any $y \in \mathcal{P}_j$, thus

$$\sup_{y \in \hat{\mathcal{R}}_t^p(\mathcal{P}_j)} d(y, \hat{\mathcal{R}}_t^{p_j}(\mathcal{P}_j)) \leq Err(\mathcal{P}, \mathcal{P}_j).$$

Similarly, we can show that

$$\sup_{x \in \hat{\mathcal{R}}_t^{p_j}(\mathcal{P}_j)} d(x, \hat{\mathcal{R}}_t^p(\mathcal{P}_j)) \leq Err(\mathcal{P}, \mathcal{P}_j)$$

which proves the result. \square

The quantity $Err(\mathcal{P}, \mathcal{P}_j)$ can be easily computed from trajectories ξ_p and ξ_{p_j} , their corresponding sensitivity matrices, and $\|\mathcal{P}_j\|$. It has the following properties:

- If the dynamics is affine, then $Err(\mathcal{P}, \mathcal{P}_j) = 0$. Indeed, in this case, we have $\hat{\xi}_{p_j}^p = \xi_{p_j}$ so the first term vanishes and $S_p = S_{p_j}$ so the second term vanishes as well;
- If limit $\|\mathcal{P}\|$ is 0, then limit $Err(\mathcal{P}, \mathcal{P}_j)$ is also 0. Indeed, as $\|\mathcal{P}\|$ decreases, so does $\|p - p_j\|$, and thus $\|\xi_{p_j}(t) - \hat{\xi}_{p_j}^p(t)\|$ and $\|\mathcal{P}_j\|$, since \mathcal{P}_j is a subset of \mathcal{P} . We can show that the convergence is quadratic.

The computation of reachable sets at a given time t can be extended to time intervals. Assume that \mathcal{T} is a time interval of the form $\mathcal{T} = [t_0, t_f]$. The set reachable from \mathcal{P} during \mathcal{T} is $\mathcal{R}_{\mathcal{T}}(\mathcal{P}) = \cup_{t \in \mathcal{T}} \mathcal{R}_t(\mathcal{P})$. It can be approximated by simple interpolation between $\mathcal{R}_{t_0}(\mathcal{P})$ and $\mathcal{R}_{t_f}(\mathcal{P})$. Of course, it may be necessary to subdivide \mathcal{T} into smaller intervals to improve the precision of the interpolation. A reasonable choice for this subdivision is to use the time steps taken by the numerical solver to compute the solution of the ODE and the sensitivity matrices.

3.4 Parameter Synthesis Algorithm

In this section, we state a parameter synthesis problem and propose an algorithm that provides an approximate solution. Let \mathcal{F} be a set of "bad" states. Our goal is to partition the set \mathcal{P} into safe and bad parameters. That is, we want to partition the parameters into those that induce trajectories that intersect \mathcal{F} during some time interval \mathcal{T} , and those that do not.

Definition 2. A solution of the parameter synthesis problem $(Sys = (f, \mathcal{P}), \mathcal{F}, \mathcal{T})$ where \mathcal{F} is a set states and \mathcal{T} a subset of $\mathbb{R}_{\geq 0}$, is a partition $\mathcal{P}_{bad} \cup \mathcal{P}_{saf}$ of \mathcal{P} such that for all $p \in \mathcal{P}_{bad}$ (resp. $p \in \mathcal{P}_{saf}$), $\xi_p(t) \cap \mathcal{F} \neq \emptyset$ (resp. $\xi_p(t) \cap \mathcal{F} = \emptyset$) for all $t \in \mathcal{T}$. An approximate solution is a partition $\mathcal{P} = \mathcal{P}_{saf} \cup \mathcal{P}_{unc} \cup \mathcal{P}_{bad}$ where \mathcal{P}_{saf} and \mathcal{P}_{bad} are defined as before and \mathcal{P}_{unc} (i.e., uncertain) may contain both safe and bad parameters.

Exact solutions cannot be obtained in general, but we can try to compute an approximate solution with the uncertain subset being as small as possible. The idea is to iteratively refine \mathcal{P} and to classify the subsets into the three categories. A subset \mathcal{P}_j qualifies as safe (resp. bad) if:

1. $\hat{\mathcal{R}}_{\mathcal{T}}^{p_j}(\mathcal{P}_j)$ is a reliable estimation of $\mathcal{R}_{\mathcal{T}}(\mathcal{P}_j)$ based on the Err function;
2. $\hat{\mathcal{R}}_{\mathcal{T}}^{p_j}(\mathcal{P}_j)$ does not (resp. does) intersect with \mathcal{F} .

To guarantee that the process ends, we need to ensure that each refinement introduces only subsets that are strictly smaller than the refined set.

Definition 3 (Refining Partition). A refining partition of a set \mathcal{P} is a finite set of sets $\{\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_l\}$ such that

- $\mathcal{P} = \bigcup_{j=1}^l \mathcal{P}_j$;
- There exists $\gamma < 1$ such that $\max_{j \in \{1, \dots, l\}} \|\mathcal{P}_j\| \leq \gamma \|\mathcal{P}\|$.

Let ρ be a function that maps a set to one of its refining partitions. Our algorithm stops whenever the uncertain partition is empty, or it contains only subsets with a diameter smaller than some user-specified value, δp . The complete algorithm is given by Algorithm 1 below.

Algorithm 1 Parameter Synthesis Algorithm

procedure SAFE($\mathcal{P}, \mathcal{F}, t, \delta p, Tol$)

 $\mathcal{P}_{\text{saf}} = \mathcal{P}_{\text{bad}} = \emptyset, \mathcal{P}_{\text{unc}} = \{\mathcal{P}\}$
repeat

 Pick and remove \mathcal{Q} from \mathcal{P}_{unc} and let $q \in \mathcal{Q}$
for each $(q_j, \mathcal{Q}_j) \in \rho(\mathcal{Q})$ **do**
if $Err(\mathcal{Q}, \mathcal{Q}_j) \leq Tol$ **then**
if $\hat{\mathcal{R}}_T^q(\mathcal{Q}_j) \cap \mathcal{F} = \emptyset$ **then**
 $\mathcal{P}_{\text{saf}} = \mathcal{P}_{\text{saf}} \cup \mathcal{Q}_j$
else if $\hat{\mathcal{R}}_T^q(\mathcal{Q}_j) \subset \mathcal{F}$ **then**
 $\mathcal{P}_{\text{bad}} = \mathcal{P}_{\text{bad}} \cup \mathcal{Q}_j$
else
 $\mathcal{P}_{\text{unc}} = \mathcal{P}_{\text{unc}} \cup \{(q_j, \mathcal{Q}_j)\}$
end if
else
 $\mathcal{P}_{\text{unc}} = \mathcal{P}_{\text{unc}} \cup \{(q_j, \mathcal{Q}_j)\}$
end if
end for
until $\mathcal{P}_{\text{unc}} = \emptyset$ or $\max_{P_j \in \mathcal{P}_{\text{unc}}} \|\mathcal{P}_j\| \leq \delta p$
return $\mathcal{P}_{\text{saf}}, \mathcal{P}_{\text{unc}}, \mathcal{P}_{\text{bad}}$
end procedure

 \triangleright Reach set estimation is reliable

 \triangleright Reach set away from \mathcal{F}
 \triangleright Reach set inside \mathcal{F}
 \triangleright Some intersection with the bad set

 \triangleright Reach set estimation not enough precise

The algorithm has been implemented within the Matlab toolbox Breach [12], which combines Matlab routines to manipulate partitions with the CVODES numerical solver, which can efficiently compute ODEs solutions with sensitivity matrices.

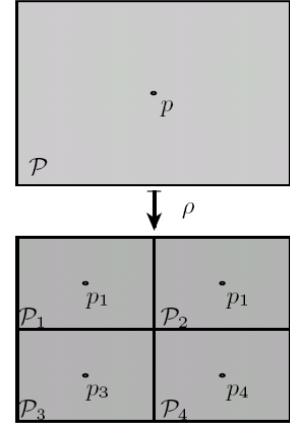
It uses rectangular partitions of the form

$$\mathcal{P}(p, \epsilon) = \{p' : p - \epsilon \leq p' \leq p + \epsilon\}$$

The refinement operator ρ is such that

$$\rho(\mathcal{P}(p, \epsilon)) = \{\mathcal{P}(p^1, \epsilon^1), \mathcal{P}(p^2, \epsilon^2), \dots, \mathcal{P}(p^l, \epsilon^l)\},$$

with $\epsilon^k = \epsilon/2$ and $p^k = p + (\pm \frac{\epsilon_1}{2}, \pm \frac{\epsilon_2}{2}, \dots, \pm \frac{\epsilon_n}{2})$. This operation is illustrated in the Figure on the right.



4 Application to Models of Acute Inflammation

We applied our method to two models of the acute inflammatory response to infection. The first is the 4-equation, 22-parameter model presented in [26], and the second is the 17-equation, 79-parameter model presented in [20]. The primary difference between these models is one of detail, and the first model can be thought of as a reduced dimensional version of the second.

The acute inflammatory response to infection has evolved to promote healing by ridding the organism of the pathogen. The actual response is a complex and carefully regulated combination of molecular and cellular cascades that exhibit both pro and anti-inflammatory behaviors. The pro-inflammatory elements are primarily responsible for eliminating the pathogen, but bacterial killing can cause collateral tissue damage.

Tissue damage, in turn, triggers an escalation in the pro-inflammatory response creating a positive feedback cycle (Figure 2). The anti-inflammatory elements counteract this cycle, thereby minimizing tissue damage and promoting healing. However, in cases of extreme infection, the delicate balance between pro and anti-inflammatory elements is destroyed, resulting in a potentially lethal amount of tissue damage.

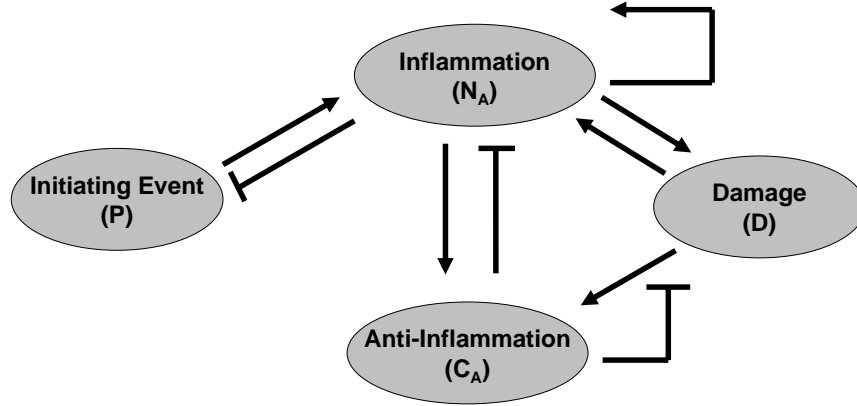


Fig. 2. Cartoon representation of the 4-equation model of the acute immune response. Arrows represent up-regulation, bars represent down-regulation. Figure is adapted from Figure 1 in [26].

The 4-equation model is as follows:

$$\begin{aligned}\frac{dP}{dt} &= k_{pg}P\left(1 - \frac{P}{p_\infty}\right) - \frac{k_{pm}s_mP}{\mu_m + k_{mp}P} - k_{pm}f(N_A)P, \\ \frac{dN_A}{dt} &= \frac{s_{nr}R}{\mu_{nr} + R} - \mu_n N_A, \\ \frac{dD}{dt} &= k_{dn}f_s(f(N_A)) - \mu_d D, \\ \frac{dC_A}{dt} &= s_c + \frac{k_{cn}f(N_A + k_{cmd}D)}{1 + f(N_A + k_{cmd}D)} - \mu_c C_A,\end{aligned}$$

where

$$R = f(k_{nn}N_A + k_{np}P + k_{nd}D), \quad f(V) = \frac{V}{(1 + (C_A/c_\infty)^2)} \quad \text{and} \quad f_s(V) = \frac{V^6}{x_{dn}^6 + V^6}.$$

Here, k_* , μ_* , s_* , p_* are parameters, as defined in [26]. The state variables P , N_A , D , and C_A , correspond to the amounts of pathogen, pro-inflammatory mediators (e.g., activated neutrophils), tissue damage, and anti-inflammatory mediators (e.g., cortisol and interleukin-10), respectively. The 17-equation model, naturally, is far more detailed in terms of which mediators are modeled.

In each model, there are 3 clinically relevant outcomes: (i) a return to health, (ii) aseptic death, and (iii) septic death. Death is defined as a sustained amount of tissue damage (D) above a specified threshold value and constitutes the undesirable or “bad” outcome we wish to avoid. Aseptic and septic death are

distinguished by whether the pathogen (P) is cleared below a specified threshold value. Let \mathcal{F}_{alive} (resp. \mathcal{F}_{dead}) refer to the set of states such that D is below (resp. above) some threshold D_{death} , and let \mathcal{F}_{septic} (resp. $\mathcal{F}_{aseptic}$) refer to the set of states such that P is above (resp. below) some threshold P_{septic} . We can now define three sets of states corresponding to the three clinically relevant outcomes as follows: (i) $Health = \mathcal{F}_{alive} \cap \mathcal{F}_{aseptic}$; (ii) $Aseptic\ death = \mathcal{F}_{dead} \cap \mathcal{F}_{aseptic}$; and (iii) $Septic\ death = \mathcal{F}_{dead} \cap \mathcal{F}_{septic}$. In Figure 3 we present sample traces for both the 4 and 17 equation models.

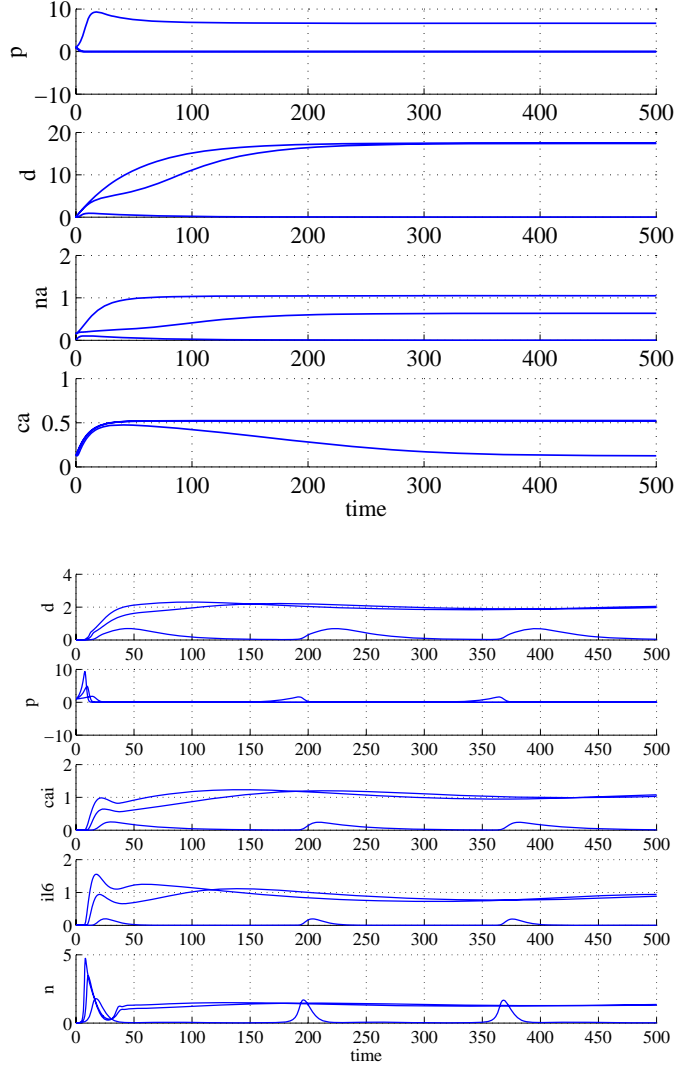


Fig. 3. (Top) Examples trace from the 4-equation model. There are three different traces corresponding to septic death, aseptic death and health. (Bottom) Example traces from the 17-equation model; 5 of the 17 variables are shown. There are also three traces, illustrating the richer dynamics of the model. Two traces corresponds to aseptic death and the third to health with a periodic small resurgence of the pathogen. Time is measured in hours.

4.1 Experiments

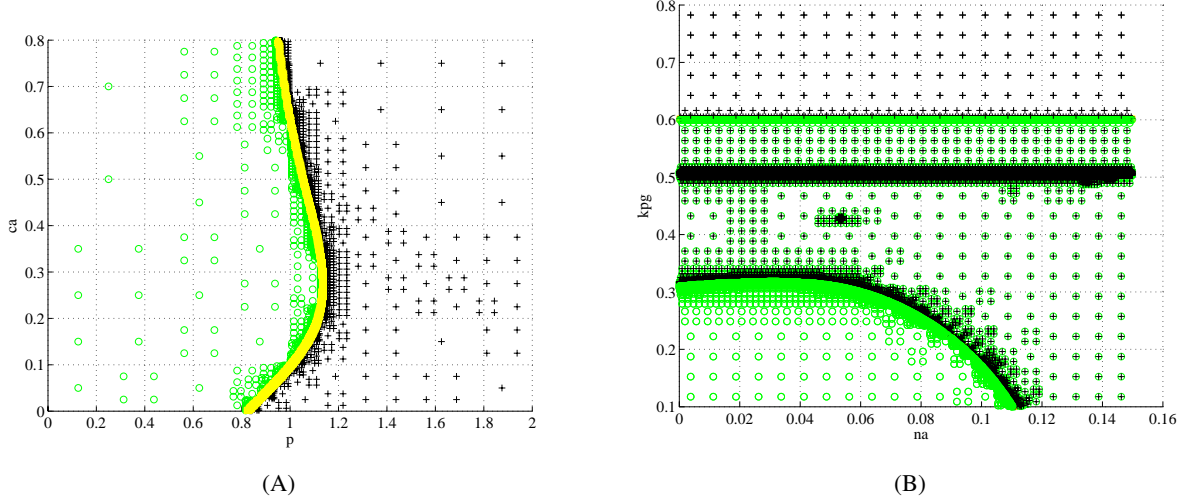


Fig. 4. Results obtained for the 4-equation model. Figure (A) reproduces results presented in Figure 8 of [26] with $k_{pg} = 0.3$, which was obtained using classical bifurcation analysis. Circles are parameter values leading to Health while crosses represent values leading to Death. Figure (B) illustrates how a pair of parameters (N_A and k_{pg}) can be partitioned into the three possible outcomes. Circles alone lead to Health, crosses and circles lead to Aseptic Death and crosses alone lead to Septic Death. The separation between regions is induced from small uncertain regions computed by the algorithm.

We performed several experiments. In the first experiment, we validated our method by reproducing results previously obtained in [26] using bifurcation analysis. Figure 4-A contrasts the initial amount of pathogen, P_0 , and the initial amount of anti-inflammatory mediators, C_{A0} . The growth rate of pathogen, k_{pg} , was set to 0.3 and other parameters to their nominal values. The region \mathcal{F}_{death} given by $D \geq 5$ was used in our algorithm and we checked the intersection with reachable set at time 300 hours. Crosses correspond to initial values leading to death while circles lead to a healthy outcome. We can see that the resulting partition is quantitatively consistent with Figure 8 in [26]. In our second experiment, we varied growth rate of pathogen, k_{pg} , and N_A . Figure 4-(B) shows that there are three distinct regions in the k_{pg} - N_A plane, corresponding to the three clinical outcomes.

We then performed several experimentations with the 17-equation model. Figures 5 (A) and (B) depict the k_{pg} - N_A and k_{pg} - C_{AI} planes, respectively. C_{AI} is a generic anti-inflammatory mediator. We partitioned the region using $\mathcal{F}_{death} = D > 1.5$ and checked after time 300 hours. The 17-equation model exhibits an interesting behavior in the k_{pg} - C_{AI} plane. Namely, that the separation between health and death is not monotone in the growth rate of the pathogen.

As previously mentioned, our algorithm is implemented in Matlab and uses the CVODES numerical solver. Figures 4 (A) and (B) were generated in a few seconds, and Figures 5 (A) and (B) were generated in about an hour on a standard laptop (Intel Dual Core 1.8GHz with 2 Gb of memory). We note that the algorithm could easily be parallelized.

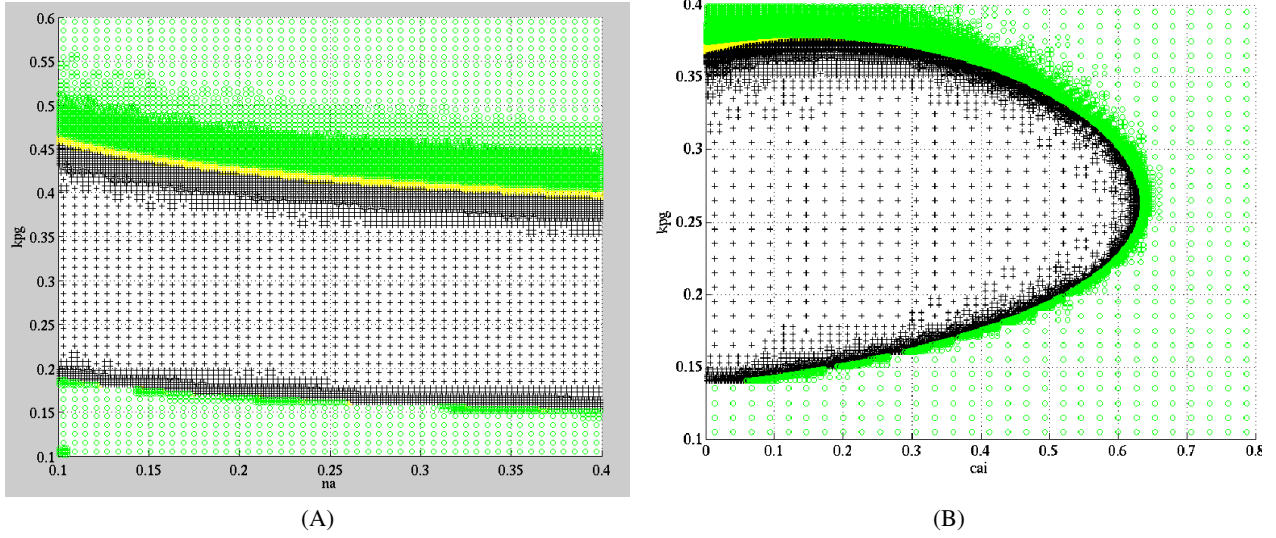


Fig. 5. Results for the 17-equation model. Figure (A) illustrates the k_{pg} - N_A plane, partitioned into regions leading to death (here aseptic death, represented by crosses) and regions leading to health (represented by circles). Figure (B) illustrates the k_{pg} - C_{AI} plane. Interestingly, the separation is not monotone with the growth of pathogen k_{pg} .

5 Discussion and Conclusions

Complex models are increasingly being used to make predictions about complex phenomena in biology and medicine (e.g., [3, 25]). Such models can be potentially very useful in guiding early decisions regarding intervention, but it is often impossible to obtain accurate estimates for every parameter. Thus, it is important to have tools for explicitly examining a range of possible parameters to determine whether the behavior of the model is sensitive to those parameters that are poorly estimated. Performing this task for nonlinear models is especially challenging. We have presented an algorithm for solving the parameter synthesis problem for nonlinear dynamical models and applied it on two models of acute inflammation from the literature. The larger of the two has 17 equations and 79 parameters, demonstrating the scalability of our approach.

Our approach has several limitations. First, our refinement process implies that the number of partitions increases exponentially with the number of varying parameters. Thus, in practice, some variables must be held fixed while analyzing the behavior of the model. On the other hand, the number of state variables is not a limiting factor, as illustrated in our experiments on the 17-equation model. Second, our method relies on numerical simulations. Numerical methods are fundamentally limited in the context of verification since numerical image computation is not semi-decidable for nonlinear differential equations [23]. Moreover, there are no known methods capable of providing provable bounds on numerical errors for general nonlinear differential equations. Thus, we cannot claim to provide *formal* guarantees on the correctness of the results computed by our method. However *asymptotic* guarantees exist, meaning that results can always be improved by decreasing tolerance factors in the numerical computations. A nice feature of our approach is that it enables one to obtain qualitative results using a few simulations (e.g. a coarse partition between regions leading to qualitatively different behaviors). These qualitative results can then be made as precise as desired by focusing on smaller partitions.

There are several areas for future research. Our first order error control mechanism can be improved to make the refinements more efficient and more adaptive when nonlinear (i.e. higher order) behaviors dominate any linear dependence on parameter variations. We are also interested in developing techniques for verifying properties that are more complex than the reachability predicates we considered in this paper. For instance, temporal properties could easily be introduced in our framework. The extended system could then be used, for example, verify the possible outcomes associated with a particular medical intervention. Finally, we believe that the method could easily be used in the context of personalized medicine. In particular, given individual or longitudinal measurements from a specific patient, we could define a reachable set, \mathcal{F}_{obs} , that includes these observations (possibly convolved with a model of the measurement errors). We could then use our method to identify the set of parameters that are consistent with the observations. The refined parameters can then be used to make patient-specific predictions.

Acknowledgments

This work is supported in part by US Department of Energy Career Award and a grant from Microsoft Research to CJL.

References

1. R. Alur, C. Courcoubetis, N. Halbwachs, T. A. Henzinger, P. Ho, X. Nicollin, A. Olivero, J. Sifakis, and S. Yovine. The algorithmic analysis of hybrid systems. *Theoretical Computer Science*, 138(1):3–34, 1995.
2. R. Alur, T. Henzinger, G. Lafferriere, and G. Pappas. Discrete abstractions of hybrid systems. volume 88(7), pages 971–984. IEEE, 2000.
3. G. An. Agent based computer simulation and sirs: building a gap between basic science and clinical trials. *Shock*, 16:266273, 2001.
4. A. Annichini, E. Asarin, and A. Bouajjani. Symbolic techniques for parametric reasoning about counter and clock systems. In *CAV*, volume 1855 of *Lecture Notes in Computer Science*, pages 419–434. Springer, 2000.
5. E. Asarin, T. Dang, and A. Girard. Hybridization methods for verification of non-linear systems. In *ECC-CDC’05 joint conference: Conference on Decision and Control CDC and European Control Conference ECC*, 2005.
6. G. Batt, C. Belta, and R. Weiss. Model checking genetic regulatory networks with parameter uncertainty. In *HSCC*, volume 4416 of *Lecture Notes in Computer Science*, pages 61–75. Springer, 2007.
7. G. Batt, C. Belta, and R. Weiss. Model checking liveness properties of genetic regulatory networks. In *TACAS*, volume 4424 of *Lecture Notes in Computer Science*, pages 323–338. Springer, 2007.
8. G. Batt, D. Ropers, H. de Jong, J. Geiselmann, R. Mateescu, M. Page, and D. Schneider. Analysis and verification of qualitative models of genetic regulatory networks: A model-checking approach. In *IJCAI*, pages 370–375, 2005.
9. E. Cinquemani, A. Miliadis-Argeitis, and J. Lygeros. Identification of genetic regulatory networks: A stochastic hybrid approach. In *IFAC World Congress*, 2008.
10. E. Clarke, O. Grumberg, and D. Peled. *Model Checking*. MIT Press, 1999.
11. A. Donz, B. Krogh, and A. Rajhans. Parameter synthesis for hybrid systems with an application to simulink models. In *Proceedings of the 12th International Conference on Hybrid Systems: Computation and Control (HSCC’09)*, LNCS. Springer-Verlag, April 2009.
12. A. Donzé. *Trajectory-Based Verification and Controller Synthesis for Continuous and Hybrid Systems*. PhD thesis, University Joseph Fourier, June 2007.
13. A. Donzé and O. Maler. Systematic simulations using sensitivity analysis. In *HSCC’07*, LNCS, April 2007.
14. S. Drulhe, G. Ferrari-Trecate, H. de Jong, and A. Viari. Reconstruction of switching thresholds in piecewise-affine models of genetic regulatory networks. In *HSCC*, volume 3927 of *Lecture Notes in Computer Science*, pages 184–199. Springer, 2006.
15. G. E. Fainekos and G. J. Pappas. Robust sampling for mtl specifications. In *FORMATS*, Lecture Notes in Computer Science, pages 147–162. Springer, 2007.
16. E. Fargot and J.-L. Gouze. How to control a biological switch: a mathematical framework for the control of piecewise affine models of gene networks. Technical report, INRIA Sophia Antipolis, 2006.
17. G. Frehse, S. K. Jha, and B. H. Krogh. A counterexample-guided approach to parameter synthesis for linear hybrid automata. In *HSCC*, volume 4981, pages 187–200, 2008.

18. R. Ghosh and C. Tomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-notch signalling. *System Biology*, 2004.
19. T. A. Henzinger, B. Horowitz, R. Majumdar, and H. Wong-Toi. Beyond hytech: Hybrid systems analysis using interval numerical methods. In *HSCC*, volume 1790 of *Lecture Notes in Computer Science*, pages 130–144. Springer, 2000.
20. R. Kumar. *The Dynamics of Acute Inflammation*. PhD thesis, University of Pittsburgh, 2004.
21. I. Mitchell and C. Tomlin. Level set methods for computation in hybrid systems. In *HSCC*, pages 310–323, 2000.
22. I. M. Mitchell and C. J. Tomlin. Overapproximating reachable sets by hamilton-jacobi projections. *J. Symbolic Computation*, 19:1–3, 2002.
23. A. Platzer and E. M. Clarke. The image computation problem in hybrid systems model checking. In A. Bemporad, A. Bicchi, and G. Buttazzo, editors, *Hybrid Systems: Computation and Control, 10th International Conference, HSCC 2007, Pisa, Italy, Proceedings*, volume 4416 of *LNCS*, pages 473–486. Springer-Verlag, 2007.
24. A. Pnueli. The temporal logic of programs. In *Proc. 18th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 46–57, 1977.
25. D. Polidori and J. Trimmer. Bringing advanced therapies to marker faster: a role for bio-simulation? *Diabetes Voice*, 48(2):28–30, 2003.
26. A. Reynolds, J. Rubin, G. Clermont, J. Day, Y. Vodovotz, and B. Ermentrout. A reduced mathematical model of the acute inflammatory response: I. derivation of model and analysis of anti-inflammation. *J Theor Biol*, 242(1):220–236, 2006.
27. R. Serban and A. C. Hindmarsh. Cvodes: the sensitivity-enabled ode solver in sundials. In *Proceedings of IDETC/CIE 2005*, Long Beach, CA., Sept. 2005.
28. O. Stursberg and B. H. Krogh. Efficient representation and computation of reachable sets for hybrid systems. In *HSCC*, pages 482–497, 2003.

Biological Populations Obeying Difference Equations: Stable Points, Stable Cycles, and Chaos

ROBERT M. MAY

Biology Department, Princeton University, Princeton, N.J. 08540, U.S.A.

(Received 28 June 1974)

For biological populations with nonoverlapping generations, population growth takes place in discrete time steps and is described by difference equations. Some of the simplest such nonlinear difference equations can exhibit a remarkable spectrum of dynamical behavior, from stable equilibrium points, to stable cyclic oscillations between two population points, to stable cycles with four points, then eight, 16, etc., points, through to a chaotic regime in which (depending on the initial population value) cycles of *any* period, or even totally aperiodic but bounded population fluctuations, can occur. This rich dynamical structure is overlooked in conventional linearized stability analyses; its existence in the simplest and fully deterministic nonlinear (“density dependent”) difference equations is a fact of considerable mathematical and ecological interest.

1. Introduction

In some biological situations (such as man), population growth is a continuous process and generations overlap; the appropriate mathematical description involves nonlinear differential equations. In other biological situations (such as 13 year periodical cicadas), population growth takes place at discrete intervals of time and generations are completely nonoverlapping; the appropriate mathematical description is in terms of nonlinear difference equations. For a single species, the simplest such differential equations, with no time-delays, lead to very simple dynamics: a familiar example is the logistic, $dN/dt = rN(1 - N/K)$, with a globally stable equilibrium point at $N = K$ for all $r > 0$. But the corresponding simplest difference equations, with their built-in time lag in the operation of regulatory mechanisms, can have a complicated dynamical structure, the great richness of which is not commonly appreciated either in the ecological literature, or in elementary mathematical discussions of difference equations.

For a single species, the difference equations arising in population biology are usually discussed as having either a stable equilibrium point or unstable, growing oscillations. In fact, some of the most elementary of these nonlinear

difference equations exhibit a spectrum of dynamical behavior, which, as the intrinsic growth rate r increases, goes from a stable equilibrium point, to stable cyclic oscillations between two population points, to stable cycles with four points, then eight points, and so on, through to a regime which can only be described as "chaotic" (an apt term coined by Li & Yorke, 1974). For any given value of r , in this chaotic regime there are cycles of period, 2, 3, 4, 5, ..., n , ..., where n is any positive integer, along with an uncountable number of initial points for which the system does not eventually settle into any finite cycle; whether the system converges on a cycle, and, if so, which cycle, depends on the initial population point (and some of the cycles may be attained only from infinitely unlikely initial points). Figure 1 aims to illustrate this range of behavior.

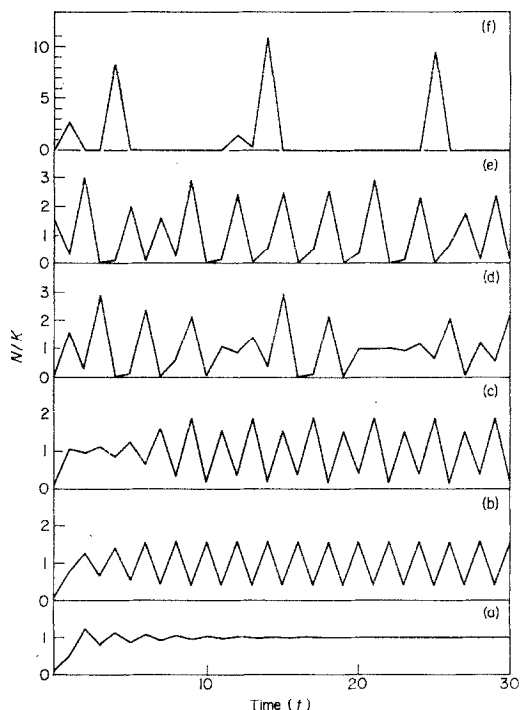


FIG. 1. Spectrum of dynamical behavior of the population density, N_t/K , as a function of time, t , as described by the difference equation (1) for various values of r . Specifically: (a) $r = 1.8$, stable equilibrium point; (b) $r = 2.3$, stable two-point cycle; (c) $r = 2.6$, stable four-point cycle; (d), (e), (f) are in the chaotic regime, where the detailed character of the solution depends on the initial population value, with (d) $r = 3.3$ ($N_0/K = 0.075$) (e) $r = 3.3$ ($N_0/K = 1.5$), (f) $r = 5.0$ ($N_0/K = 0.02$).

Specifically, consider the simple nonlinear equation

$$N_{t+1} = N_t \exp [r(1 - N_t/K)]. \quad (1)$$

This is considered by some people (Macfadyen, 1963; Cooke, 1965) to be the difference equation analogue of the logistic differential equation, with r and K the usual growth rate and carrying capacity, respectively. The stability character of this equation, as a function of increasing r , is set out in Table 1, and illustrated by Fig. 1.

TABLE 1

Dynamics of a population described by the difference equation (1)

Dynamical behavior	Value of growth rate, r	Illustration
Globally stable equilibrium point	$2 > r > 0$	Fig. 1(a)
Globally stable two-point cycle	$2.526 > r > 2$	Fig. 1(b)
Globally stable four-point cycle	$2.656 > r > 2.526$	Fig. 1(c)
Stable cycle, period 8, giving way in turn to cycles of period 16, 32, etc. as r increases	$2.692 > r > 2.656$	
Chaos (cycles of arbitrary period, or aperiodic behavior, depending on initial condition)	$r > 2.692$	Fig. 1(d), (e), (f)

Another example is

$$N_{t+1} = N_t[1 + r(1 - N_t/K)] \quad (2a)$$

or, equivalently, defining $x = (r/1+r)(N/K)$,

$$x_{t+1} = (1+r)x_t(1-x_t). \quad (2b)$$

In the form (2b), this is probably the simplest nonlinear difference equation one could write down. Although discussed by various people (Maynard Smith, 1968; May, 1972; Krebs, 1972; Scudo & Leviné, 1974) as the analogue of the logistic differential equation, equation (2) is less satisfactory than (1) by virtue of its unbiological feature that the population can become negative if at any point N_t exceeds $K(1+r)/r$. Thus stability properties here refer to stability within some specific neighborhood, unlike equation (1) where, for example, the stable equilibrium point at $N = K$ is globally stable (for all $N > 0$) for $2 > r > 0$. With this proviso, the stability behavior of equation (2) is strikingly similar to that of equation (1): see Table 2.

The one-parameter difference equations (1) and (2) are treated in detail to give specificity to the discussion. It is to be emphasized, however, that the

TABLE 2

Dynamics of a population described by the difference equation (2)

Dynamical behavior	Value of growth rate, r
Stable equilibrium point	$2 > r > 0$
Stable two-point cycle	$2.449 > r > 2$
Stable four-point cycle	$2.544 > r > 2.449$
Stable cycles, period 8, then 16, 32, etc.	$2.570 > r > 2.544$
Chaos	$r > 2.570$

phenomenon of a threefold regime of a stable point, giving way to stable cycles of period 2^n , giving way to chaotic behavior, is a generic one which is liable to occur in any model for discrete generations with the possibility of strongly density-dependent population growth. Some other simple difference equations, which are mainly culled from the entomological literature and which exhibit the phenomenon, are as follows. (i) The equation

$$N_{t+1} = \lambda[1 + aN_t]^{-b}N_t$$

has been used by Hassell (1974) to provide a two-parameter fit to a wide range of field and laboratory data on single-species population growth: for relatively small values of b or of λ there is a globally stable point; the conjunction of moderate values of b and λ produces stable cycles; relatively large values of both b and λ leads to chaos. (ii) The density dependent form

$$N_{t+1} = \left[\lambda_1 + \frac{\lambda_2}{1 + \exp\{A(N_t - B)\}} \right] N_t$$

discussed by Pennycuik, Compton & Beckingham (1968), by Usher (1972), and (in a limiting step-function form) by Williamson (1974), can also exhibit all three regimes as the two parameters A and B are varied. (iii) Similarly the class of models

$$N_{t+1} = \frac{\lambda N_t}{1 + (aN_t)^b},$$

the possible stable points of which have been discussed by Maynard Smith (1974), can show all three types of behavior as λ and b vary. (iv) The density dependent equation

$$\begin{aligned} N_{t+1} &= \lambda N_t && [\text{if } N_t < B] \\ N_{t+1} &= \lambda(N_t/B)^{-b}N_t && [\text{if } N_t > B], \end{aligned}$$

discussed by Varley, Gradwell & Hassell (1973), is an interesting example. Here, as a consequence of the pathological discontinuity, the stable point regime ($0 < b < 2$) gives way directly to the chaotic regime ($b > 2$), with no

intervening regime of stable cycles. (v) In all the above examples, the parameters can take values such that the curve relating N_{t+1} to N_t has a hump. To the contrary, the form

$$N_{t+1} = \frac{\lambda N_t}{1 + aN_t}$$

which is sometimes called *the* logistic difference equation (Skellam, 1952; Leslie, 1957; Utida, 1967; Pielou, 1969), gives a monotonic curve relating N_{t+1} and N_t , and consequently it always leads simply to a globally stable equilibrium point.

That such single species difference equations should describe populations going from stable equilibrium points to stable cycles as r increases is not surprising, in view of the general engineering precept that excessively long time delays in otherwise stabilizing feedback mechanisms can lead to "instability" or, more precisely, to stable limit cycles (see May, 1973, pp. 27-30 and chap. 4; May, Conway, Hassell & Southwood, 1974). What is remarkable, and disturbing, is that the simplest, *purely deterministic*, single species models give essentially arbitrary dynamical behavior once r is big enough [$r > 2.692$ for equation (1), $r > 2.570$ for equation (2)]. Such behavior has previously been noted in a meteorological context (Lorenz, 1963, 1964), and doubtless has other applications elsewhere.

For population biology in general, and for temperate-zone insects in particular, the implication is that even if the natural world was 100% predictable, the dynamics of populations with "density dependent" regulation could nonetheless in some circumstances be indistinguishable from chaos, if the intrinsic growth rate r is large enough.

Section 2 presents the stability analysis for the model (1) at smaller r , up to the regime of chaos; this section contains an explicit Lyapunov function to show the stable equilibrium point (for $2 > r > 0$) is globally stable, and introduces some novel mathematical tricks to study the regime of stable cycles (for $2.692 > r > 2$). Section 3 similarly gives the analysis of the model (2) at smaller r . Section 4 briefly outlines an abstract mathematical theorem, very recently proved by Li & Yorke (1974), which shows that if the system

$$N_{t+1} = N_t f(N_t) \tag{3}$$

has a cycle of period 3, then it also has cycles of period n , where n is any positive integer, so that its behavior is chaotic (in the sense defined above). This general theorem is then applied to the specific equations (1) and (2) to elucidate their behavior at larger r . Section 4 also speculates upon some of the tendencies evidenced by Fig. 1(d), (e), (f), which suggest the need for further general mathematical analysis of such systems. Sections 5 and 6 briefly discuss some other biological and mathematical aspects of the problem.

2. Equation (1): Stable Points and Stable Cycles

We begin by quickly recapitulating the standard linearized analysis for stable equilibrium points of difference equations such as (1), (2) or (3), because these general methods underlie the tricks subsequently introduced in the derivation of stable cycles.

We first find the possible equilibrium points, and then study their stability. Using the general form of equation (3), equilibrium points where $N_{t+1} = N_t = N^*$ are the solutions of

$$f(N^*) = 1. \quad (4)$$

To examine the stability of such an equilibrium point with respect to small perturbations, write $N_t = N^* + x_t$, and linearize about the equilibrium point (neglecting initially small quantities of order x^2) to get an equation for the population perturbation x_t :

$$x_{t+1} = (1 - \mu)x_t. \quad (5)$$

Here, for notational convenience, we have introduced the definition

$$\mu = -N^* \left(\frac{df}{dN} \right)^* = - \left(\frac{d \ln f}{d \ln N} \right)^*. \quad (6)$$

Neighborhood stability clearly requires $|1 - \mu| < 1$, which leads to the criterion

$$2 > \mu > 0. \quad (7)$$

More specifically, if $1 > \mu > 0$ the perturbations are monotonically damped, and if $2 > \mu > 1$ they are damped in an oscillatory manner.

Applied to equation (1), where $f(N) = \exp[r(1 - N/K)]$, equation (4) leads to a unique equilibrium point at

$$N^* = K. \quad (8)$$

Next, equation (6) reduces to $\mu = r$, whence the requirement for this equilibrium point to be a stable one is

$$2 > r > 0. \quad (9)$$

More particularly, perturbations are exponentially damped if $1 > r > 0$, oscillatorily damped if $2 > r > 1$.

The above constitutes a linearized stability analysis. However, in this instance we can construct a nonlinear Lyapunov function; that is, a function V_t with the properties $V_t \geq 0$ and $\Delta V_t \equiv V_{t+1} - V_t \leq 0$. Such a function is

$$V_t = (N_t - K)^2. \quad (10)$$

This clearly has the property $V_t \geq 0$, and for the quantity ΔV_t we have

$\Delta V_t = (N_{t+1} - N_t)(N_{t+1} + N_t - 2K)$ for which it may be shown that

$$\Delta V_t \leq 0 \quad [\text{for all } N_t > 0], \quad (11)$$

if, and only if, $2 > r > 0$. Therefore the linearized stability analysis is a valid characterization of the global, nonlinear stability properties, and the equilibrium point of equation (8) is globally stable for $2 > r > 0$. This is a useful, if special, result.

To study what happens when $r > 2$ it is helpful to have recourse to the apparently novel trick of expressing N_{t+2} as a function of N_t :

$$N_{t+2} = N_t g(N_t), \quad (12)$$

where clearly the function $g(N)$ is defined in terms of the $f(N)$ of the general equation (3) by

$$g(N) = f(N)f(Nf(N)). \quad (13)$$

The analysis outlined three paragraphs above may now be repeated, step by step, using $g(N)$ instead of $f(N)$, to seek stable solutions ($N_{t+2} = N_t = N_{t-2}$, etc.) of the equation (12). Such solutions will lead to stable two-point cycles of the kind depicted in Fig. 1(b).

Applying this trick specifically to equation (1), we have

$$g(N) = \exp \left[r \left(2 - \frac{N}{K} \left\{ \exp \left[r \left(1 - \frac{N}{K} \right) \right] + 1 \right\} \right) \right]. \quad (14)$$

Possible equilibrium solutions, N^* , follow from

$$g(N^*) = 1, \quad (15)$$

that is,

$$2 = (N/K)(\exp [r(1 - N/K)] + 1). \quad (16)$$

By writing

$$N^* \equiv K(1 + y), \quad (17)$$

equation (16) may be manipulated into the form

$$y = \tanh \left(\frac{1}{2} r y \right). \quad (18)$$

A graphical way of solving this transcendental equation is indicated in Fig. 2. The essential point is that if $r < 2$, there is only *one* real solution, namely $y = 0$ ($N^* = K$), corresponding to the globally stable equilibrium point already discovered for $r < 2$. However, for $r > 2$ there are *three* real solutions: the trivial solution $y = 0$, and a pair of solutions $y = \pm y_0$ (with $y_0 < 1$) as indicated in Fig. 2. It may further be shown by the techniques discussed above that the solution $y = 0$ is always unstable for $r > 2$, but that each of the pair of solutions $N^* = K(1 \pm y_0)$ is stable provided

$$2 > r[2 - r(1 - y_0^2)] > 0. \quad (19)$$

The quantities r and y_0 are themselves related by equation (18), whence equation (19) eventually comes down to the constraint $r < 2.526$.

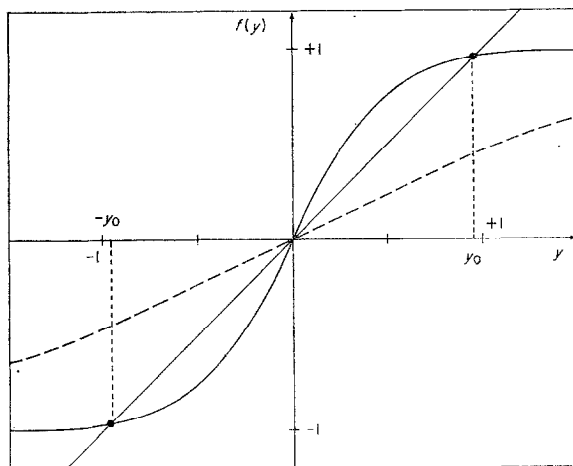


FIG. 2. Graphical solution of equation (18). The straight line depicts the function $f(y) = y$ [the left-hand side of equation (18)]; the dashed curve illustrates the function $\tanh(\frac{1}{2}ry)$ if $r < 2$ (specifically, $r = 1$), so that the slope near the origin is less than y , and here the only solution of equation (18) is at $y = 0$; the solid curved line illustrates $\tanh(\frac{1}{2}ry)$ if $r > 2$ (specifically, $r = 4$), so that the slope near the origin is greater than y , and here there are necessarily three real solutions of equation (18), at $y = 0$ and $y = \pm y_0$.

Thus for $r < 2$ equation (1) has a stable equilibrium point at $N^* = K$, but as r increases beyond 2 this solution becomes unstable, and bifurcates into a pair of points $N^* = K(1 \pm y_0)$, between which the population alternates in a two-point cycle which is stable provided $2 < r < 2.526$. Extensive numerical studies suggest this two-point cycle is globally stable for these values of r .

Beyond $r = 2.526$, the two-point cycle in turn becomes unstable, and each of the points bifurcates into two further points, giving a stable four-point cycle, as illustrated in Fig. 1(c). The details of this process can be elucidated by studying the relationship

$$N_{t+4} = N_t h(N_t) \quad (20)$$

which follows from the general equations (3) and (12) with

$$h(N) = g(N)g(Ng(N)). \quad (21)$$

For equation (1), $h(N)$ then follows from equation (14), and we can compute the equilibrium solutions and their stability, along the lines laid down above. We find that for $r < 2.526$ the only real solutions of $h(N^*) = 1$ are the three points just discussed. However, for larger r not only are all these points unstable, but there emerge four further real solutions, which lead to a stable four-point cycle for $r < 2.656$.

Beyond $r = 2.656$ there lies a stable eight-point cycle, then a stable 16-point cycle, and so on. Notice that as the period of the various cycles increases, the range of r values for which they are stable decreases: the equilibrium point is stable for a band of r values of width two, the two-point cycle is stable in a band of width 0.526, the four-point cycle in a band of width 0.130. In section 4 it will be shown that eventually, for $r > 3.102$, the system has entered the chaotic regime; however, the full details of the transition from the regime of stable cycles (with systematically increasing periods of length 2^n), to the chaotic regime, has so far defied analysis.†

Note that, independent of the details of the dynamics, the population variations must eventually lie between finite upper and lower bounds, N_+ and N_- respectively, for all values of r and for all initial population values. For the upper bound, observe that regardless of the initial value, equation (1) implies the population in the subsequent generation cannot exceed the maximum of the function $f(x) = Kx \exp[r(1-x)]$, which maximum occurs at $x = 1/r$; that is

$$N_+ = \frac{K \exp(r-1)}{r}. \quad (22)$$

Although extreme initial conditions can keep the population low at first, ultimately the smallest possible population is that attained one step after the N_+ of equation (22), and thus has the value

$$N_- = N_+ \exp[r(1 - N_+/K)]. \quad (23)$$

The ratio between these upper and lower bounds provides a measure of the population variations liable to occur once r increases substantially beyond two:

$$N_+/N_- = \exp[\exp(r-1) - r]. \quad (24)$$

The magnitude of this ratio is obviously sensitively dependent on the value of r as r increases.

3. Equation (2): Stable Points and Stable Cycles

Following the recipes outlined in the previous section, the analysis of equation (2) is analogous to that of equation (1).

First observe that the only possible equilibrium point, given by the solution of equation (4), is $N^* = K$. The formula (6) gives $\mu = r$, whence from equation (7) this point is stable if, and only if, $2 > r > 0$. In contrast with the global result obtained in the previous section, this equilibrium point is stable only to perturbations which are not too large. For all $r > 0$, a disturbance to $N_t > K(r+1)/r$ leads in the next time step to a *negative* N_{t+1} , and all subsequent N_{t+k} are necessarily negative, diverging towards $-\infty$. Biologically,

† See note added in proof.

of course, such negative values of N imply extinction; but these features of the model (2) make it in some respects less satisfying than (1).

For $r > 2$, we again turn to study the possibility of stable two-point cycles, using equations (12) and (13): here $g(N)$ may usefully be written in the form

$$g(N) = 1 - r^3 K^{-3} (N - K)(N - N_A)(N - N_B), \quad (25)$$

with the definition

$$N_{A,B} = (K/2r)[r + 2 \pm (r^2 - 4)^{1/2}]. \quad (26)$$

We see that if $r < 2$, there is only one real solution of the equation $g(N^*) = 1$, namely the familiar point $N^* = K$. But once $r > 2$, this solution becomes unstable, and two new solutions of equation (12) split off on either side of it, at $N^* = N_A, N_B$. These two new points will be stable if equation (7) is satisfied, where here

$$\mu = - \left(N \frac{dg}{dN} \right)^* = r^2 - 4. \quad (27)$$

Thus equation (2) will have a stable two-point cycle if, and only if, $\sqrt{6} > r > 2$, as set out in Table 2.

Again, as r increases beyond $\sqrt{6}$, each of these two points bifurcates, to give stable four-point cycles if $2.544 > r > 2.449$; and so on. As for the previous example, the next section shows that eventually a regime of chaos is established for $r > 2.828$. Again, the details of the transition zone where stable cycles of period 2^n merge into the chaotic regime are not yet elucidated.†

4. Three-point Cycles and Chaotic Behavior

Motivated by earlier work of Lorenz (1963, 1964), published in the meteorological literature, Li & Yorke (1974) have very recently proved an abstract mathematical theorem which is relevant to our present discussion of ecological equations such as (1) and (2). Suppose the general difference equation (3) has a three-point cycle, that is a solution such that $N_{t+3} = N_t = N^*$, with $N_{t+1} \neq N_{t+2} \neq N^*$. It then necessarily follows that there are also cycles with period n , where n is any positive integer, and furthermore that there are an uncountable number of initial points N_0 from which the system does not eventually settle into any of these cycles (that is, is not "asymptotically periodic"). In these circumstances, whether the system will converge upon one of the cycles, and, if so, which cycle, depends on the starting point, N_0 .

We now apply this general mathematical theorem to the ecologically interesting equations (1) and (2). First we seek a cyclic solution, with period 3,

† See note added in proof.

for equation (1). For notational convenience, write the three points in such a cycle as $N_1 = aK$, $N_2 = bK$, $N_3 = cK$: then a , b , c (with $a < b < c$) are given by

$$\begin{aligned} b &= a \exp [r(1-a)] \\ c &= b \exp [r(1-b)] \\ a &= c \exp [r(1-c)]. \end{aligned} \quad (28)$$

With the help of the observation that $a+b+c = 3$, it may be seen that a is the smallest solution of the transcendental equation

$$r = \frac{\ln \{3/a - 1 - \exp [r(1-a)]\}}{2 - a - a \exp [r(1-a)]}. \quad (29)$$

Figure 3 illustrates the behavior of these solutions a , b , c as functions of r . For r above r_c (where $r_c = 3.102$) there are two distinct three-point cycles; below $r_c = 3.102$ no such cycles exist.

The dynamical behavior of equation (1) in this chaotic regime, $r > r_c$, is illustrated in Fig. 1(d), (e), (f). Figure 1(d) and (e) are for the same value of r , and differ only in their initial population value. Note that Fig. 1(d) or (e), if looked at only over particular short time intervals, could convey the impres-

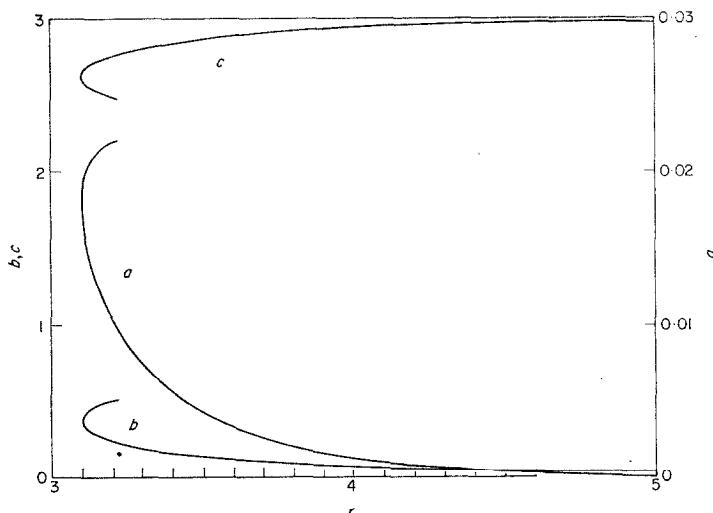


FIG. 3. Values of the three population points $N_1 = aK$, $N_2 = bK$, $N_3 = cK$, with $a < b < c$, in the three-point cyclic solution of equation (1), as functions of r . The curve for a is according to the right-hand scale (from 0.03), and the curves for b and c are according to the left-hand scale (from 0-3). Notice there is no solution for $r < 3.102$.

sion of being locked into a three-point cycle; there is a tendency to be "captured" into almost-periodic three-point cycles, in between episodes of apparently chaotic behavior. A detailed understanding of these properties remains an interesting mathematical problem, related to that of determining what fraction of the totality of initial points converge to a three-point cycle, what fraction to a five-point cycle, and so on, ending with a determination of the fraction of initial points which lead to aperiodic behavior.

For relatively large values of r beyond r_c [e.g., Fig. 1(f)], the population fluctuations become more severe, as indicated earlier by the asymptotic upper and lower limits (22) and (23). Notice, however, that the *mean* population remains around the value K , as follows from the remark that for equation (1)

$$N_{t+j} = N_t \exp \left[r \left(j - \sum_{i=0}^{j-1} N_{t+i}/K \right) \right], \quad (30)$$

that is,

$$\sum_{i=0}^{j-1} N_{t+i} = jK [1 - (1/rj) \ln (N_{t+j}/N_t)]. \quad (31)$$

Thus, for a long time sequence, $j \gg 1$, we have

$$\langle N \rangle \simeq K. \quad (32)$$

As r becomes large, this mean value is increasingly constituted of a few fairly large population values, together with long sequences of very low population values [e.g., Fig. 1(f)]. Very approximately, we may observe from equation (22) that the large fluctuations will have an amplitude around $K[\exp(r-1)]/r$, and will consequently on the average be spaced $(1/r) \exp(r-1)$ time intervals apart. This gives qualitative insight into the numerical results.

It remains to find the value of r which marks the onset of three-point cycles, and consequent chaos, in equation (2). To this end, we use the form (2b), and seek solutions $a < b < c$ such that

$$\begin{aligned} b &= (1+r)a(1-a) \\ c &= (1+r)b(1-b). \\ a &= (1+r)c(1-c) \end{aligned} \quad (33)$$

Numerical computations reveal that such real solutions can be found if, and only if, $r > 2.828$. [Li & Yorke (1974) have already remarked, by way of an example, that this difference equation has three-point cycles once $r \geq 3$.]

5. Multispecies Difference Equations

The above discussion is restricted to single species systems obeying difference equations. Similar considerations are likely to apply, *a fortiori*, to

multispecies situations. Here analytic results are hard to come by; some numerical studies are reported elsewhere (May, 1974a).

6. Discussion

Equations (1) and (2) are two of the simplest nonlinear difference equations to be found. Their rich dynamical structure is a fact of considerable mathematical interest, which deserves to be more widely appreciated.

Previous work in this general area of population biology includes, *inter alia*, remarks on the relation between differential equation models and difference equation models (Van der Vaart, 1973; May, 1972), and on the equivalence between difference equations and differential equations with explicit time-delays (May *et al.*, 1974; Maynard Smith, 1968; McMurtrie, 1974). Earlier discussions of the stability properties of equations (1) and (2) consist of linearized analyses showing the equilibrium point is in both cases only stable for $2 > r > 0$ [Cook, 1965 for (1); Maynard Smith, 1968 for (2)], with larger r usually dismissed as "unstable, with diverging oscillations". Very recently, May (1974b) has noted the stable two- and four-point cycles for these equations when $r > 2$, and Scudo & Levine (1974) have independently noted the two-point cycle behavior in equation (2).

The transition, as r increases beyond r_c , into a regime of apparent chaos, with cycles of essentially arbitrary period possible, is a result with many ecological implications. It could be particularly relevant to temperate insect populations, where the natural description is in terms of nonlinear or "density dependent" difference equations, often with relatively large r . In conclusion, it may be emphasized that without an understanding of the range of behavior latent in deterministic difference equations, one could be hard put to make sense of computer simulations or time-series analyses of such models.

I am indebted to R. E. McMurtrie, G. F. Oster, and a reviewer (J. R. Beddington) for helpful comments. I particularly thank J. A. Yorke for drawing his very elegant work to my attention.

REFERENCES

- COOK, L. M. (1965). *Nature, Lond.* **207**, 316.
 HASSELL, M. P. (1974). *J. Anim. Ecol.* (in press).
 KREBS, J. R. (1972). *Ecology: The Experimental Analysis of Distribution and Abundance*. New York: Harper and Row.
 LESLIE, P. H. (1957). *Biometrika* **44**, 314.
 LI, T.-Y. & YORKE, J. A. (1974). *SIAM J. Math. Anal.* (in press).
 LORENZ, E. N. (1963). *J. Atmos. Sci.* **20**, 448.
 LORENZ, E. N. (1964). *Tellus* **16**, 1.
 MACFADYEN, A. (1963). *Animal Ecology*. London: Pitman.
 MCMURTRIE, R. (1974). (In press.)

- MAY, R. M. (1972). *Am. Nat.* **107**, 46.
- MAY, R. M. (1973). *Stability and Complexity in Model Ecosystems*. Princeton: Princeton University Press.
- MAY, R. M. (1974a). *Science*, N.Y. (in press).
- MAY, R. M. (1974b). In *Progress in Theoretical Biology*, vol. 3 (R. Rosen & F. Snell, eds). New York: Academic Press.
- MAY, R. M., CONWAY, G. R., HASSELL, M. P. & SOUTHWOOD, T. R. E. (1974). *J. Anim. Ecol.* (in press).
- MAYNARD SMITH, J. (1968). *Mathematical Ideas in Biology*. Cambridge: Cambridge University Press.
- MAYNARD SMITH, J. (1974). *Models in Ecology*, p. 53. Cambridge: Cambridge University Press.
- PENNYCUIK, C. J., COMPTON, R. M. & BECKINGHAM, L. (1968). *J. theor. Biol.* **18**, 316.
- PIELOU, E. C. (1969). *An Introduction to Mathematical Ecology*. New York: Wiley.
- SCUDO, F. M. & LEVINE, F. (1974). (In press.)
- SKELLAM, J. F. (1952). *Biometrika* **39**, 346.
- USHER, M. B. (1972). In *Mathematical Models in Ecology* (J. N. R. Jeffers, ed.), pp. 29–60. Oxford: Blackwells.
- UTIDA, S. (1967). *Res. Pop. Ecol.* **9**, 1.
- VAN DER VAART, H. R. (1973). *Bull. math. Biophys.* **35**, 195.
- VARLEY, G. C., GRADWELL, G. R. & HASSELL, M. P. (1973). *Insect Population Ecology*, pp. 20–25. Oxford: Blackwells.
- WILLIAMSON, M. (1974). In *Ecological Stability* (M. B. Usher & M. Williamson, eds), pp. 17–34. London: Chapman and Hall.

Note added in proof:

These analytic difficulties have recently been resolved (May & Oster, to be published). Tables 1 and 2 incorporate these latest results, and give an accurate description of the dynamical behavior of equations (1) and (2).

Biology 504: Modeling Evolutionary Dynamics

Instructors:

Scott Nuismer
email: snuismer@uidaho.edu

Office: Life Sciences South 266C
Phone: 885-4096

Ben Ridenhour
email: bridenho@uidaho.edu

Office: Life Sciences South 281A
Phone: 885-4414

The goal of this course is to familiarize you with the basic analytical and numerical techniques necessary for developing and analyzing your own models of evolutionary processes.

Lectures and Homework

The first five weeks of the course will consist of interactive lectures where models of specific evolutionary processes are developed and analyzed using *Mathematica*. At the end of each lecture, a specific analysis of the model developed in lecture will be assigned as homework and “due” the next class period. Randomly selected students (2-3) will start the next lecture by presenting their analysis of the homework.

Final Mathematica Notebook

The remainder of the semester will be used for the development and analysis of student models. Students can work as individuals or as groups of any size. The only requirement is a shared interest in a particular evolutionary question. Instructors will be available by appointment throughout the remainder of the semester to work with students on their modeling and analysis. We strongly recommend meeting with one of the instructors at least every other week to make sure you are on the right track. The minimum level of interaction is to turn in a draft notebook by April 13 for the instructors to evaluate and make suggestions/corrections. The final notebook is due April 27, and should include the following sections: 1) Introduction to the problem and why we should care; 2) Thorough discussion of model assumptions; 3) Table of parameters/variables and their biological definitions; 3) Derivation of dynamical equations showing all steps; 4) Analysis of dynamical equations; 5) Biological interpretation of mathematical results; 6) Discussion of biological implications/relevance of mathematical results; 7) References.

Week 1

Organizational meeting

Week 2

Lecture 1: Philosophy of modeling

Lecture 2: Single locus models: introduction to *Mathematica*, finding equilibria, local stability analyses, and weak selection approximations

→ Homework: analyze single locus model with intraspecific competition/facilitation

Week 3

Lecture 1: Single trait quantitative genetics (QG) models: assumptions, two different approaches approximation, and abstract integration.

→ Homework: analyze the single trait QG model under stabilizing selection

Lecture 2: Two trait quantitative genetics models: assumptions, approximating fitness functions, and abstract integration.

→ Homework: analyze the two trait QG model under stabilizing selection

Week 4

Lecture 1: Two locus population genetic models and quasi-linkage equilibrium (QLE) approximation

→ Homework: analyze two loci under stabilizing selection.

Lecture 2: Modifier models and QLE approximation

→ Homework: when would a modifier of dominance spread?

Week 5

Lecture 1: Spatial Structure: single locus and single trait models

→ Homework: how much genetic variation is maintained?

Lecture 2: Multi-locus models: moment based approaches (intra-specific competition) and QLE approximations

→ Homework: intraspecific competition

Week 6

Lecture 1: Students choice or overflow

Lecture 2: Students choice or overflow

Weeks 7-14

Working on independent or group projects.

Important Deadlines:

March 2: Topics and Groups due

April 13: Draft *Mathematica* notebooks due

April 27: Final *Mathematica* notebooks due

Grading:

Homework: 50%

Final *Mathematica* notebook 50%

- Lu JC (1976) Singularity theory with an introduction to catastrophe theory. Springer, Berlin
- Majthay A (1985) Foundations of catastrophe theory. Pitman, Boston
- Mather JN (1969) Stability of C^∞ -mappings. I: Annals Math 87:89–104; II: Annals Math 89(2):254–291
- Poston T, Stewart J (1976) Taylor expansions and catastrophes. Pitman, London
- Stewart I (1977) Catastrophe theory. Math Chronicle 5:140–165
- Thom R (1975) Structural stability and morphogenesis. Benjamin Inc., Reading
- Thompson JMT (1982) Instabilities and catastrophes in science and engineering. Wiley, Chichester
- Triebel H (1989) Analysis und mathematische Physik. Birkhäuser, Basel
- Ursprung HW (1982) Die elementare Katastrophentheorie: Eine Darstellung aus der Sicht der Ökonomie. Lecture Notes in Economics and Mathematical Systems, vol 195. Springer, Berlin
- Woodcock A, Davis M (1978) Catastrophe theory. Dutton, New York
- Zeeman C (1976) Catastrophe theory. Sci Am 234(4):65–83

Cell Biology: Networks, Regulation and Pathways

GAŠPER TKAČIK¹, WILLIAM BIALEK^{1,2}

¹ Joseph Henry Laboratories of Physics,
Lewis–Sigler Institute for Integrative Genomics,
Princeton University, Princeton, USA

² Princeton Center for Theoretical Physics,
Princeton University, Princeton, USA

Article Outline

[Glossary](#)

[Definition of the Subject](#)

[Introduction](#)

[Biological Networks and Their Building Blocks](#)

[Models of Biological Networks](#)

[Network Properties and Operating Principles](#)

[Future Directions](#)

[Acknowledgments](#)

[Bibliography](#)

Glossary

Dynamical system is a set of components the properties of which (e. g. their quantity, activity level etc.) change in time because the components interact among themselves and are also influenced by external forces.

Network node is a constituent component of the network, in biological networks most often identified with a molecular species.

Interaction is a connection between network nodes; in biological networks an interaction means that two

nodes chemically react, regulate each other, or effectively influence each other's activities. Interactions are mostly pairwise, but can be higher-order as well; they can be directed or undirected, and are usually characterized by an interaction strength.

Network is a system of interacting nodes. A network can be represented mathematically as a graph, where vertices denote the nodes and edges denote the interactions. Biological networks often are understood to be dynamical systems as well, because the activities of network nodes evolve in time due to the graph of interactions.

Network state is the vector of activities of all nodes that fully characterizes the network at any point in time; since a biological network is a dynamical system, this state generally changes through time according to a set of dynamical equations.

Biological function refers to the role that a specific network plays in the life of the organism; the network can be viewed as existing to perform a task that enables the cell to survive and reproduce, such as the detection or transduction of a specific chemical signal.

Pathway is a subset of nodes and interactions in a network along which information or energy and matter flow in a directed fashion; pathways can be coupled through interactions or unwanted cross-talk.

Curse of dimensionality is the rapid increase of complexity encountered when analyzing or experimentally observing network states, as more and more network nodes are added. If there are N network nodes each of which only has two states (for example *on* and *off*), the number of states that the network can be in grows as 2^N .

Design principle is an (assumed) constraint on the network architecture, stating that a biological network, in addition to performing a certain function, implements that function in a particular way, usually to maximize or minimize some further objective measure, for instance robustness, information transmission, or designability.

Definition of the Subject

In cell biology, networks are systems of interacting molecules that implement cellular functions, such as the regulation of gene expression, metabolism or intracellular signaling. While on a molecular level a biological network is a mesh of chemical reactions between, for example, enzymes and their substrates, or DNA-binding proteins and the genes that they regulate, the collective effect of these reactions can often be thought of as the enabling and regulat-

ing the flow of matter and energy (in metabolic networks), or of information (in signaling and transcriptional regulatory networks). The field is concerned primarily with the description and properties of such flows and with their emergence from network constituent parts – the molecules and their physical interactions. An important focus is also the question of how network function and operating principles can be inferred despite the limited experimental access to network states and building blocks.

Introduction

Biological network has come to mean a system of interacting molecules that jointly perform cellular tasks such as the regulation of gene expression, information transmission, or metabolism [28]. Specific instances of biological networks include, for example, the DNA and DNA binding proteins comprising the transcriptional regulatory network; signaling proteins and small molecules comprising various signaling networks; or enzymes and metabolites comprising the metabolic network. Two important assumptions shape our current understanding of such systems: first, that the biological networks have been under selective evolutionary pressure to perform specific cellular functions in a way that furthers the overall reproductive success of the individual; and second, that these functions often are not implemented on a microscopic level by single molecules, but are rather a collective property of the whole interaction network. The question of how complex behavior emerges in a network of (simple) nodes under a functional constraint is thus central [144].

To start off with a concrete example, consider chemotaxis in the bacterium *Escherichia coli* [16,40], one of the paradigmatic examples of signal transduction. This system is dedicated to steering the bacteria towards areas high in nutrient substances and away from repellents. Chemo-effector molecules in the solution outside the bacterium bind to receptor molecules on the cell surface, and the resulting structural changes in the receptors are relayed in turn by the activities of the intracellular signaling proteins to generate a control signal for molecular motors that drive the bacterial flagella. The chemotactic network consists of about 10 nodes (here, signaling proteins), and the interactions between the nodes are the chemical reactions of methylation or phosphorylation. Notable features of this system include its extreme sensitivity, down to the limits set by counting individual molecules as they arrive at the cell surface [17], and the maintenance of this sensitivity across a huge dynamic range, through an adaptation mechanism that provides nearly perfect compensation of background concentrations [27]. More recently it has been

appreciated that aspects of this functionality, such as perfect adaptation, are also robust against large variations in the concentrations of the network components [6].

Abstractly, different kinds of signaling proteins, such as those in chemotaxis, can be thought of as the building blocks of a network, with their biochemical interactions forming the wiring diagram of the system, much like the components and wiring diagram of, for instance, a radio receiver. In principle, these wiring diagrams are hugely complex; for a network composed of N species, there are $\sim C_k^N$ possible connections among any set of k components, and typically we don't have direct experimental guidance about the numbers associated with each 'wire.' One approach is to view this as giant fitting problem: once we draw a network, there is a direct translation of this graph into dynamical equations, with many parameters, and we should test the predictions of these dynamics against whatever data are available to best determine the underlying parameters. Another approach is to ask whether this large collection of parameters is special in any way other than that it happens to fit the data – are there principles that allow us to predict how these systems *should* work? In the context of chemotaxis, we might imagine that network parameters have been selected to optimize the average progress of bacteria up the chemical gradients of nutrients, or to maximize the robustness of certain functions against extreme parameter variations. These ideas of design principles clearly are not limited to bacterial chemotaxis.

An important aspect of biological networks is that the same components (or components that have an easily identifiable evolutionary relationship) can be (re)used in different modules or used for the same function in a different way across species, as discussed for example by Rao et al. [118] for the case of bacterial chemotaxis. Furthermore, because evolutionary selection depends on function and not directly on microscopic details, different wiring diagrams or even changes in components themselves can result in the same performance; evolutionary process can gradually change the structure of the network as long as its function is preserved; as an example see the discussion of transcriptional regulation in yeast by Tanay et al. [148]. On the other hand, one can also expect that signal processing problems like gain control, noise reduction, ensuring (bi)stability etc, have appeared and were solved repeatedly, perhaps even in similar ways across various cellular functions, and we might be able to detect the traces of their commonality in the network structure, as for example in the discussion of local connectivity in bacterial transcriptional regulation by Shen–Orr et al. [136]. Thus there are reasons to believe that in addition to design prin-

ciples at the network level, there might also be local organizing principles, similar to common wiring motifs in electronic circuitry, yet still independent of the identity of the molecules that implement these principles.

Biological networks have been approached at many different levels, often by investigators from different disciplines. The basic wiring diagram of a network – the fact that a kinase phosphorylates these particular proteins, and not all others, or that a transcription factor binds to the promoter regions of particular genes – is determined by classical biochemical and structural concepts such as binding specificity. At the opposite extreme, trying to understand the collective behavior of the network as a whole suggests approaches from statistical physics, often looking at simplified models that leave out many molecular details. Analyses that start with design principles are yet a different approach, more in the ‘top-down’ spirit of statistical physics but leaving perhaps more room for details to emerge as the analysis is refined. Eventually, all of these different views need to converge: networks really are built out of molecules, their functions emerge as collective behaviors, and these functions must really be functions of use to the organism. At the moment, however, we seldom know enough to bridge the different levels of description, so the different approaches are pursued more or less independently, and we follow this convention here. We will start with the molecular building blocks, then look at models for networks as a whole, and finally consider design principles. We hope that this sequence doesn’t leave the impression that we actually know how to build up from molecules to function!

Before exploring our subject in more detail, we take a moment to consider its boundaries. Our assignment from the editors was to focus on phenomena at the level of molecular and cellular biology. A very different approach attempts to create a ‘science of networks’ that searches for common properties in biological, social, economic and computer networks [104]. Even within the biological world, there is a significant divide between work on networks in cell biology and networks in the brain. As far as we can see this division is an artifact of history, since there are many issues which cut across these different fields. Thus, some of the most beautiful work on signaling comes from photoreceptors, where the combination of optical inputs and electrical outputs allowed, already in the 1970s, for experiments with a degree of quantitative analysis that even today is hard to match in systems which take chemical inputs and give outputs that modulate the expression levels of genes [14,121]. Similarly, problems of noise in the control of gene expression have parallels in the long history of work on noise in ion channels, as we have

discussed elsewhere [156], and the problems of robustness have also been extensively explored in the network of interactions among the multiple species of ion channels in the membrane [51,88]. Finally, the ideas of collective behavior are much better developed in the context of neural networks than in cellular networks, and it is an open question how much can be learned by studying these different systems in the same language [151].

Biological Networks and Their Building Blocks

Genetic Regulatory Networks

Cells constantly adjust their levels of gene expression. One central mechanism in this regulatory process involves the control of transcription by proteins known as transcription factors (TFs), which locate and bind short DNA sequences in the regulated genes’ promoter or enhancer regions. A given transcription factor can regulate either a few or a sizable proportion of the genes in a genome, and a single gene may be regulated by more than one transcription factor; different transcription factors can also regulate each other [166].

In the simplest case of a gene regulated by a single TF, the gene might be expressed whenever the factor – in this case called an activator – is bound to the cognate sequence in the promoter (which corresponds to the situation when the TF concentration in the nucleus is high), whereas the binding of a repressor would shut a normally active gene down. The outlines of these basic control principles were established long ago, well before the individual transcription factors could be isolated, in elegant experiments on the *lactose* operon of *Escherichia coli* [69] and even simpler model systems such as phage λ [115]. To a great extent the lessons learned from these experiments have provided the framework for understanding transcriptional control more generally, in prokaryotes [114], eukaryotes [75], and even during the development of complex multicellular organisms [8].

The advent of high throughput techniques for probing gene regulation has extended our reach beyond single genes. In particular, microarrays [30] and the related data analysis tools, such as clustering [36], have enabled researchers to find sets of genes, or *modules*, that are *coexpressed*, i. e. up- or down-regulated in a correlated fashion when the organism is exposed to different external conditions, and are thus probably regulated by the same set of transcription factors. Chromatin immunoprecipitation (ChIP) assays have made it possible to directly screen for short segments of DNA that known TFs bind; using microarray technology it is then possible to locate the intergenic regions which these segments belong to, and hence

find the regulated genes, as has recently been done for the *Saccharomyces cerevisiae* DNA-TF interaction map [86].

These high throughput experimental approaches, combined with traditional molecular biology and complemented by sequence analysis and related mathematical tools [139], provide a large scale, topological view of the transcriptional regulatory network of a particular organism, where each link between two nodes (genes) in the regulatory graph implies either activation or repression [5]. While useful for describing causal interactions and trying to predict responses to mutations and external perturbations [89], this picture does not explain how the network operates on a physical level: it lacks dynamics and specifies neither the strengths of the interactions nor how all the links converging onto a given node jointly exercise control over it. To address these issues, representative wild-type or simple synthetic regulatory elements and networks consisting of a few nodes have been studied extensively to construct quantitative models of the network building blocks.

For instance, combinatorial regulation of a gene by several transcription factors that bind and interact on the promoter has been considered by Buchler et al. [31] as an example of (binary) biological computation and synthetic networks implementing such computations have been created [56,170]. Building on classical work describing allosteric proteins such as hemoglobin, thermodynamic models have been used with success to account for combinatorial interactions on the operator of the λ phage [2]. More recently Bintu et al. [24,25] have reviewed the equilibrium statistical mechanics of such interactions, Setty et al. [134] have experimentally and systematically mapped out the response surface of the *lac* promoter to combinations of its two regulatory inputs, cAMP and IPTG, and Kuhlman et al. [85] have finally provided a consistent picture of the known experimental results and the thermodynamic model for the combinatorial regulation of the lactose operon. There have also been some successes in eukaryotic regulation, where Schroeder et al. [132] used thermodynamically motivated models to detect clusters of binding sites that regulate the gap genes in morphogenesis of the fruit fly.

Gene regulation is a dynamical process composed of a number of steps, for example the binding of TF to DNA, recruitment of transcription machinery and the production of the messenger RNA, post-transcriptional regulation, splicing and transport of mRNA, translation, maturation and possible localization of proteins. While the extensive palette of such microscopic interactions represents a formidable theoretical and experimental challenge for each detailed study, on a network level it primarily induces three effects. First, each node – usually understood

as the amount of gene product – in a graph of regulatory interactions is really not a single dynamical variable, but has a nontrivial internal state representing the configuration on the associated promoter, concentration of the corresponding messenger RNA etc.; the relation of these quantities to the concentration of the output protein is not necessarily straightforward, as emphasized in recent work comparing mRNA and protein levels in yeast [46]. Second, collapsing multiple chemical species onto a single node makes it difficult to include non-transcriptional regulation of gene expression in the same framework. Third, the response of the target gene to changes in the concentrations of its regulators will be delayed and extended in time, as in the example explored by Rosenfeld and Alon [123].

Perhaps the clearest testimonies to the importance of dynamics in addition to network topology are provided by systems that involve regulatory loops, in which the output of a network feeds back on one of the inputs as an activator or repressor. McAdams and Shapiro [99] have argued that the time delays in genetic regulatory elements are essential for the proper functioning of the phage λ switch, while Elowitz and Leibler [38] have created a synthetic circuit made up of three mutually repressing genes (the “repressilator”), that exhibits spontaneous oscillations. Circadian clocks are examples of naturally occurring genetic oscillators [171].

In short, much is known about the skeleton of genetic regulatory interactions for model organisms, and physical models exist for several well studied (mostly prokaryotic) regulatory elements. While homology allows us to bridge the gap between model organisms and their relatives, it is less clear how and at which level of detail the knowledge about regulatory elements must be combined into a network to explain and predict its function.

Protein–Protein Interaction Networks

After having been produced, proteins often assemble into complexes through direct contact interactions, and these complexes are functionally active units participating in signal propagation and other pathways. Proteins also interact through less persistent encounters, as when a protein kinase meets its substrate. It is tempting to define a link in the network of protein–protein interactions by such physical associations, and this is the basis of several experimental methods which aim at a genome-wide survey of these interactions. Although starting out being relatively unreliable (with false positive rates of up to 50%), high throughput techniques like the yeast two hybrid assay [68,161] or mass spectrometry [45,61] are providing data of increasing quality about protein–protein in-

teractions, or the “interactome” [84]. While more reliable methods are being developed [5] and new organisms are being analyzed in this way [49,91,125], the existing interaction data from high throughput experiments and curated databases has already been extensively studied.

Interpretation of the interactions in the protein network is tricky, however, due to the fact that different experimental approaches have various biases – for example, mass spectrometry is biased towards detecting interactions between proteins of high abundance, while two hybrid methods seem to be unbiased in this regard; on the other hand, all methods show some degree of bias towards different cellular localizations and evolutionary novelty of the proteins. Assessing such biases, however, currently depends not on direct calibration of the methods themselves but on comparison of the results with manually curated databases, although the databases surely have their own biases [70]. It is reassuring that the intersection of various experimental results shows significantly improved agreement with the databases, but this comes at the cost of a substantial drop in coverage of the proteome [100].

In contrast to the case of transcriptional regulation, the relationship between two interacting proteins is symmetric: if protein A binds to protein B, B also binds to A, so that the network is described by an undirected graph. Most of the studies have been focused on binary interactions that yeast two hybrid and derived approaches can probe, although spectrometry can detect multiprotein complexes as well. Estimates of number of links in these networks vary widely, even in the yeast *Saccharomyces cerevisiae*: Krogan et al. [84] directly measure around 7100 interactions (between 2700 proteins), while Tucker et al. [158] estimate the total to be around 13 000–17 000, and von Mering et al. [100] would put the lower estimate at about 30 000. Apart from the experimental biases that can influence such estimates and have been discussed already, it is important to realize that each experiment can only detect interactions between proteins that are expressed under the chosen external conditions (e. g. the nutrient medium); moreover, interactions can vary from being transient to permanent, to which various measurement methods respond differently. It will thus become increasingly important to qualify each interaction in a graph by specifying how it depends on context in which the interaction takes place.

Proteins ultimately carry out most of the cellular processes such as transcriptional regulation, signal propagation and metabolism, and these processes can be modeled by their respective network and dynamical system abstractions. In contrast, the interactome is not a dynamical system itself, but instead captures specific reactions (like pro-

tein complex assembly) and structural and/or functional relations that are present in all of the above processes. In this respect it has an important practical role of annotating currently unknown proteins through ‘guilt by association,’ by tying them into complexes and processes with a previously known function.

Metabolic Networks

Metabolic networks organize our knowledge about anabolic and catabolic reactions between the enzymes, their substrates and co-factors (such as ATP), by reducing the set of reactions to a graph representation where two substrates are joined by a link if they participate in the same reaction. For model organisms like the bacterium *Escherichia coli* the metabolic networks have been studied in depth and are publicly available [77,78], and an increasing number of analyzed genomes offers sufficient sampling power to make statistical statements about the network properties across different domains of life [72].

Several important features distinguish metabolic from protein–protein interaction and transcriptional regulation networks. First, for well studied systems the coverage of metabolic reactions is high, at least for the central routes of energy metabolism and small molecule synthesis; notice that this is a property of our knowledge, not a property of the networks (!). Second, cellular concentrations of metabolites usually are much higher than those of transcription factors, making the stochasticity in reactions due to small molecular counts irrelevant. Third, knowledge of the stoichiometry of reactions allows one to directly write down a system of first order differential equations for the metabolite fluxes [60], which in steady state reduces to a set of linear constraints on the space of solutions. These chemical constraints go beyond topology and can yield strong and testable predictions; for example, Ibarra et al. [66] have shown how computationally maximizing the growth rate of *Escherichia coli* within the space of allowed solutions given by flux balance constraints can correctly predict measurable relationships between oxygen and substrate uptake, and that bacteria can be evolved towards the predicted optimality for growth conditions in which the response was initially suboptimal.

Signaling Networks

Signaling networks consist of receptor and signaling proteins that integrate, transmit and route information by means of chemical transformations of the network constituents. One class of such transformations, for example, are post-translational modifications, where targets are phosphorylated, methylated, acetylated, ... on spe-

cific residues, with a resulting change in their enzymatic (and thus signaling) activity. Alternatively, proteins might form stable complexes or dissociate from them, again introducing states of differential activity. The ability of cells to modify or tag proteins (possibly on several residues) can increase considerably the cell's capacity to encode its state and transmit information, assuming that the signaling proteins are highly specific not only for the identity but also the modification state of their targets; for a review see [110].

Despite the seeming overlap between the domains of protein–protein network and signaling networks, the focus of the analysis is substantially different. The interactome is simply a set of possible protein–protein interactions and thus a topological (or connectivity) map; in contrast, signaling networks aim to capture signal transduction and therefore need to establish a causal map, in which the nature of the protein–protein interaction, its direction and timescale, and its quantitative effect on the activity of the target protein matter. As an example, see the discussion by Kolch et al. [83] on the role of protein–protein interactions in MAPK signaling cascade.

Experiments on some signaling systems, such as the *Escherichia coli* chemotactic module, have generated enough experimental data to require detailed models in the form of dynamical equations. Molecular processes in a signaling cascade extend over different time scales, from milliseconds required for kinase and phosphatase reactions and protein conformational changes, to minutes or more required for gene expression control, cell movement and receptor trafficking; this fact, along with the (often essential) spatial effects such as the localization of signaling machinery and diffusion of chemical messengers, can considerably complicate analyses and simulations.

Signaling networks are often factored into pathways that have specific inputs, such as the ligands of the G protein coupled receptors on the cell surface, and specific outputs, as with pathways that couple to the transcriptional regulation apparatus or to changes in the intracellular concentration of messengers such as calcium or cyclic nucleotides. Nodes in signaling networks can participate in several pathways simultaneously, thus enabling signal integration or potentially inducing damaging “crosstalk” between pathways; how junctions and nodes process signals is an area of active research [74].

The components of signaling networks have long been the focus of biochemical research, and genetic methods allow experiments to assess the impact of knocking out or over-expressing particular components. In addition, several experimental approaches are being designed specifically for elucidating signaling networks. Ab-chips localize

various signaling proteins on chips reminiscent of DNA microarrays, and stain them with appropriate fluorescent antibodies [105]. Multicolor flow cytometry is performed on cells immuno-stained for signaling protein modifications and hundreds of single cell simultaneous measurements of the modification state of pathway nodes are collected [113]. Indirect inference of signaling pathways is also possible from genomic or proteomic data.

One well studied signal transduction system is the mitogen activated protein kinase (MAPK) cascade that controls, among other functions, cell proliferation and differentiation [32]. Because this system is present in all eukaryotes and its structural components are used in multiple pathways, it has been chosen as a paradigm for the study of specificity and crosstalk. Similarly, the TOR system, identified initially in yeast, is responsible for integrating the information on nutrient availability, growth factors and energy status of the cell and correspondingly regulating the cell growth [95]. Another interesting example of signal integration and both intra- and inter-cellular communication is observed in the quorum sensing circuit of the bacterium *Vibrio harveyi*, where different kinds of species- and genus-specific signaling molecules are detected by their cognate receptors on the cell surface, and the information is fed into a common *Lux* phosphorelay pathway which ultimately regulates the quorum sensing genes [165].

Models of Biological Networks

Topological Models

The structural features of a network are captured by its connectivity graph, where interactions (reactions, structural relations) are depicted as the links between the interacting nodes (genes, proteins, metabolites). Information about connectivity clearly cannot and does not describe the network behavior, but it might influence and constrain it in revealing ways, similar to effect that the topology of the lattice has on the statistical mechanics of systems living on it.

Theorists have studied extensively the properties of regular networks and random graphs starting with Erdős and Rényi in 1960s. The first ones are characterized by high symmetry inherent in a square, triangular, or all-to-all (mean field) lattice; the random graphs were without such regularity, constructed simply by distributing K links at random between N nodes. The simple one-point statistical characterization that distinguishes random from regular networks looks at the node degree, that is the probability $P(k)$ that any node has k incoming and/or outgoing links. For random graphs this distribution is Poisson,

meaning that most of the nodes have degrees very close to the mean, $\langle k \rangle = \sum_k k P(k)$, although there are fluctuations; for regular lattices every node has the same connectivity to its neighbors.

The first analyses of the early reconstructions of large metabolic networks revealed a surprising “scale free” node degree distribution, that is $P(k) \sim k^{-\gamma}$, with γ between 2 and 3 for most networks. For the physics community, which had seen the impact of such scale invariance on our understanding of phase transitions, these observations were extremely suggestive. It should be emphasized that for many problems in areas as diverse as quantum field theory, statistical mechanics and dynamical systems, such scaling relations are much more than curiosities. Power laws relating various experimentally observable quantities are exact (at least in some limit), and the exponents (here, γ) really contain everything one might want to know about the nature of order in the system. Further, some of the first thoughts on scaling emerged from phenomenological analyses of real data. Thus, the large body of work on scaling ideas in theoretical physics set the stage for people to be excited by the experimental observation of power laws in much more complex systems, although it is not clear to us whether the implied promise of connection to a deeper theoretical structure has been fulfilled. For divergent views on these matters see Barabási et al. [10] and Keller et al. [81].

The most immediate practical consequence of a scale free degree distribution is that – relative to expectations based on random graphs – there will be an over-representation of nodes with very large numbers of links, as with pyruvate or co-enzyme A in metabolic networks [72,163]. These are sometimes called hubs, although another consequence of a scale free distribution is that there is no ‘critical degree of connection’ that distinguishes hubs from non-hubs. In the protein–protein interaction network of *Saccharomyces cerevisiae*, nodes with higher degree are more likely to represent essential proteins [73], suggesting that node degree does have some biological meaning. On the theoretical side, removal of a sizable fraction of nodes from a scale free network will neither increase the network diameter much, nor partition the network into equally sized parts [3], and it is tempting to think that this robustness is also biologically significant. The scale free property has been observed in many non-biological contexts, such as the topology of social interactions, World Wide Web links, electrical power grid connectivity ... [144]. A number of models have been proposed for how such scaling might arise, and some of these ideas, such as growth by preferential attachment, have a vaguely biological flavor [11,12]. Finding the properties of networks that actually discrimi-

nate among different mechanisms of evolution or growth turns out to be surprisingly subtle [173].

Two other revealing measures are regularly computed for biological networks. The mean path length, $\langle l \rangle$, is the shortest path between a pair of nodes, averaged over all pairs in the graph, and measures the network’s overall ‘navigability.’ Intuitively, short path lengths correspond to, for example, efficient or fast flow of information and energy in signaling or metabolic networks, quick spread of diseases in a social network and so on. The clustering coefficient of a node i is defined as $C_i = 2n_i/k_i(k_i - 1)$, where n_i is the number of links connecting the k_i neighbors of node i to each other; equivalently, C_i is the ratio between the number of triangles passing through two neighbors of i and node i itself, divided by the maximum possible number of such triangles. Random networks have low path lengths and low clustering coefficients, whereas regular lattices have long path lengths and are locally clustered. Watts and Strogatz [167] have constructed an intermediate regime of “small world” networks, where the regular lattice has been perturbed by a small number of random links connecting distant parts of the network together. These networks, although not necessarily scale free, have short path lengths and high clustering coefficients, a property that was subsequently observed in metabolic and other biological networks as well [163].

A high clustering coefficient suggests the existence of densely connected groups of nodes within a network, which seems contradictory to the idea of scale invariance, in which there is no inherent group or cluster size; Ravasz et al. [120] addressed this problem by introducing hierarchical networks and providing a simple construction for synthetic hierarchical networks exhibiting both scale free and clustering behaviors. Although there is no unique scale for the clusters, clusters will appear at any scale one chooses to look at, and this is revealed by the scaling of clustering coefficient $C(k)$ with the node degree k , $C(k) \sim k^{-1}$, on both synthetic as well as natural metabolic networks of organisms from different domains of life [120]. Another interesting property of some biological networks is an anti-correlation of node degree of connected nodes [96], which we can think of as a ‘dissociative’ structure; in contrast, for example, with the associative character of social networks, where well connected people usually know one another.

As we look more finely at the structure of the graph representing a network, there is of course a much greater variety of things to look at. For example, Spirin and Mirny [142] have focused on high clustering coefficients as a starting point and devised algorithms to search for modules, or densely connected subgraphs within the yeast

protein–protein interaction network. Although the problem has combinatorial complexity in general, the authors found about 50 modules (of 5–10 proteins in size, some of which were unknown at the time) that come in two types: the first represents dynamic functional units (e.g. signaling cascades), and the second protein complexes. A similar conclusion was reached by Han et al. [57], after having analyzed the interactome in combination with the temporal gene expression profiles and protein localization data; the authors argue that nodes of high degree can sit either at the centers of modules, which are simultaneously expressed (“party hubs”), or they can be involved in various pathways and modules at different times (“date hubs”). The former kind is at a lower level of organization, whereas the latter tie the network into one large connected component.

Focusing on even a smaller scale, Shen-Orr et al. [136] have explored motifs, or patterns of connectivity of small sets of nodes that are over-represented in a given network compared to the randomized networks of the same degree distribution $P(k)$. In the transcriptional network of the bacterium *E. coli*, three such motifs were found: feed forward loops (in which gene X regulates Y that regulates Z, but X directly regulates Z as well), single input modules (where gene X regulates a large number of other genes in the same way and usually auto-regulates itself), and dense overlapping regulons (layers of overlapping interactions between genes and a group of transcription factors, much denser than in randomized networks). The motif approach has been extended to combined network of transcriptional regulation and protein–protein interactions [169] in yeast, as well as to other systems [101].

At the risk of being overly pessimistic, we should conclude this section with a note of caution. It would be attractive to think that a decade of work on network topology has resulted in a coherent picture, perhaps of the following form: on the smallest scale, the nodes of biological networks are assembled into motifs, these in turn are linked into modules, and this continues in a hierarchical fashion until the entire network is scale free. As we will discuss again in the context of design principles, the notion of such discrete substructure – motifs and modules – is intuitively appealing, and some discussions suggest that it is essential either for the function or the evolution of networks. On the other hand, the evidence for such structure usually is gathered with reference to some null model (e.g., a random network with the same $P(k)$), so we don’t even have an absolute definition of these structures, much less a measure of their sufficiency as a characterization of the whole system; for attempts at an absolute definition of modularity see Ziv et al. [174] and Hofman and Wiggins [62]. Similarly, while it is appealing to think about

scale free networks, the evidence for scaling almost always is confined to less than two decades, and in practice scaling often is not exact. It is then not clear whether the idealization of scale invariance captures the essential structure in these systems.

Boolean Networks

A straightforward extension of the topological picture that also permits the study of network dynamics assumes that the entities at the nodes – for example, genes or signaling proteins – are either ‘on’ or ‘off’ at each moment of time, so that for node i the state at time t is $\sigma_i(t) \in \{0, 1\}$. Time is usually discretized, and an additional prescription is needed to implement the evolution of the system: $\sigma_i(t+1) = f_i(\{\sigma_\mu(t)\})$, where f_i is a function that specifies how the states of the nodes μ that are the inputs to node i in the interaction graph combine to determine the next state at node i . For instance, f_A might be a Boolean function for gene A, which needs to have its activator gene B present and repressor gene C absent, so that $\sigma_A(t+1) = \sigma_B(t) \wedge \bar{\sigma}_C(t)$. Alternatively, f might be a function that sums the inputs at state t with some weights, and then sets $\sigma_i = 1(0)$ if the result is above (below) a threshold, as in classical models of neural networks.

Boolean networks are amenable both to analytical treatment and to efficient simulation. Early on, Kauffman [80] considered the family of random boolean networks. In these models, each node is connected at random to K other nodes on average, and it computes a random Boolean function of its inputs in which a fraction ρ of the 2^K possible input combinations leads to $\sigma_i(t+1) = 1$. In the limit that the network is large, the dynamics are either regular (settling into a stable fixed cycle) or chaotic, and these two phases are separated by a separatrix $2\rho(1-\rho)K = 1$ in the phase space (ρ, K) .

Aldana and Cluzel [4] have shown that for connectivities of $K \sim 20$ that could reasonably be expected in e.g. transcriptional regulatory networks, the chaotic regime dominates the phase space. They point out, however, that if the network is scale free, there is no ‘typical’ K as the distribution $P(k) \sim k^{-\gamma}$ does not have a well-defined mean for $\gamma \leq 3$ and the phase transition criterion must be restated. It turns out, surprisingly, that regular behavior is possible for values of γ between 2 and 2.5, observed in most biological networks, and this is exactly the region where the separatrix lies. Scale free architecture, at least for Boolean networks, seems to prevent chaos.

Several groups have used Boolean models to look at specific biological systems. Thomas [150] has established a theoretical framework in which current states of the

genes (as well as the states in the immediate past) and the environmental inputs are represented by Boolean variables that evolve through the application of Boolean functions. This work has been continued by, for example, Sanchez and Thieffry [128] who analyzed the gap-gene system of the fruit fly *Drosophila* by building a Boolean network that generates the correct levels of gene expression for 4 gap genes in response to input levels of 3 maternal morphogens with spatially varying profiles stretched along the anterior-posterior axis of the fly embryo. Interestingly, to reproduce the observed results and correctly predict the known *Drosophila* segmentation mutants, the authors had to introduce generalized Boolean variables that can take more than two states, and have identified the smallest necessary number of such states for each gene.

In a similar spirit, Li et al. [91] studied the skeleton of the budding yeast cell cycle, composed of 11 nodes, and a thresholding update rule. They found that the topology of this small network generates a robust sequence of transitions corresponding to known progression through yeast cell-cycle phases G1 (growth), S (DNA duplication), G2 (pause) and M (division), triggered by a known ‘cell-size checkpoint.’ This progression is robust, in the sense that the correct trajectory is the biggest dynamical attractor of the system, with respect to various choices of update rules and parameters, small changes in network topology, and choice of triggering checkpoints.

The usefulness of Boolean networks stems from the relative ease of implementation and simple parametrization of network topology and dynamics, making them suitable for studying medium or large networks. In addition to simplifying the states at the nodes to two (or more) discrete levels, which is an assumption that has not been clearly explored, one should be cautious that the discrete and usually synchronous dynamics in time can induce unwanted artifacts.

Probabilistic Models

Suppose one is able to observe simultaneously the activity levels of several proteins comprising a signaling network, or the expression levels of a set of genes belonging to the same regulatory module. Because they are part of a functional whole, the activity levels of the components will be correlated. Naively, one could build a network model by simply computing pairwise correlation coefficients between pairs of nodes, and postulating an interaction, and therefore a link, between the two nodes whenever their correlation is above some threshold. However, in a test case where $A \rightarrow B \rightarrow C$ (gene A induces B which induces C), one expects to see high positive correlation among all

three elements, even though there is no (physical) interaction between A and C. Correlation therefore is not equal to interaction or causation. Constructing a network from the correlations in this naive way also does not lead to a generative model that would predict the probabilities for observing different states of the network as a whole. Another approach is clearly needed; see Markowitz and Spang [94] for a review.

In the simple case where the activity of a protein/gene i can either be ‘on’ ($\sigma_i = 1$) or ‘off’ ($\sigma_i = 0$), the state of a network with N nodes will be characterized by a binary word of N bits, and because of interaction between nodes, not all these words will be equally likely. For example, if node A represses node B, then combinations such as $1_A 0_B \dots$ or $0_A 1_B \dots$ will be more likely than $1_A 1_B \dots$. In the case of deterministic Boolean networks, having node A be ‘on’ would imply that node B is ‘off’ with certainty, but in probabilistic models it only means that there is a positive *bias* for node B to be ‘off,’ quantified by the probability that node B is ‘off’ given that the state of node A is known. Having this additional probabilistic degree of freedom is advantageous, both because the network itself might be noisy, and because the experiment can induce errors in the signal readout, making the inference of deterministic rules from observed binary patterns an ill-posed problem.

Once we agree to make a probabilistic model, the goal is to find the distribution over all network states, which we can also think of as the joint distribution of all the N variable that live on the nodes of the network, $P(\sigma_1, \dots, \sigma_N | C)$, perhaps conditioned on some fixed set of environmental or experimental factors C . The activities of the nodes, σ_i , can be binary, can take on a discrete set of states, or be continuous, depending on our prior knowledge about the system and experimental and numerical constraints. Even for a modest N , experiments of realistic scale will not be enough to directly estimate the probability distribution, since even with binary variable the number of possible states, and hence the number of parameters required to specify the general probability distribution, grows as $\sim 2^N$. Progress thus depends in an essential way on simplifying assumptions.

Returning to the three gene example $A \rightarrow B \rightarrow C$, we realize that C depends on A only through B, or in other words, C is *conditionally independent* of A and hence no interaction should be assigned between nodes A and C. Thus, the joint distribution of three variables can be factorized,

$$P(\sigma_A, \sigma_B, \sigma_C) = P(\sigma_C | \sigma_B) P(\sigma_B | \sigma_A) P(\sigma_A).$$

One might hope that, even in a large network, these sorts

of conditional independence relations could be used to simplify our model of the probability distribution. In general this doesn't work, because of feedback loops which, in our simple example, would include the possibility that C affects the state of A, either directly or through some more circuitous path. Nonetheless one can try to make an approximation in which loops either are neglected or (more sensibly) taken into account in some sort of average way; in statistical mechanics, this approximation goes back at least to the work of Bethe [19].

In the computer science and bioinformatics literature, the exploitation of Bethe-like approximations has come to be known as 'Bayesian network modeling' [43]. In practice what this approach does is to search among possible network topologies, excluding loops, and then for fixed topology one uses the conditional probability relationships to factorize the probability distribution and fit the tables of conditional probabilities at each node that will best reproduce some set of data. Networks with more links have more parameters, so one must introduce a trade-off between the quality of the fit to the data and this increasing complexity. In this framework there is thus an explicit simplification based on conditional independence, and an implicit simplification based on a preference for models with fewer links or sparse connectivity.

The best known application of this approach to a biological network is the analysis of the MAPK signaling pathway in T cells from the human immune system [127]. The data for this analysis comes from experiments in which the phosphorylated states of 11 proteins in the pathway are sampled simultaneously by immunostaining [113], with hundreds of cells sampled for each set of external conditions. By combining experiments from multiple conditions, the Bayesian network analysis was able to find a network of interactions among the 11 proteins that has high overlap with those known to occur experimentally.

A very different approach to simplification of probabilistic models is based on the maximum entropy principle [71]. In this approach one views a set of experiments as providing an estimate of some set of correlations, for example the $\sim N^2$ correlations among all pairs of elements in the network. One then tries to construct a probability distribution which matches these correlations but otherwise has as little structure – as much entropy – as possible. We recall that the Boltzmann distribution for systems in thermal equilibrium can be derived as the distribution which has maximum entropy consistent with a given average energy, and maximum entropy modeling generalizes this to take account of other average properties. In fact one can construct a hierarchy of maximum entropy distributions which are consistent with higher and higher orders

of correlation [130]. Maximum entropy models for binary variables that are consistent with pairwise correlations are exactly the Ising models of statistical physics, which opens a wealth of analytic tools and intuition about collective behavior in these systems.

In the context of biological networks (broadly construed), recent work has shown that maximum entropy models consistent with pairwise correlations are surprisingly successful at describing the patterns of activity among populations of neurons in the vertebrate retina as it responds to natural movies [131,153]. Similar results are obtained for very different retinas under different conditions [137], and these successes have touched a flurry of interest in the analysis of neural populations more generally. The connection to the Ising model has a special resonance in the context of neural networks, where the collective behavior of the Ising model has been used for some time as a prototype for thinking about the dynamics of computation and memory storage [64]; in the maximum entropy approach the Ising model emerges directly as the least structured model consistent with the experimentally measured patterns of correlation among pairs of cells. A particularly striking result of this analysis is that the Ising models which emerge seem to be poised near a critical point [153]. Returning to cell biology, the maximum entropy approach has also been used to analyze patterns of gene expression in yeast [90] as well as to revisit the MAPK cascade [151].

Dynamical Systems

If the information about a biological system is detailed enough to encompass all relevant interacting molecules along with the associated reactions and estimated reaction rates, and the molecular noise is expected to play a negligible role, it is possible to describe the system with rate equations of chemical kinetics. An obvious benefit is the immediate availability of mathematical tools, such as steady state and stability analyses, insight provided by nonlinear dynamics and chaos theory, well developed numerical algorithms for integration in time and convenient visualization with phase portraits or bifurcation diagrams. Moreover, analytical approximations can be often exploited productively when warranted by some prior knowledge, for example, in separately treating 'fast' and 'slow' reactions. In practice, however, reaction rates and other important parameters are often unknown or known only up to order-of-magnitude estimations; in this case the problem usually reduces to the identification of phase space regions where the behavior of the system is qualitatively the same, for example, regions where the system exhibits limit-cycle oscil-

lations, bistability, convergence into a single steady state etc.; see Tyson et al. [159] for a review. Despite the difficulties, deterministic chemical kinetic models have been very powerful tools in analyzing specific network motifs or regulatory elements, as in the protein signaling circuits that achieve perfect adaptation, homeostasis, switching and so on, described by Tyson et al. [160], and more generally in the analysis of transcriptional regulatory networks as reviewed by Hasty et al. [59].

In the world of bacteria, some of the first detailed computer simulations of the chemotaxis module of *Escherichia coli* were undertaken by Bray et al. [29]. The signaling cascade from the Tar receptor at the cell surface to the modifications in the phosphorylation state of the molecular motor were captured by Michaelis–Menten kinetic reactions (and equilibrium binding conditions for the receptor), and the system of equations was numerically integrated in time. While slow adaptation kinetics was not studied in this first effort, the model nevertheless qualitatively reproduces about 80 percent of examined chemotactic protein deletion and overexpression mutants, although the extreme sensitivity of the system remained unexplained.

In eukaryotes, Novak and Tyson [107] have, for instance, constructed an extensive model of cell cycle control in fission yeast. Despite its complexity (~ 10 proteins and ~ 30 rate constants), Novak and colleagues have provided an interpretation of the system in terms of three interlocking modules that regulate the transitions from G1 (growth) into S (DNA synthesis) phase, from G2 into M (division) phase, and the exit from mitosis, respectively. The modules are coupled through cdc2/cdc13 protein complex and the system is driven by the interaction with the cell size signal (proportional to the number of ribosomes per nucleus). At small size, the control circuit can only support one stable attractor, which is the state with low cdc2 activity corresponding to G1 phase. As the cell grows, new stable state appears and the system makes an irreversible transition into S/G2 at a bifurcation point, and, at an even larger size, the mitotic module becomes unstable and executes limit cycles in cdc2-cdc13 activity until the M phase is completed and the cell returns to its initial size. The basic idea is that the cell, driven by the the size readout, progresses through robust cell states created by bistability in the three modules comprising the cell cycle control – in this way, once it commits to a transition from G2 state into M, small fluctuations will not flip it back into G2. The mathematical model has in this case successfully predicted the behaviors of a number of cell cycle mutants and recapitulated experimental observations collected during 1970s and 1980s by Nurse and collaborators [108].

The circadian clock is a naturally occurring transcriptional module that is particularly amenable to dynamical systems modeling. Leloup and Goldbeter [87] have created a mathematical model of a mammalian clock (with ~ 20 rate equations) that exhibits autonomous sustained oscillations over a sizable range of parameter values, and reproduces the entrainment of the oscillations to the light–dark cycles through light-induced gene expression. The basic mechanism that enables the cyclic behavior is negative feedback transcriptional control, although the actual circuit contains at least two coupled oscillators. Studying circadian clock in mammals, the fruit fly *Drosophila* or *Neurospora* is attractive because of the possibility of connecting a sizable catalogue of physiological disorders in circadian rhythms to malfunctions in the clock circuit and direct experimentation with light–dark stimuli [171]. Recent experiments indicate that at least in cyanobacteria the circadian clock can be reconstituted from a surprisingly small set of biochemical reactions, without transcription or translation [102,157], and this opens possibilities for even simpler and highly predictive dynamical models [126].

Dynamical modeling has in addition been applied to many smaller systems. For example, the construction of a synthetic toggle switch [44], and the ‘repressilator’ – oscillating network of three mutually repressing genes [38] – are examples where mathematical analysis has stimulated the design of synthetic circuits. A successful reaction–diffusion model of how localization and complex formation of Min proteins can lead to spatial limit cycle oscillations (used by *Escherichia coli* to find its division site) was constructed by Huang et al. [65]. It remains a challenge, nevertheless, to navigate in the space of parameters as it becomes ever larger for bigger networks, to correctly account for localization and count various forms of protein modifications, especially when the signaling networks also couple to transcriptional regulation, and to find a proper balance between models that capture all known reactions and interactions and phenomenological models that include coarse-grained variables.

Stochastic Dynamics

Stochastic dynamics is in principle the most detailed level of system description. Here, the (integer) count of every molecular species is tracked and reactions are drawn at random with appropriate probabilities per unit time (proportional to their respective reaction rates) and executed to update the current tally of molecular counts. An algorithm implementing this prescription, called the stochastic simulation algorithm or SSA, was devised by Gille-

spie [47]; see Gillespie [48] for a review of SSA and a discussion of related methods. Although slow, this approach for simulating chemical reactions can be made exact. In general, when all molecules are present in large numbers and continuous, well-mixed concentrations are good approximations, the (deterministic) rate dynamics equations and stochastic simulation give the same results; however, when molecular counts are low and, consequently, the stochasticity in reaction timing and ordering becomes important, the rate dynamics breaks down and SSA needs to be used. In biological networks and specifically in transcriptional regulation, a gene and its promoter region are only present in one (or perhaps a few) copies, while transcription factors that regulate it can also be at nanomolar concentrations (i. e. from a few to a few hundred molecules per nucleus), making stochastic effects possibly very important [97,98].

One of the pioneering studies of the role of noise in a biological system was a simulation of the phage λ lysogenic switch by Arkin et al. [7]. The life cycle of the phage is determined by the concentrations of two transcription factors, *cI* (lambda repressor) and *cro*, that compete for binding to the same operator on the DNA. If *cI* is prevalent, the phage DNA is integrated into the host's genome and no phage genes except for *cI* are expressed (the lysogenic state); if *cro* is dominant, the phage is in lytic state, using cell's DNA replication machinery to produce more phages and ultimately lyse the host cell [115]. The switch is bistable and the fate of the phage depends on the temporal and random pattern of gene expression of two mutually antagonistic transcription factors, although the balance can be shifted by subjecting the host cell to stress and thus flipping the toggle into lytic phase. The stochastic simulation correctly reproduces the experimentally observed fraction of lysogenic phages as a function of multiplicity-of-infection. An extension of SSA to spatially extended models is possible.

Although the simulations are exact, they are computationally intensive and do not offer any analytical insight into the behavior of the solutions. As a result, various theoretical techniques have been developed for studying the effects of stochasticity in biological networks. These are often operating in a regime where the deterministic chemical kinetics is a good approximation, and noise (i. e. fluctuation of concentrations around the mean) is added into the system of differential equations as a perturbation; these Langevin methods have been useful for the study of noise propagation in regulatory networks [76,111,149]. The analysis of stochastic dynamics is especially interesting in the context of design principles which consider the reliability of network function, to which we return below.

Network Properties and Operating Principles

Modularity

Biological networks are said to be modular, although the term has several related but nevertheless distinct meanings. Their common denominator is the idea that there exist a partitioning of the network nodes into groups, or modules, that are largely independent of each other and perform separate or autonomous functions. Independence can be achieved through spatial isolation of the module's processes or by chemical specificity of its components. The ability to extract the module from the cell and reconstitute it in vitro, or transplant it to another type of cell is a powerful argument for the existence of modularity [58]. In the absence of such strong and laborious experimental verifications, however, measures of modularity that depend on a particular network model are frequently used.

In topological networks, the focus is on the module's independence: nodes within a module are densely connected to each other, while inter-modular links are sparse [57,120,142] and the tendency to cluster is measured by high clustering coefficients. As a caveat to this view note that despite their sparseness the inter-module links could represent strong dynamical couplings. Modular architecture has been studied in Boolean networks by Kashtan and Alon [79], who have shown that modularity can evolve by mutation and selection in a time-varying fitness landscape where changeable goals decompose into a set of fixed subproblems. In the example studied they computationally evolve networks implementing several Boolean formulae and observe the appearance of a module – a circuit of logical gates implementing a particular Boolean operator (like XOR) in a reusable way. This work makes clear that modularity in networks is plausibly connected to modularity in the kinds of problems that these networks were selected to solve, but we really know relatively little about the formal structure of these problems.

There are also ways of inferring a form of modularity directly without assuming any particular network model. Clustering tools partition genes into co-expressed groups, or clusters, that are often identified with particular modules [36,133,140]. Ihmels et al. [67] have noted that each node can belong to more than one module depending on the biological state of the cell, or the context, and have correspondingly reexamined the clustering problem. Elemento et al. [37] have recently presented a general information theoretic approach to inferring regulatory modules and the associated transcription factor binding sites from various kinds of high-throughput data. While clustering methods have been widely applied in the exploration of

gene expression, it should be emphasized that merely finding clusters does not by itself provide evidence for modularity. As noted above, the whole discussion would be much more satisfying if we had independent definitions of modularity and, we might add, clearly stated alternative hypotheses about the structure and dynamics of these networks.

Focusing on the functional aspect of the module, we often observe that the majority of the components of a system (for instance, a set of promoter sites or a set of genes regulating motility in bacteria) are conserved together across species. These observations support the hypothesis that the conserved components are part of a very tightly coupled sub-network which we might identify as a module. Bioinformatic tools can then use the combined sequence and expression data to give predictions about modules, as reviewed by Siggia et al. [139]. Purely phylogenetic approaches that infer module components based on inter-species comparisons have also been productive and can extract candidate modules based only on phylogenetic footprinting, that is, studying the presence or absence of homologous genes across organisms and correlating their presence with hand annotated phenotypic traits [141].

Robustness

Robustness refers to a property of the biological network such that some aspect of its function is not sensitive to perturbations of network parameters, environmental variables (e.g. temperature), or initial state; see de Visser et al. [162] for a review of robustness from an evolutionary perspective and Goulian [53] for mechanisms of robustness in bacterial circuits. Robustness encompasses two very different ideas. One idea has to do with a general principle about the nature of explanation in the quantitative sciences: qualitatively striking facts should not depend on the fine tuning of parameters, because such a scenario just shifts the problem to understanding why the parameters are tuned as they are. The second idea is more intrinsic to the function of the system, and entails the hypothesis that cells cannot rely on precisely reproducible parameters or conditions and must nonetheless function reliably and reproducibly.

Robustness has been studied extensively in the chemotactic system of the bacterium *Escherichia coli*. The systematic bias to swim towards chemoattractants and away from repellents can only be sustained if the bacterium is sensitive to the spatial gradients of the concentration and not to its absolute levels. This discriminative ability is ensured by the mechanism of perfect adaptation, with which the proportion of bacterial straight runs and tum-

bles (random changes in direction) always returns to the same value in the absence of gradients [27]. Naively, however, the ability to adapt perfectly seems to be sensitive to the amounts of intracellular signaling proteins, which can be tuned only approximately by means of transcriptional regulation. Barkai and Leibler [13] argued that there is integral feedback control in the chemotactic circuit which makes it robust against changes in these parameters, and Alon et al. [6] showed experimentally that precision of adaptation truly stays robust, while other properties of the systems (such as the time to adapt and the steady state) show marked variations with intracellular signaling protein concentrations.

One seemingly clear example of robust biological function is embryonic development. We know that the spatial structure of the fully developed organism follows a ‘blueprint’ laid out early in development as a spatial pattern of gene expression levels. von Dassow et al. [34] studied one part of this process in the fruit fly *Drosophila*, the ‘segment polarity network’ that generates striped patterns of expression. They considered a dynamical system based on the wiring diagram of interactions among a small group of genes and signaling molecules, with ~ 50 associated constants parametrizing production and degradation rates, saturation response and diffusion, and searched the parameter space for solutions that reproduce the known striped patterns. They found that, with their initial guess at network topology, such solutions do not exist, but adding a particular link – biologically motivated though unconfirmed at the time – allowed them to find solutions by random sampling of parameter space. Although they presented no rigorous measure for the volume of parameter space in which correct solutions exist, it seems that a wide variety of parameter choices and initial conditions indeed produce striped expression patterns, and this was taken to be a signature of robustness.

Robustness in dynamical models is the ability of the biological network to sustain its trajectory through state space despite parameter or state perturbations. In circadian clocks the oscillations have to be robust against both molecular noise inherent in transcriptional regulation, examined in stochastic simulations by Gonze et al. [52], as well as variation in rate parameters [143]; in the latter work the authors introduce integral robustness measures along the trajectory in state space and argue that the clock network architecture tends to concentrate the fragility to perturbations into parameters that are global to the cell (maximum overall translation and protein degradation rates) while increasing the robustness to processes specific to the circadian oscillator. As was mentioned earlier, robustness to state perturbations was demonstrated by Li et al. [91] in

the threshold binary network model of the yeast cell cycle, and examined in scale-free random Boolean networks by Aldana and Cluzel [4].

As with modularity, robustness has been somewhat resistant to rigorous definitions. Importantly, robustness has always been used as a relational concept: function X is robust to variations in Y . An alternative to robustness is for the organism to exert precise control over Y , perhaps even using X as a feedback signal. This seems to be how neurons stabilize a functional mix of different ion channels [93], following the original theoretical suggestion of LeMasson et al. [88]. Pattern formation during embryonic development in *Drosophila* begins with spatial gradients of transcription factors, such as Bicoid, which are established by maternal gene expression, and it has been assumed that variations in these expression levels are inevitable, requiring some robust readout mechanism. Recent measurements of Bicoid in live embryos, however, demonstrate that the absolute concentrations are actually reproducible from embryo to embryo with $\sim 10\%$ precision [54]. While there remain many open questions, these results suggest that organisms may be able to exert surprisingly exact control over critical parameters, rather than having compensation schemes for initially sloppy mechanisms. The example of ion channels alerts us to the possibility that cells may even ‘know’ which combinations of parameters are critical, so that variations in a multidimensional parameter space are large, but confined to a low dimensional manifold.

Noise

A dynamical system with constant reaction rates, starting repeatedly from the same initial condition in a stable environment, always follows a deterministic time evolution. When the concentrations of the reacting species are low enough, however, the description in terms of time (and possibly space) dependent concentration breaks down, and the stochasticity in reactions, driven by random encounters between individual molecules, becomes important: on repeated trials from the same initial conditions, the system will trace out different trajectories in the state space. As has been pointed out in the section on stochastic dynamics, biological networks in this regime need to be simulated with the Gillespie algorithm [47], or analyzed within approximate schemes that treat noise as perturbation of deterministic dynamics. Recent experimental developments have made it possible to observe this noise directly, spurring new research in the field. Noise in biological networks fundamentally limits the organism’s ability to sense, process and respond to environmental and

internal signals, suggesting that analysis of noise is a crucial component in any attempt to understand the design of these networks. This line of reasoning is well developed in the context of neural function [20], and we draw attention in particular to work on the ability of the visual system to count single photons, which depends upon the precision of the G-protein mediated signaling cascade in photo receptors; see, for example, [117].

Because transcriptional regulation inherently deals with molecules, such as DNA and transcription factors, that are present at low copy numbers, most noise studies were carried out on transcriptional regulatory elements. The availability of fluorescent proteins and their fusions to wild type proteins have been the crucial tools, enabling researchers to image the cells expressing these probes in a controllable manner, and track their number in time and across the population of cells. Elowitz et al. [39] pioneered the idea of observing the output of two identical regulatory elements driving the expression of two fluorescent proteins of different colors, regulated by a common input in a single *Escherichia coli* cell. In this ‘two-color experiment,’ the correlated fluctuations in both colors must be due to the *extrinsic* fluctuations in the common factors that influence the production of both proteins, such as overall RNA polymerase or transcription factor levels; on the other hand, the remaining, uncorrelated fluctuation is due to the *intrinsic* stochasticity in the transcription of the gene and translation of the messenger RNA into the fluorescent protein from each of the two promoters [147]. Ozbudak et al. [109] have studied the contributions of stochasticity in transcription and translation to the total noise in gene expression in prokaryotes, while Pedraza and van Oudenaarden [112] and Hooshangi et al. [63] have looked at the propagation of noise from transcription factors to their targets in synthetic multi-gene cascades. Rosenfeld et al. [124] have used the statistics of binomial partitioning of proteins during the division of *Escherichia coli* to convert their fluorescence measurements into the corresponding absolute protein concentrations, and also were able to observe the dynamics of these fluctuations, characterizing the correlation times of both intrinsic and extrinsic noise.

Theoretical work has primarily been concerned with disentangling and quantifying the contributions of different steps in transcriptional regulation and gene expression to the total noise in the regulated gene [111,146,149], often by looking for signatures of various noise sources in the behavior of the measured noise as a function of the mean expression level of a gene. For many of the examples studied in prokaryotes, noise seemed to be primarily attributable to the production of proteins in bursts from single messenger RNA molecules, and to pulsatile and ran-

dom activation of genes and therefore bursty translation into mRNA [50]. In yeast [26,119] and in mammalian cells [116] such stochastic synthesis of mRNA was modeled and observed as well. Simple scaling of noise with the mean was observed in ~ 40 yeast proteins under different conditions by Bar-Even et al. [9] and interpreted as originating in variability in mRNA copy numbers or gene activation.

Bialek and Setayeshgar [22] have demonstrated theoretically that at low concentrations of transcriptional regulator, there should be a lower bound on the noise set by the basic physics of diffusion of transcription factor molecules to the DNA binding sites. This limit is independent of (possibly complex, and usually unknown) molecular details of the binding process; as an example, cooperativity enhances the ‘sensitivity’ to small changes in concentration, but doesn’t lower the physical limit to noise performance [23]. This randomness in diffusive flux of factors to their ‘detectors’ on the DNA must ultimately limit the precision and reliability of transcriptional regulation, much like the randomness in diffusion of chemoattractants to the detectors on the surface of *Escherichia coli* limits its chemotactic performance [17]. Interestingly, one dimensional diffusion of transcription factors along the DNA can have a big impact on the speed with which TFs find their targets, but the change in noise performance that one might expect to accompany these kinetic changes is largely compensated by the extended correlation structure of one dimensional diffusion [152]. Recent measurements of the regulation of the *hunchback* gene by Bicoid during early fruit fly development by Gregor et al. [54] have provided evidence for the dominant role of such input noise, which coexists with previously studied output noise in production of mRNA and protein [156]. These results raise the possibility that initial decisions in embryonic development are made with a precision limited by fundamental physical principles.

Dynamics, Attractors, Stability and Large Fluctuations

The behavior of a dynamical system as the time tends to infinity, in response to a particular input, is interesting regardless of the nature of the network model. Both discrete and continuous, or deterministic and noisy, systems can settle into a number of fixed points, exhibit limit-cycle oscillations, or execute chaotic dynamics. In biological networks it is important to ask whether these qualitatively different outcomes correspond to distinct phenotypes or behaviors. If so, then a specific stable gene expression profile in a network of developmental genes, for example, encodes that cell’s developmental fate, as the amount of lambda re-

pressor encodes the state of lysis vs lysogeny switch in the phage. The history of the system that led to the establishment of a specific steady state would not matter as long as the system persisted in the same attractor: the dynamics could be regarded as a ‘computation’ leading to the final result, the identity of the attractor, with the activities of genes in this steady state in turn driving the downstream pathways and other modules; see Kauffman [80] for genetic networks and Hopfield [64] for similar ideas in neural networks for associative memory. Alternatively, such partitioning into transient dynamics and ‘meaningful’ steady states might not be possible: the system must be analyzed as a whole while it moves in state space, and parts of it do not separately and sequentially settle into their attractors.

It seems, for example, that qualitative behavior of the cell cycle can be understood by progression through well-defined states or checkpoints: after transients die away, the cell cycle proteins are in a ‘consistent’ state that regulates division or growth related activities, so long as the conditions do not warrant a new transition into the next state [33,103]. In the fruit fly *Drosophila* development it has been suggested that combined processes of diffusion and degradation first establish steady-state spatial profiles of maternal morphogens along the major axis of the embryo, after which this stable ‘coordinate system’ is read out by gap and other downstream genes to generate the body segments. Recent measurements by Gregor et al. [55] have shown that there is a rich dynamics in the Bicoid morphogens concentration, prompting Bergmann et al. [18] to hypothesize that perhaps downstream genes read out and respond to morphogens even before the steady state has been reached. On another note, an interesting excitable motif, called the “feedback resistor,” has been found in HIV Tat system – instead of having a bistable switch like the λ phage, HIV (which lacks negative feedback capability) implements a circuit with a single stable ‘off’ lysogenic state, that is perturbed in a pulse of trans activation when the virus attacks. The pulse probably triggers a threshold-crossing process that drives downstream events, and subsequently decays away; the feedback resistor is thus again an example of a dynamic, as opposed to the steady-state, readout [168]. Excitable dynamics are of course at the heart of the action potential in neurons, which results from the coupled dynamics of ion channel proteins, and related dynamical ideas are now emerging other cellular networks [145].

If attractors of the dynamical system correspond to distinct biological states of the organism, it is important to examine their stability against noise-induced spontaneous flipping. Bistable elements are akin to the ‘flip-flop’

switches in computer chips – they form the basis of cellular (epigenetic) memory. While this mechanism for remembering the past is not unique – for example, a very slow, but not bistable, dynamics will also retain ‘memory’ of the initial condition through protein levels that persist on a generation time scale [138], it has the potential to be the most stable mechanism. The naturally occurring bistable switch of the λ phage was studied using stochastic simulation by Arkin et al. [7], and a synthetic toggle switch was constructed in *Escherichia coli* by Gardner et al. [44]. Theoretical studies of systems where large fluctuations are important are generally difficult and restricted to simple regulatory elements, but Bialek [21] has shown that a bistable switch can be created with as few as tens of molecules yet remain stable for years. A full understanding of such stochastic switching brings in powerful methods from statistical physics and field theory [122,129,164], ultimately with the hope of connecting to quantitative experiments [1].

Optimization Principles

If the function of a pathway or a network module can be quantified by a scalar measure, it is possible to explore the space of networks that perform the given function optimally. An example already given was that of maximizing the growth rate of the bacterium *Escherichia coli*, subject to the constraints imposed by the known metabolic reactions of the cell; the resulting optimal joint usage of oxygen and food could be compared to the experiments [66]. If enough constraints exist for the problem to be well posed, and there is sufficient reason to believe that evolution drove the organism towards optimal behavior, optimization principles allow us to both tune the otherwise unknown parameters to achieve the maximum, and also to compare the wild type and optimal performances.

Dekel and Alon [35] have performed the cost/benefit analysis of expressing *lac* operon in bacteria. On one hand *lac* genes allow *Escherichia coli* to digest lactose, but on the other there is the incurred metabolic cost to the cell for expressing them. That the cost is not negligible to the bacterium is demonstrated best by the fact that it shuts off the operon if no lactose is present in the environment. The cost terms are measured by inducing the *lac* operon with changeable amount of IPTG that provides no energy in return; the benefit is measured by fully inducing *lac* with IPTG and supplying variable amounts of lactose; both cost and benefit are in turn expressed as the change in the growth rate compared to the wild-type grown at fixed conditions. Optimal levels of *lac* expression were then predicted as a function of lactose concentration and bacteria

were evolved for several hundred generations to verify that evolved organisms lie close to the predicted optimum.

Zaslaver et al. [172] have considered a cascade of amino-acid biosynthesis reactions in *Escherichia coli*, catalyzed by their corresponding enzymes. They have then optimized the parameters of the model that describes the regulation of enzyme gene expression, such that the total metabolic cost for enzyme production was balanced against the benefit of achieving a desired metabolic flux through the biosynthesis pathway. The resulting optimal on-times and promoter activities for the enzymes were compared to the measured activities of amino-acid biosynthesis promoters exposed to different amino-acids in the medium. The authors conclude that the bacterium implements a ‘just-in-time’ transcription program, with enzymes catalyzing initial steps in the pathway being produced from strong and low-latency promoters.

In signal transduction networks the definition of an objective function to be maximized is somewhat more tricky. The ability of the cell to sense its environment and make decisions, for instance about which genes to up- or down-regulate, is limited by several factors: scarcity of signals coming from the environment, perhaps because of the limited time that can be dedicated to data collection; noise inherent in the signaling network that degrades the quality of the detected signal; (sub-)optimality of the decision strategy; and noise in the effector systems at the output. A first idea would be to postulate that networks are designed to lower the noise, and intuitively the ubiquity of mechanisms such as negative feedback [15,53] is consistent with such an objective. There are various definitions for noise, however, which in addition are generally a function of the input, raising serious issues about how to formulate a principled optimization criterion.

When we think about energy flow in biological systems, there is no doubt that our thinking must at least be consistent with thermodynamics. More strongly, thermodynamics provides us with notions of efficiency that place the performance of biological systems on an absolute scale, and in many cases this performance really is quite impressive. In contrast, most discussions of information in biological systems leave “information” as a colloquial term, making no reference to the formal apparatus of information theory as developed by Shannon and others more than fifty years ago [135]. Although many aspects of information theory that are especially important for modern technology (e.g., sophisticated error-correcting codes) have no obvious connection to biology, there is something at the core of information theory that is vital: Shannon proved that if we want to quantify the intuitive concept that “ x provides information about y ,” then there

is only one way to do this that is guaranteed to work under all conditions and to obey simple intuitive criteria such as the additivity of independent information. This unique measure of “information” is Shannon’s mutual information. Further, there are theorems in information theory which, in parallel to results in thermodynamics, provide us with limits to what is possible and with notions of efficiency.

There is a long history of using information theoretic ideas to analyze the flow of information in the nervous system, including the idea that aspects of the brain’s coding strategies might be chosen to optimize the efficiency of coding, and these theoretical ideas have led directly to interesting experiments. The use of information to think about cellular signaling and its possible optimization is more recent [154,175]. An important aspect of optimizing information flow is that the input/output relations of signaling devices must be matched to the distribution of inputs, and recent measurements on the control of *hunchback* by Bicoid in the early fruit fly embryo [54] seem remarkably consistent with the (parameter free) predictions from these matching relations [155].

In the context of neuroscience there is a long tradition of forcing the complex dynamics of signal processing into a setting where the subject needs to decide between a small set of alternatives; in this limit there is a well developed theory of optimal Bayesian decision making, which uses prior knowledge of the possible signals to help overcome noise intrinsic to the signaling system; Libby et al. [92] have recently applied this approach to the *lac* operon in *Escherichia coli*. The regulatory element is viewed as an inference module that has to ‘decide,’ by choosing its induction level, if the environmental lactose concentration is high or low. If the bacterium detects a momentarily high sugar concentration, it has to discriminate between two situations: either the environment really is at low overall concentration but there has been a large fluctuation; or the environment has switched to a high concentration mode. The authors examine how plausible regulatory element architectures (e. g. activator vs repressor, cooperative binding etc.) yield different discrimination performance. Intrinsic noise in the *lac* system can additionally complicate such decision making, but can be included into the theoretical Bayesian framework.

The question of whether biological systems are optimal in any precise mathematical sense is likely to remain controversial for some time. Currently opinions are stronger than the data, with some investigators using ‘optimized’ rather loosely and others convinced that what we see today is only a historical accident, not organizable around such lofty principles. We emphasize, however, that attempts to

formulate optimization principles require us to articulate clearly what we mean by “function” in each context, and this is an important exercise. Exploration of optimization principles also exposes new questions, such as the nature of the distribution of inputs to signaling systems, that one might not have thought to ask otherwise. Many of these questions remain as challenges for a new generation of experiments.

Evolvability and Designability

Kirschner and Gerhart [82] define *evolvability* as an organism’s capacity to generate heritable phenotypic variation. This capacity may have two components: first, to reduce the lethality of mutations, and second, to reduce the number of mutations needed to produce phenotypically novel traits. The systematic study of evolvability is hard because the genotype-to-phenotype map is highly non-trivial, but there have been some qualitative observations relevant to biological networks. Emergence of *weak linkage* of processes, such as the co-dependence of transcription factors and their DNA binding sites in metazoan transcriptional regulation, is one example. Metazoan regulation seems to depend on combinatorial control by many transcription factors with weak DNA-binding specificities and the corresponding binding sites (called cis-regulatory modules) can be dispersed and extended on the DNA. This is in stark contrast to the strong linkage between the factors and the DNA in prokaryotic regulation or in metabolism, energy transfer or macromolecular assembly, where steric and complementarity requirements for interacting molecules are high. In protein signaling networks, strongly conserved but flexible proteins, like calmodulin, can bind weakly to many other proteins, with small mutations in their sequence probably affecting such binding and making the establishment of new regulatory links possible and perhaps easy.

Some of the most detailed attempts to follow the evolution of network function have been by Francois and coworkers [41,42]. In their initial work they showed how simple functional circuits, performing logical operations or implementing bistable or oscillatory behavior, can be reliably created by a mutational process with selection by an appropriate fitness function. More recently they have considered fitness functions which favor spatial structure in patterns of gene expression, and shown how the networks that emerge from dynamics in this fitness landscape recapitulate the outlines of the segmentation networks known to be operating during embryonic development.

Instead of asking if there *exists* a network of nodes such that they perform a given computation, and if it can be

found by mutation and selection as in the examples above, one can ask how many network topologies perform a given computation. In other words, one is asking whether there is only one (fine tuned?) or many topologies or solutions to a given problem. The question of how many network topologies, proxies for different genotypes, produce the same dynamics, a proxy for phenotype, is a question of designability, a concept originally proposed to study the properties of amino-acid sequences comprising functional proteins, but applicable also to biological regulatory networks [106]. The authors examine three- and four-node binary networks with threshold updating rule and show that all networks with the shared phenotype have a common 'core' set of connections, but can differ in the variable part, similar to protein folding where the essential set of residues is necessary for the fold, with numerous variations in the nonessential part.

Future Directions

The study of biological networks is at an early stage, both on the theoretical as well as on the experimental side. Although high-throughput experiments are generating large data sets, these can suffer from serious biases, lack of temporal or spatial detail, and limited access to the component parts of the interacting system. On a theoretical front, general analytical insights that would link dynamics with network topology are few, although for specific systems with known topology computer simulation can be of great assistance. There can be confusion about which aspects of the dynamical model have biological significance and interpretation, and which aspects are just 'temporary variables' and the 'envelope' of the proverbial back-of-the-envelope calculations that cells use to perform their biological computations on; which parts of the trajectory are functionally constrained and which ones could fluctuate considerably with no ill-effects; how much noise is tolerable in the nodes of the network and what is its correlation structure; or how the unobserved, or 'hidden,' nodes (or their modification/activity states) influence the network dynamics.

Despite these caveats, cellular networks have some advantages over biological systems of comparable complexity, such as neural networks. Due to technological developments, we are considerably closer to the complete census of the interacting molecules in a cell than we are generally to the picture of connectivity of the neural tissue. Components of the regulatory networks are simpler than neurons, which are capable of a range of complicated behaviors on different timescales. Modules and pathways often comprise smaller number of interacting elements than in

neural networks, making it possible to design small but interesting synthetic circuits. Last but not least, sequence and homology can provide strong insights or be powerful tools for network inference in their own right.

Those of us who come from the traditionally quantitative sciences, such as physics, were raised with experiments in which crucial elements are isolated and controlled. In biological systems, attempts at such isolation may break the regulatory mechanisms that are essential for normal operation of the system, leaving us with a system which is in fact more variable and less controlled than we would have if we faced the full complexity of the organism. It is only recently that we have seen the development of experimental techniques that allow fully quantitative, real time measurements of the molecular events inside individual cells, and the theoretical framework into which such measurements will be fit still is being constructed. The range of theoretical approaches being explored is diverse, and it behooves us to search for those approaches which have the chance to organize our understanding of many different systems rather than being satisfied with models of particular systems. Again, there is a balance between the search for generality and the need to connect with experiments on specific networks. We have tried to give some examples of all these developments, hopefully conveying the correct combination of enthusiasm and skepticism.

Acknowledgments

We thank our colleagues and collaborators who have helped us learn about these issues: MJ Berry, CG Callan, T Gregor, JB Kinney, P Mehta, SE Palmer, E Schneidman, JJ Hopfield, T Mora, S Setayeshgar, N Slonim, GJ Stephens, DW Tank, N Tishby, A Walczak, EF Wieschaus, CH Wiggins and NS Wingreen. Our work was supported in part by NIH grants P50 GM071508 and R01 GM077599, by NSF Grants IIS-0613435 and PHY-0650617, by the Swartz Foundation, and by the Burroughs Wellcome Fund.

Bibliography

1. Acar M, Becksei A, van Oudenaarden A (2005) Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435:228–232
2. Ackers GK, Johnson AD, Shea MA (1982) Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci (USA)* 79(4):1129–33
3. Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. *Nature* 406(6794):378–82
4. Aldana M, Cluzel P (2003) A natural class of robust networks. *Proc Natl Acad Sci (USA)* 100:8710–4
5. Alm E, Arkin AP (2003) Biological networks. *Curr Opin Struct Biol* 13(2):193–202

6. Alon U, Surette MG, Barkai N, Leibler S (1999) Robustness in bacterial chemotaxis. *Nature* 397(6715):168–71
7. Arkin A, Ross J, McAdams HH (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 149(4):1633–48
8. Arnosti DN, Kulkarni MM (2005) Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* 94(5):890–8
9. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N (2006) Noise in protein expression scales with natural protein abundance. *Nat Genet* 38(6):636–43
10. Barabási AL (2002) *Linked: The New Science of Networks*. Perseus Publishing, Cambridge
11. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–12
12. Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5(2):101–13
13. Barkai N, Leibler S (1997) Robustness in simple biochemical networks. *Nature* 387(6636):913–7
14. Baylor DA, Lamb TD, Yau KW (1979) Responses of retinal rods to single photons. *J Physiol (Lond)* 288:613–634
15. Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. *Nature* 405(6786):590–3
16. Berg HC (1975) Chemotaxis in bacteria. *Annu Rev Biophys Bioeng* 4(00):119–36
17. Berg HC, Purcell EM (1977) Physics of chemoreception. *Biophys J* 20(2):193–219
18. Bergmann S, Sandler O, Sberro H, Shnider S, Schejter E, Shilo BZ, Barkai N (2007) Pre-steady-state decoding of the bicoid morphogen gradient. *PLoS Biol* 5(2):e46
19. Bethe HA (1935) Statistical theory of superlattices. *Proc R Soc London Ser A* 150:552–575
20. Bialek W (1987) Physical limits to sensation and perception. *Annu Rev Biophys Biochem* 16:455–78
21. Bialek W (2001) Stability and noise in biochemical switches. *Adv Neur Info Proc Syst* 13:103
22. Bialek W, Setayeshgar S (2005) Physical limits to biochemical signaling. *Proc Natl Acad Sci (USA)* 102(29):10040–5
23. Bialek W, Setayeshgar S (2006) Cooperativity, sensitivity and noise in biochemical signaling. arXiv.org:q-bio.MN/0601001
24. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R (2005a) Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev* 15(2):125–35
25. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Phillips R (2005b) Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev* 15(2):116–24
26. Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* 422(6932):633–7
27. Block SM, Segall JE, Berg HC (1983) Adaptation kinetics in bacterial chemotaxis. *J Bacteriol* 154(1):312–23
28. Bray D (1995) Protein molecules as computational elements in living cells. *Nature* 376(6538):307–12
29. Bray D, Bourret RB, Simon MI (1993) Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol Biol Cell* 4(5):469–82
30. Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21(1 Suppl):33–7
31. Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. *Proc Natl Acad Sci (USA)* 100(9):5136–41
32. Chang L, Karin M (2001) Mammalian map kinase signalling cascades. *Nature* 410(6824):37–40
33. Chen KC, Csikasz-Nagy A, Gyorffy B, Val J, Novak B, Tyson JJ (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol Biol Cell* 11(1):369–91
34. von Dassow G, Meir E, Munro EM, Odell GM (2000) The segment polarity network is a robust developmental module. *Nature* 406(6792):188–92
35. Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436(7050):588–92
36. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci (USA)* 95(25):14863–8
37. Elemento O, Slonim N, Tavazoie S (2007) A universal framework for regulatory element discovery across all genomes and data types. *Mol Cell* 28(2):337–50
38. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403(6767):335–8
39. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297(5584):1183–6
40. Falke JJ, Bass RB, Butler SL, Chervitz SA, Danielson MA (1997) The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu Rev Cell Dev Biol* 13:457–512
41. Francois P, Hakim V (2004) Design of genetic networks with specified functions by evolution in silico. *Proc Natl Acad Sci (USA)* 101(2):580–5
42. Francois P, Hakim V, Siggia ED (2007) Deriving structure from evolution: metazoan segmentation. *Mol Syst Bio* 3: Article 154
43. Friedman N (2004) Inferring cellular networks using probabilistic graphical models. *Science* 303(5659):799–805
44. Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403(6767):339–42
45. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415(6868):141–7
46. Ghaemmighami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* 425(6959):737–41
47. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361
48. Gillespie DT (2007) Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem* 58:35–55
49. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamodar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrola S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley J R L, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets

- RA, McKenna MP, Chant J, Rothberg JM (2003) A protein interaction map of *Drosophila melanogaster*. *Science* (5651): 1727–36
50. Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* 123(6): 1025–36
51. Goldman MS, Golowasch J, Marder E, Abbott LF (2001) Global structure robustness and modulation of neural models. *J Neurosci* 21:5229–5238
52. Gonze D, Halloy J, Goldbeter A (2002) Robustness of circadian rhythms with respect to molecular noise. *Proc Natl Acad Sci (USA)* 99(2):673–8
53. Goulian M (2004) Robust control in bacterial regulatory circuits. *Curr Opin Microbiol* 7(2):198–202
54. Gregor T, Tank DW, Wieschaus EF, Bialek W (2007a) Probing the limits to positional information. *Cell* 130(1):153–64
55. Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW (2007b) Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* 130(1):141–52
56. Guet CC, Elowitz MB, Hsing W, Leibler S (2002) Combinatorial synthesis of genetic networks. *Science* 296(5572):1466–70
57. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430(6995): 88–93
58. Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* 402(6761 Suppl):C47–52
59. Hasty J, McMillen D, Isaacs F, Collins JJ (2001) Computational studies of gene regulatory networks: in numero molecular biology. *Nat Rev Genet* 2(4):268–79
60. Heinrich R, Schuster S (1996) *The Regulation of Cellular Systems*. Chapman and Hall, New York
61. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreau M, Muskut B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415(6868):180–3
62. Hofman J, Wiggins C (2007) A bayesian approach to network modularity. *arXiv.org*:07093512
63. Hooshangi S, Thiberge S, Weiss R (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci (USA)* 102(10):3581–6
64. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci (USA)* 79(8):2554–8
65. Huang KC, Meir Y, Wingreen NS (2003) Dynamic structures in *Escherichia coli*: spontaneous formation of mine rings and mind polar zones. *Proc Natl Acad Sci (USA)* 100(22):12724–8
66. Ibarra RU, Edwards JS, Palsson BO (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420(6912):186–9
67. Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N (2002) Revealing modular organization in the yeast transcriptional network. *Nat Genet* 31(4):370–7
68. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci (USA)* 98(8):4569–74
69. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 3:318–56
70. Jansen R, Gerstein M (2004) Analyzing protein function on a genomic scale: the importance of gold-standard positives and negatives for network prediction. *Curr Opin Microbiol* 7(5):535–45
71. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 106:62–79
72. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. *Nature* 407(6804):651–4
73. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411(6833):41–2
74. Jordan JD, Landau EM, Iyengar R (2000) Signaling networks: the origins of cellular multitasking. *Cell* 103(2):193–200
75. Kadonaga JT (2004) Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* 116(2):247–57
76. van Kampen NG (2007) *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam
77. Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) The KEGG databases at genomet. *Nucleic Acids Res* 30(1):42–6
78. Karp PD, Riley M, Saier M, Paulsen IT, Collado-Vides J, Paley SM, Pellegrini-Toole A, Bonavides C, Gama-Castro S (2002) The ecocyc database. *Nucleic Acids Res* 30(1):56–8
79. Kashtan N, Alon U (2005) Spontaneous evolution of modularity and network motifs. *Proc Natl Acad Sci (USA)* 102(39):13773–8
80. Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 22(3):437–67
81. Keller EF (2005) Revisiting "scale-free" networks. *Bioessays* 27(10):1060–8
82. Kirschner M, Gerhart J (1998) Evolvability. *Proc Natl Acad Sci (USA)* 95(15):8420–7
83. Kolch W (2000) Meaningful relationships: the regulation of the ras/raf/mek/erk pathway by protein interactions. *Biochem J* 351 Pt 2:289–305
84. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrin-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandhi K, Thompson NJ, Musso G, St Onge P, Ghanny S, Lam MH, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440(7084):637–43
85. Kuhlman T, Zhang Z, Saier J M H, Hwa T (2007) Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc Natl Acad Sci (USA)* 104(14):6043–8
86. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young

- RA (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298(5594):799–804
87. Leloup JC, Goldbeter A (2003) Toward a detailed computational model for the mammalian circadian clock. *Proc Natl Acad Sci (USA)* 100(12):7051–6
88. LeMasson G, Marder E, Abbott LF (1993) Activity-dependent regulation of conductances in model neurons. *Science* 259:1915–1917
89. Levine M, Davidson EH (2005) Gene regulatory networks for development. *Proc Natl Acad Sci (USA)* 102(14):4936–42
90. Lezon TR, Banavar JR, Cieplak M, Maritan A, Federoff NV (2006) Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc Natl Acad Sci (USA)* 103:19033–19038
91. Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci (USA)* 101(14):4781–6
92. Libby E, Perkins TJ, Swain PS (2007) Noisy information processing through transcriptional regulation. *Proc Natl Acad Sci (USA)* 104(17):7151–6
93. Marder E, Bucher D (2006) Variability, compensation and homeostasis in neuron and network function. *Nature Rev Neurosci* 7:563–574
94. Markowitz F, Spang R (2007) Inferring cellular networks – a review. *BMC Bioinformatics* 8: 6–55
95. Martin DE, Hall MN (2005) The expanding tor signaling network. *Curr Opin Cell Biol* 17(2):158–66
96. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. *Science* 296(5569):910–3
97. McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. *Proc Natl Acad Sci (USA)* 94(3):814–9
98. McAdams HH, Arkin A (1999) It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet* 15(2):65–9
99. McAdams HH, Shapiro L (1995) Circuit simulation of genetic networks. *Science* 269(5224):650–6
100. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417(6887):399–403
101. Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M, Alon U (2004) Superfamilies of evolved and designed networks. *Science* 303(5663):1538–42
102. Nakajima M, Imai K, Ito H, Nishiwaki T, Murayama Y, Iwasaki H, Oyama T, Kondo T (2005) Reconstitution of circadian oscillation of cyanobacterial *kaic* phosphorylation in vitro. *Science* 308:414–415
103. Nasmyth K (1996) At the heart of the budding yeast cell cycle. *Trends Genet* 12(10):405–12
104. Newman M, Watts D, Barabási AL (2006) *The Structure and Dynamics of Networks*. Princeton University Press, Princeton
105. Nielsen UB, Cardone MH, Sinskey AJ, MacBeath G, Sorger PK (2003) Profiling receptor tyrosine kinase activation by using ab microarrays. *Proc Natl Acad Sci (USA)* 100(16):9330–5
106. Nochomovitz YD, Li H (2006) Highly designable phenotypes and mutational buffers emerge from a systematic mapping between network topology and dynamic output. *Proc Natl Acad Sci (USA)* 103(11):4180–5
107. Novak B, Tyson JJ (1997) Modeling the control of DNA replication in fission yeast. *Proc Natl Acad Sci (USA)* 94(17): 9147–52
108. Nurse P (2001) Cyclin dependent kinases and cell cycle control. *Les Prix Nobel*
109. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31(1):69–73
110. Papin JA, Hunter T, Palsson BO, Subramaniam S (2005) Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol* 6(2):99–111
111. Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427(6973):415–8
112. Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* 307(5717):1965–9
113. Perez OD, Nolan GP (2002) Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. *Nat Biotechnol* 20(2):155–62
114. Ptashne M (2001) *Genes and Signals*. CSHL Press, Cold Spring Harbor, USA
115. Ptashne M (2004) *A Genetic Switch: Phage lambda Revisited*. CSHL Press
116. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4(10):e309
117. Ramanathan S, Detwiler PB, Sengupta AM, Shraiman BI (2005) G-protein-coupled enzyme cascades have intrinsic properties that improve signal localization and fidelity. *Biophys J* 88(5):3063–71
118. Rao CV, Kirby JR, Arkin AP (2004) Design and diversity in bacterial chemotaxis: a comparative study in *Escherichia coli* and *Bacillus subtilis*. *PLoS Biol* 2(2):E49
119. Raser JM, O'Shea EK (2005) Noise in gene expression: origins, consequences, and control. *Science* 309(5743):2010–3
120. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297(5586):1551–5
121. Rieke F, Baylor DA (1998) Single photon detection by rod cells of the retina. *Rev Mod Phys* 70:1027–1036
122. Roma DM, O'Flanagan R, Ruckenstein AE, Sengupta AM (2005) Optimal path to epigenetic switching. *Phys Rev E* 71: 011902
123. Rosenfeld N, Alon U (2003) Response delays and the structure of transcription networks. *J Mol Biol* 329(4):645–54
124. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307(5717):1962–5
125. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Towards a proteome-scale map of the human protein–protein interaction network. *Nature* 437(7062): 1173–8
126. Rust MJ, Markson JS, Lane WS, Fisher DS, O'Shea EK (2007) Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science* 318:809–812
127. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308(5721):523–9

128. Sanchez L, Thieffry D (2001) A logical analysis of the *Drosophila* gap-gene system. *J Theor Biol* 211(2):115–41
129. Sasai M, Wolynes PG (2003) Stochastic gene expression as a many-body problem. *Proc Natl Acad Sci (USA)* 100(5):2374–9
130. Schneidman E, Still S, Berry II MJ, Bialek W (2003) Network information and connected correlations. *Phys Rev Lett* 91(23):238701
131. Schneidman E, Berry II MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440(7087):1007–12
132. Schroeder MD, Pearce M, Fak J, Fan H, Unnerstall U, Emberly E, Rajewsky N, Siggia ED, Gaul U (2004) Transcriptional control in the segmentation gene network of *Drosophila*. *PLoS Biol* 2(9):E271
133. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, Friedman N (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34(2):166–76
134. Setty Y, Mayo AE, Surette MG, Alon U (2003) Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci (USA)* 100(13):7702–7
135. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423 & 623–656
136. Shen-Orr SS, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 31(1):64–8
137. Shlens J, Field GD, Gauthier JL, Grivich MI, Petrusca D, Sher A, Litke AM, Chichilnisky EJ (2006) The structure of multi-neuron firing patterns in primate retina. *J Neurosci* 26(32):8254–66
138. Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, Rosenfeld N, Danon T, Perzov N, Alon U (2006) Variability and memory of protein levels in human cells. *Nature* 444(7119):643–6
139. Siggia ED (2005) Computational methods for transcriptional regulation. *Curr Opin Genet Dev* 15(2):214–21
140. Slonim N, Atwal GS, Tkačik G, Bialek W (2005) Information-based clustering. *Proc Natl Acad Sci (USA)* 102(51):18297–302
141. Slonim N, Elemento O, Tavazoie S (2006) Ab initio genotype-phenotype association reveals intrinsic modularity in genetic networks. *Mol Syst Biol* 2 (2006) 0005
142. Spirin V, Mirny LA (2003) Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci (USA)* 100(21):12123–8
143. Stelling J, Gilles ED, Doyle 3rd FJ (2004) Robustness properties of circadian clock architectures. *Proc Natl Acad Sci (USA)* 101(36):13210–5
144. Strogatz SH (2001) Exploring complex networks. *Nature* 410(6825):268–76
145. Süel GM, Garcia-Ojalvo J, Liberman L, Elowitz MB (2006) An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440:545–550
146. Swain PS (2004) Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J Mol Biol* 344(4):965–76
147. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci (USA)* 99(20):12795–800
148. Tanay A, Regev A, Shamir R (2005) Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc Natl Acad Sci (USA)* 102(20):7203–8
149. Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci (USA)* 98(15):8614–9
150. Thomas R (1973) Boolean formalization of genetic control circuits. *J Theor Biol* 42(3):563–85
151. Tkačik G (2007) Information Flow in Biological Networks. Dissertation, Princeton University, Princeton
152. Tkačik G, Bialek W (2007) Diffusion, dimensionality and noise in transcriptional regulation. [arXiv.org:07121852](https://arxiv.org/abs/07121852) [q-bio.MN]
153. Tkačik G, Schneidman E, Berry II MJ, Bialek W (2006) Ising models for networks of real neurons. [arXiv.org:q-bio.NC/0611072](https://arxiv.org/abs/q-bio.NC/0611072)
154. Tkačik G, Callan Jr CG, Bialek W (2008) Information capacity of genetic regulatory elements. *Phys Rev E* 78:011910. [arXiv.org:0709.4209](https://arxiv.org/abs/0709.4209) [q-bio.MN]
155. Tkačik G, Callan Jr CG, Bialek W (2008) Information flow and optimization in transcriptional regulation. *Proc Natl Acad Sci* 105(34):12265–12270. [arXiv.org:0705.0313](https://arxiv.org/abs/0705.0313) [q-bio.MN]
156. Tkačik G, Gregor T, Bialek W (2008) The role of input noise in transcriptional regulation. *PLoS One* 3, e2774 [arXiv.org:q-bio.MN/0701002](https://arxiv.org/abs/q-bio.MN/0701002)
157. Tomita J, Nakajima M, Kondo T, Iwasaki H (2005) No transcription-translation feedback in circadian rhythm of kaic phosphorylation. *Science* 307:251–254
158. Tucker CL, Gera JF, Uetz P (2001) Towards an understanding of complex protein networks. *Trends Cell Biol* 11(3):102–6
159. Tyson JJ, Chen K, Novak B (2001) Network dynamics and cell physiology. *Nat Rev Mol Cell Biol* 2(12):908–16
160. Tyson JJ, Chen KC, Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* 15(2):221–31
161. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamar G, Yang M, Johnston M, Fields S, Rothberg JM (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403(6770):623–7
162. de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, Gibson G, Hansen TF, Krakauer D, Lewontin RC, Ofria C, Rice SH, von Dassow G, Wagner A, Whitlock MC (2003) Perspective: Evolution and detection of genetic robustness. *Evolution Int J Org Evolution* 57(9):1959–72
163. Wagner A, Fell DA (2001) The small world inside large metabolic networks. *Proc Biol Sci* 268(1478):1803–10
164. Walczak AM, Sasai M, Wolynes PG (2005) Self-consistent proteomic field theory of stochastic gene switches. *Biophys J* 88(2):828–50
165. Waters CM, Bassler BL (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* 21:319–46
166. Watson JD, Baker TA, Belin SP, Gann A, Levine M, Losick R (2003) *Molecular Biology of the Gene*: 5th edn. Benjamin Cummings, Menlo Park
167. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–2
168. Weinberger LS, Shenk T (2007) An hiv feedback resistor: auto-regulatory circuit deactivator and noise buffer. *PLoS Biol* 5(1):e9
169. Yeager-Lotem E, Sattath S, Kashtan N, Itzkovitz S, Milo R, Pinter RY, Alon U, Margalit H (2004) Network motifs in integrated cellular networks of transcription-regulation and protein-

- protein interaction. *Proc Natl Acad Sci (USA)* 101(16):5934–5939
170. Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Nat Acad Sci (USA)* 99(26):16587–91
 171. Young MW, Kay SA (2001) Time zones: a comparative genetics of circadian clocks. *Nat Rev Genet* 2(9):702–15
 172. Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U (2004) Just-in-time transcription program in metabolic pathways. *Nat Genet* 36(5):486–91
 173. Ziv E, Koytcheff R, Middendorf M, Wiggins C (2005a) Systematic identification of statistically significant network measures. *Phys Rev E* 71:016110
 174. Ziv E, Middendorf M, Wiggins C (2005b) Information theoretic approach to network modularity. *Phys Rev E* 71:046117
 175. Ziv E, Nemenman I, Wiggins CH (2006) Optimal signal processing in small stochastic biochemical networks. *arXiv.org:q-bio/0612041*

Cellular Automata as Models of Parallel Computation

THOMAS WORSCH

Lehrstuhl Informatik für Ingenieure
und Naturwissenschaftler, Universität Karlsruhe,
Karlsruhe, Germany

Article Outline

Glossary

Definition of the Subject

Introduction

Time and Space Complexity

Measuring and Controlling the Activities

Communication in CA

Future Directions

Bibliography

Glossary

Cellular automaton The classical fine-grained parallel model introduced by John von Neumann.

Hyperbolic cellular automaton A cellular automaton resulting from a tessellation of the hyperbolic plane.

Parallel Turing machine A generalization of Turing's classical model where several control units work cooperatively on the same tape (or set of tapes).

Time complexity Number of steps needed for computing a result. Usually a function $t: \mathbb{N}_+ \rightarrow \mathbb{N}_+$, $t(n)$ being the maximum ("worst case") for any input of size n .

Space complexity Number of cells needed for computing a result. Usually a function $s: \mathbb{N}_+ \rightarrow \mathbb{N}_+$, $s(n)$ being the maximum for any input of size n .

State change complexity Number of proper state changes of cells during a computation. Usually a function $sc: \mathbb{N}_+ \rightarrow \mathbb{N}_+$, $sc(n)$ being the maximum for any input of size n .

Processor complexity Maximum number of control units of a parallel Turing machine which are simultaneously active during a computation. Usually a function $sc: \mathbb{N}_+ \rightarrow \mathbb{N}_+$, $sc(n)$ being the maximum for any input of size n .

\mathbb{N}_+ The set $\{1, 2, 3, \dots\}$ of positive natural numbers.

\mathbb{Z} The set $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ of integers.

Q^G The set of all (total) functions from a set G to a set Q .

Definition of the Subject

This article will explore the properties of cellular automata (CA) as a parallel model.

The Main Theme

We will first look at the standard model of CA and compare it with Turing machines as the standard sequential model, mainly from a computational complexity point of view. From there we will proceed in two directions: by removing computational power and by adding computational power in different ways in order to gain insight into the importance of some ingredients of the definition of CA.

What Is Left Out

There are topics which we will not cover although they would have fit under the title.

One such topic is *parallel algorithms* for CA. There are algorithmic problems which make sense only for parallel models. Probably the most famous for CA is the so-called *Firing Squad Synchronization Problem*. This is the topic of Umeo's article (► [Firing Squad Synchronization Problem in Cellular Automata](#)), which can also be found in this encyclopedia.

Another such topic in this area is the Leader election problem. For CA it has received increased attention in recent years. See the paper by Stratmann and Worsch [29] and the references therein for more details.

And we do want to mention the most exciting (in our opinion) CA algorithm: Tougne has designed a CA which, starting from a single point, after t steps has generated the discretized circle of radius t , for all t ; see [5] for this gem.

There are also models which generalize standard CA by making the cells more powerful. Kutrib has introduced *push-down cellular automata* [14]. As the name indicates,



Dynamical Stabilization of an Unstable Equilibrium in Chemical and Biological Systems

H. MALCHOW*

Institute of Environmental Systems Research, Department of Mathematics and Computer Science
University of Osnabrück, Artilleriestr. 34, D-49069 Osnabrück, Germany
malchow@uos.de

S. V. PETROVSKII

Shirshov Institute of Oceanology, Russian Academy of Sciences
Nakhimovsky Prospect 36, Moscow 117218, Russia
spetrovs@sio.rssi.ru

Abstract—The dynamics of two-component diffusion-reaction systems is considered. Using well-known models from population dynamics and chemical physics, it is shown that for certain parameter values the systems exhibit a rather unusual behaviour: a locally unstable equilibrium may become stable during a certain transition process. Both the analytical and numerical investigations of this phenomenon are presented in one and two spatial dimensions. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords—Reaction-diffusion systems, Wave propagation, Predator-prey model, Gray-Scott model.

1. INTRODUCTION

Diffusion-reaction systems have been attracting significant interest during the last decades because of their numerous applications in chemistry and chemical physics [1], biology [2], ecology [3,4], and many other scientific fields [5]. Particularly, instabilities and related transition processes are thought to be responsible for the formation of spatial “dissipative structures” in a chemical reactor [6], in a cell community (morphogenesis) [7], and in population dynamics [8].

Now days, a number of different theoretical tools for the investigation of pattern formation processes are known [9–11]. However, a widely used analytical technique is still based on partial differential equations. In many cases, the dynamics of a diffusion-reaction system is described by the following two equations:

$$\frac{\partial u(\mathbf{r}, t)}{\partial t} = D_u \nabla^2 u(\mathbf{r}, t) + f(u, v), \quad (1)$$

$$\frac{\partial v(\mathbf{r}, t)}{\partial t} = D_v \nabla^2 v(\mathbf{r}, t) + g(u, v), \quad (2)$$

*Author to whom all correspondence should be addressed.

This work is partially supported by INTAS Grant No. 96-2033, by NATO Linkage Grant No. OUTRG.LG971248, by DFG Grant No. 436 RUS 113/447, and by RFBR Grant No. 98-04-04065.

where t is the time, \mathbf{r} is the position, ∇^2 is the Laplace operator, and the functions f and g describe the local kinetics. Here and further on, we will refer to the dynamical variables u and v as the concentrations of the interacting components and to the coefficients D_u and D_v as the corresponding diffusivities. It should be noted, however, that the particular meaning of the quantities in equations (1),(2) can be somewhat different in different problems. While in chemical applications D_u and D_v are the usual molecular diffusivities, in problems of population dynamics these coefficients describe the intensity of mixing either due to the self-motion of animals [3,12] or due to turbulence, e.g., in case of plankton populations [3,13,14].

The dynamics of system (1),(2) is to a large extent controlled by the properties of the “reduced” system, i.e., equations (1),(2) without diffusion

$$\dot{u} = f(u, v), \quad \dot{v} = g(u, v), \quad (3)$$

because of the evident relation between the stationary solutions of equations (3) and the homogeneous stationary states of the full system (1),(2). However, the dynamics of the full system is remarkably more rich. Since the famous paper of Turing [7], it is well known that a linearly stable stationary point of the reduced system (3) may become unstable in the full system (1),(2). Then, after the homogeneity is broken due to the linear Turing instability, the nonlinear interactions between the components drive the system into the formation of standing spatial patterns [6]. This is an irreversible process; i.e., the broken homogeneity is never restored unless the parameters of the system are changed so that, at least, the instability conditions are not met anymore.

In a somewhat more general sense, a kind of inverse process may occur. We show here that for certain parameter values, an “anti-Turing” phenomenon takes place: a locally unstable equilibrium of the system (3) can be made dynamically stable in the full diffusion-reaction system (1),(2). In this case, for certain times and lengths, the formation of spatial patterns is suppressed and the homogeneity is restored.

The structure of the paper is as follows. In the second section, an example of a biological system is given exhibiting the dynamical stabilization of an unstable equilibrium. Both analytical and numerical results are presented. In the third section, the results are extended to the 2-D case. In the fourth section, a chemical system described by the well-known Gray-Scott model is considered. It is shown that, in spite of significantly different local kinetics of the system, its spatiotemporal dynamics can also follow the dynamical stabilization scenario. In the last section, some open problems arising in connection with this new phenomenon, are discussed and an ecological example is given where the dynamical stabilization may be underlying the system dynamics.

2. A BIOLOGICAL SYSTEM: A PREY-PREDATOR COMMUNITY

Population dynamics is one of the fields of traditional and successful applications of diffusion-reaction systems [2,4,8,14,15]. Although the spatial mixing of the system components, i.e., biological species in this case, is typically caused by the self-motion of the organisms or by the specific properties of the environment (e.g., marine turbulence in case of plankton systems) and not by diffusion in the usual physicochemical meaning, the mathematical description of the mixing stays much the same [3]. Choosing a proper parameterization for the biological “reactions”, i.e., for the processes of replication, predation, and mortality, one can arrive at the following equations for the key species, cf. [2,4]:

$$\frac{\partial u}{\partial t} = \nabla^2 u + u(1 - u) - \frac{u}{u + h}v, \quad (4)$$

$$\frac{\partial v}{\partial t} = \nabla^2 v + k \frac{u}{u + h}v - mv. \quad (5)$$

Dimensionless quantities have been introduced, and the details are omitted here in order to be brief; cf. [16]. $D_u = D_v = D$ is suggested for simplicity; however, the main results do not depend on this assumption.

A crucial point, affecting the type of the spatiotemporal system dynamics, is the choice of the initial conditions. Actually, the formation of the dissipative patterns resulting from Turing instability occurs when the initial distribution of the concentrations consists of the homogeneous stationary states *plus* a perturbation with certain wavelengths; otherwise no spatial structure will emerge. Another example is given by the propagation of diffusive fronts which is only possible in case of finite initial conditions. In this paper, the considerations are restricted to the case when at the beginning of the process the prey is spatially uniformly distributed at the level of the carrying capacity, $u(\mathbf{r}, t) \equiv 1$, whereas the predator is inhabiting a finite region. These initial conditions correspond to the problem of biological invasions [4].

Thus, we begin with 1-D problem described by the following equations:

$$u_t = u_{xx} + u(1 - u) - \frac{u}{u + h}v, \quad (6)$$

$$v_t = v_{xx} + k\frac{u}{u + h}v - mv, \quad (7)$$

and

$$u(x, 0) = 1, \quad \forall x; \quad v(x, 0) = V_0, \quad \text{if } |x| \leq \frac{\Delta}{2}; \quad v(x, 0) = 0, \quad \text{if } |x| > \frac{\Delta}{2}. \quad (8)$$

Since the dynamics of the system does not show any significant dependence on the form of the finite initial distribution of the predator, the simplest form of $v(x, 0)$ is chosen here.

Another important point is the number and the type of stationary states of the reduced system. It is readily seen that the phase plane (u, v) of equations (4),(5) without diffusion terms has the following structure: under the condition $h < (1 - p)/p$, $p = m/k$, there are three stationary points in the physically meaningful region $u \geq 0$, $v \geq 0$, namely, $(0, 0)$ (trivial extinction), $(1, 0)$ (predator extinction), and (u_*, v_*) (coexistence), where

$$u_* = \frac{ph}{1 - p}, \quad v_* = (1 - u_*)(h + u_*). \quad (9)$$

The trivial solutions $(0, 0)$ and $(1, 0)$ are always saddle-points, whereas the nontrivial point (u_*, v_*) can be either focus or node, stable, or unstable, depending on the problem parameters; cf. Figure 1 and see [16,17] for more details.

The distinctions in the dynamics of system (6),(7) for different values of k , p , h can be associated with the change of the stability of the nontrivial stationary state (u_*, v_*) which takes place when

$$h = h_{cr}(p) = \frac{1 - p}{1 + p}, \quad (10)$$

which corresponds to Curve 2 in Figure 1. Note that the position of Curves 1 and 2 does not depend on k . For all parameter values when the state (u_*, v_*) is unstable, i.e., below Curve 2, it is surrounded by a stable limit cycle.

The full account of possible dynamical regimes of system (6),(7) with finite initial conditions can be found in [17] or in [18] for the case of a somewhat different parameterization of the biological processes. Typically, the solution of problem (6)–(8) evolves, after a certain transition time, to the propagation of stationary diffusive fronts. In case the stationary state (u_*, v_*) is locally stable, the dynamics of these fronts is something one can expect: they “switch” the system from the homogeneous stationary unstable state $u \equiv 1$, $v \equiv 0$ (after some damped oscillations if (u_*, v_*) is a stable focus [2,19]) to the homogeneous stable state $u \equiv u_*$, $v \equiv v_*$.

However, the situation becomes less expected when the point (u_*, v_*) becomes unstable, i.e., when the parameters cross the Hopf bifurcation curve $h = h_{cr}(p)$ in the plane (p, h) . The results

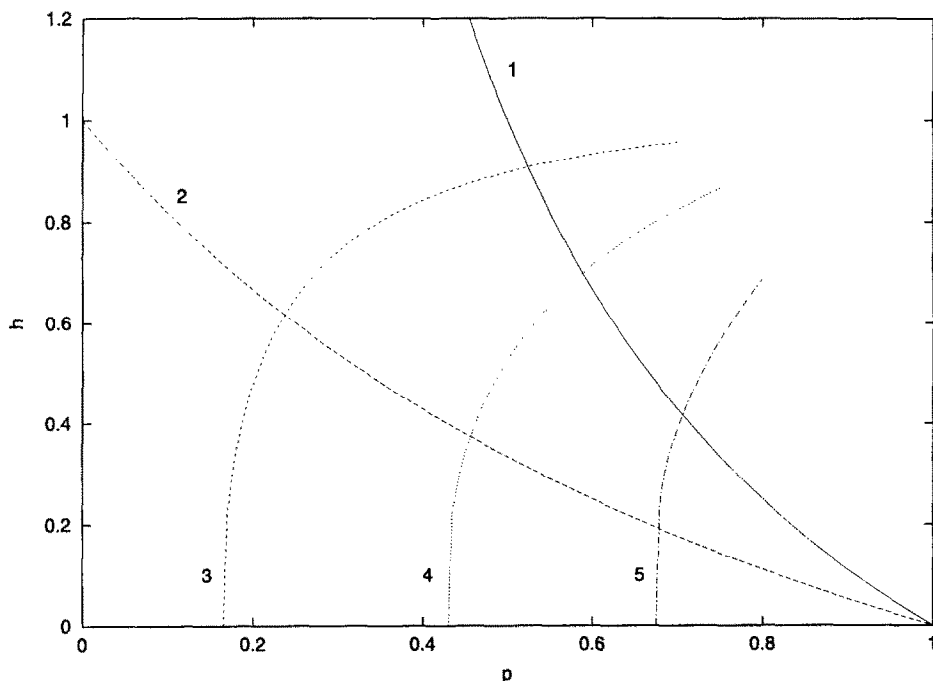


Figure 1. A map in the (p, h) parametric plane of the prey-predator system (4),(5): the nontrivial "coexistence" state exists for the parameters below Curve 1 and loses its stability when crossing Curve 2; the critical relation (17) for the dynamical stabilization is shown by Curve 3 for $k = 0.1$, Curve 4 for $k = 0.4$, and Curve 5 for $k = 1.2$.

of numerical experiments show that in this case, for parameter values "not very far" from the critical relation (10), the diffusive fronts still work as switching waves but switching the system to the state $u \equiv u_*$, $v \equiv v_*$ which is now locally unstable. Typical wave profiles are presented in Figure 2 (since the problem (6)–(8) is symmetrical with respect to $x = 0$, only half of the numerical domain is shown).

One can see that, after rather strong oscillations at the front of the wave, there comes the region (from approximately $x = 200$ to $x = 420$ in Figure 2, top, and from $x = 400$ to $x = 900$ in Figure 2, bottom) where the concentrations $u(x, t)$ and $v(x, t)$ nearly reach their stationary (but unstable!) values u_* , v_* —the dynamical stabilization takes place. We want to stress, based on our numerical results, that this unstable "plateau" exists during a remarkably long time before it is finally displaced by the irregular spatiotemporal oscillations [16]. Moreover, the length of the plateau grows with time; cf. top and bottom of Figure 2.

In order to better understand this phenomenon, the following points must be addressed.

- (i) For which restrictions on the parameter values can the dynamical stabilization of the unstable equilibrium occur? There must be some restrictions because the stabilization does not appear for arbitrary sets of parameters.
- (ii) How does the length of the plateau change (increase) with time?

The first of these problems has been considered in [17], where the conditions for dynamical stabilization were related to the change of the type of the nontrivial stationary state in four-dimensional phase space generated by system (6),(7) in case of stationary wave propagation. However, the results obtained in this way do not allow us to make any estimates concerning the length of the plateau. Besides, the approach developed in [17] somewhat lacks physical lucidity, and that might make the interpretation of the results difficult. In this paper, another physically clear approach to deal with problems (i),(ii) is proposed. Both the restrictions on the problem parameters and the equation describing the growth of the plateau length with time will appear quite naturally as a result of the comparison between the speed of different diffusive fronts.

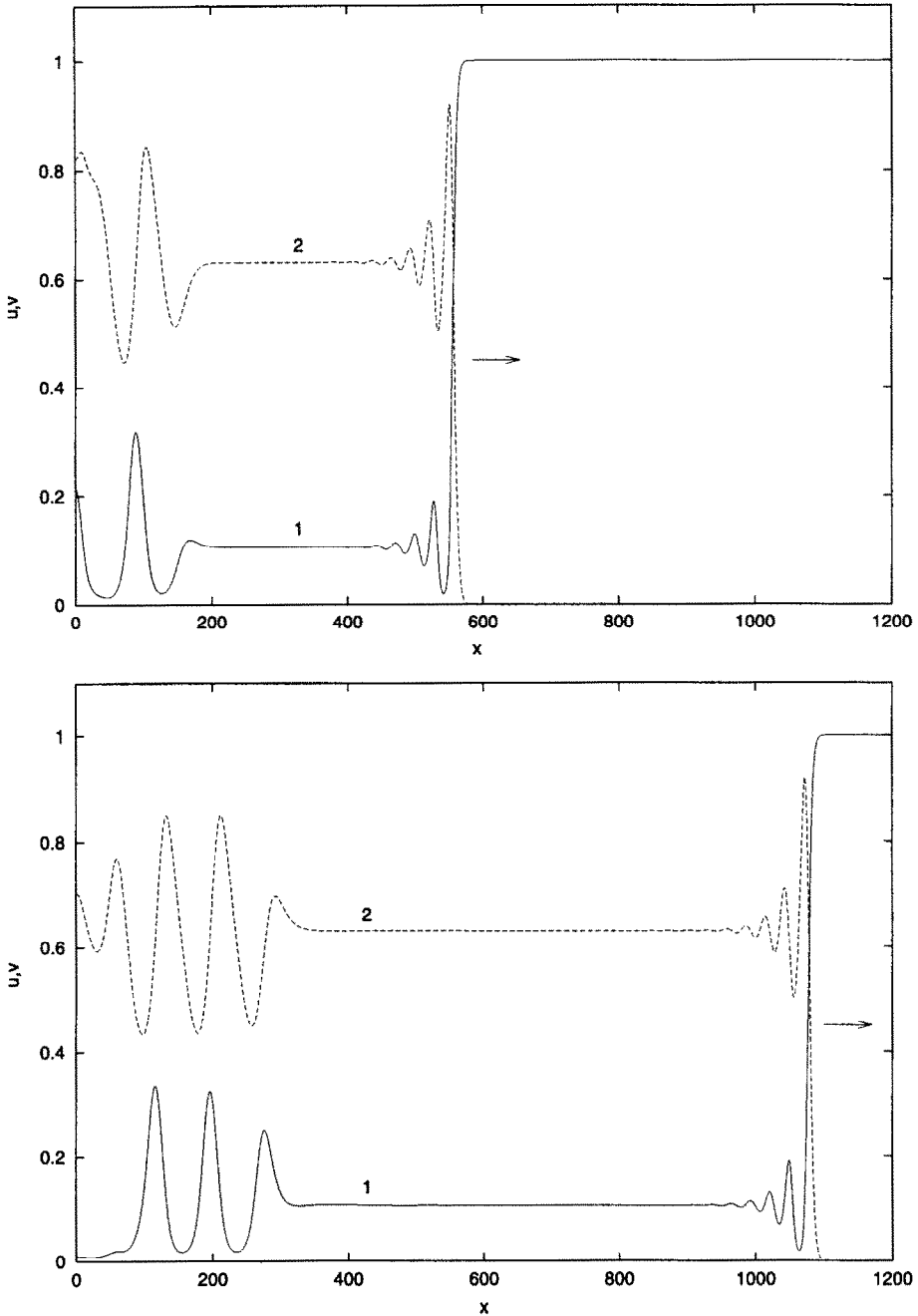


Figure 2. The profiles of the concentration of prey u (Curve 1) and predator v (Curve 2) calculated at $t = 600$ (top) and $t = 1200$ (bottom) for parameters $k = 0.4$, $p = 0.15$, and $h = 0.6$ and the initial conditions (8) with $V_0 = 1.2$ and $\Delta = 100$. The arrows show the direction of the diffusive front propagation. The region of homogeneity in the middle corresponds to the locally unstable “coexistence” state.

The idea of the method is as follows. Results of computer simulations [17,18,20] show that, generally, the stabilization of a locally unstable state occurs behind the stationary diffusive fronts (which may have an oscillating “structure”, cf. Figure 2), travelling with a certain constant speed c . The minimal possible value c_{\min} can be obtained considering the properties of the problem solution in the vicinity of the steady $(1, 0)$ state in R^4 phase space [2,17],

$$c_{\min} = \left[2 \left(T_0 + \sqrt{\Delta_0} \right) \right]^{1/2}, \quad (11)$$

with $\Delta_0 = T_0^2 - 4\delta_0$. T_0 and δ_0 are trace and determinant of the Jacobi matrix, respectively. We want to mention that, although c_{\min} is only the exact lower bound of the possible values of the speed of the front and not its actual value, it provides a good estimate for the actual value c . Moreover, it holds quite often $c = c_{\min}$ [2].

Furthermore, if (u_*, v_*) is unstable, the propagation of the diffusive fronts, no matter if with or without the dynamical stabilization in the wake, is followed by a region occupied by spatiotemporal oscillations [16,17,20–22], typically irregular. The remarkable thing is that in case of the formation of the unstable plateau, there exists a distinct boundary or interface at any time, separating the plateau from the region with irregular oscillations [16,22]. Our numerical results show that the size of the region always grows with time, and the interface propagates with a constant speed. Considering the travelling wave solutions of equations (6),(7) far ahead of the interface, where they can be regarded as small perturbations of the stationary state $u \equiv u_*$, $v \equiv v_*$, we arrive at the following estimate for the speed w of the interface [22]:

$$w_{\min} = \sqrt{2T_1} \quad (12)$$

in case of an unstable focus, and

$$w_{\min} = \left[2 \left(T_1 + \sqrt{\Delta_1} \right) \right]^{1/2} \quad (13)$$

in case of an unstable node. Here $\Delta_1 = T_1^2 - 4\delta_1$ where T_1 and δ_1 are the trace and the determinant of the matrix of the system in the vicinity of the coexistence state (u_*, v_*) . Again, although equations (12),(13) give only the minimal possible value of the speed, it is in an excellent agreement with numerical results [22].

Numerical results indicate that the dynamical stabilization is unlikely to be observed if (u_*, v_*) is an unstable node; this is also in agreement with the results of the bifurcation analysis [17]. Hence, the further considerations treat the case (u_*, v_*) as an unstable focus.

Now, the domain where the dynamical stabilization may take place is bounded by two moving boundaries: the leading edge propagating with a constant speed c and the interface between the plateau and the region of irregular oscillations propagating with a speed w . The development of the plateau is thus controlled by the relation between the values of c and w . Obviously, in case $w < c$, the length of the domain grows with time as $(c - w)t$. Let us note, however, that since the leading front behaves as a stationary travelling wave, its form and “width” (i.e., the size of the region occupied by the regular damping oscillations, cf. Figure 2 between $x = 700$ and $x = 850$) do not change with time. Then, the increase of the length of the domain locked between the two moving fronts can only mean the increase of the length of the plateau. Thus, we obtain

$$L_{\text{plateau}} = (c - w)t + L_0, \quad (14)$$

where L_0 is a constant.

On the other hand, in case $w > c$ the dynamical stabilization can hardly be observed. The length of the plateau, if it happens to appear as a result of certain specific initial conditions, would decrease with time until, finally, the region of spatiotemporal oscillations would start immediately after the stationary travelling front; cf. Figure 3. Unlike the case shown in Figure 2, the “nucleus” of the unstable plateau which can be seen just behind the damping oscillations at the front does not grow with time.

Thus, we arrive at a simple necessary condition for the dynamical stabilization

$$w < c. \quad (15)$$

However, relations (14) and (15) are still not very useful because the actual values of speed are not known. Assuming that the fronts propagate with the minimal possible speed, as it usually

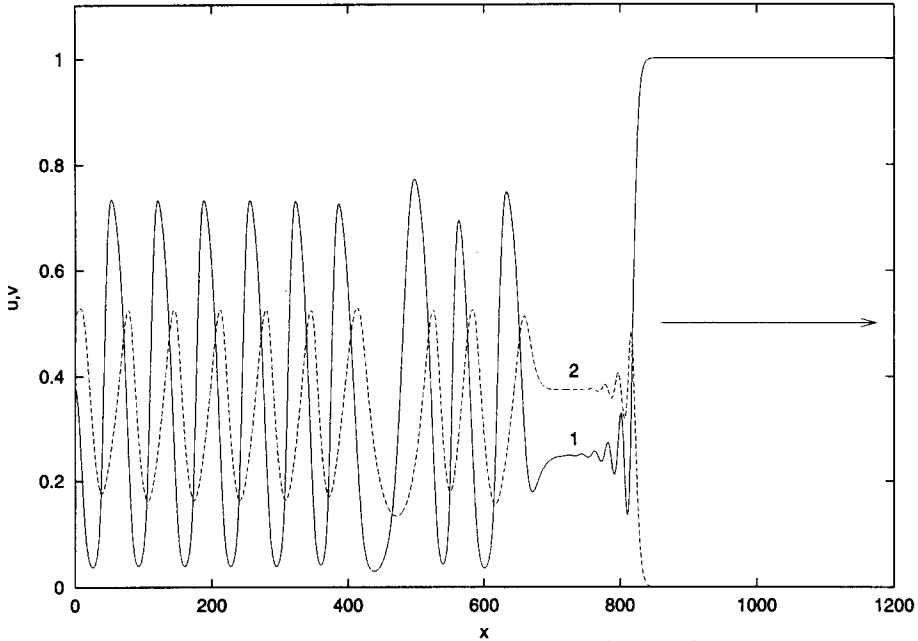


Figure 3. The profiles of the concentration of prey u (Curve 1) and predator v (Curve 2) calculated at $t = 1000$ for parameters $k = 0.4$, $p = 0.5$, and $h = 0.25$. No plateau is formed.

takes place, and taking into account equations (11) and (12), from (15) we obtain the following critical relation between the problem parameters:

$$T_1 = T_0 + \sqrt{\Delta_0}. \quad (16)$$

If the nonlinearities in the equations are chosen as in equations (6),(7), one can easily obtain that $T_1 = -2\delta_0$ (note that $\delta_0 < 0$ because the state “prey only” is a saddle-point) and, finally,

$$\frac{p}{1-p} [(1-h) - p(1+h)] = -2k \left(p - \frac{1}{1+h} \right). \quad (17)$$

The critical relation (17) is shown in Figure 1 by the dashed line for the values of parameter $k = 0.1$ (Curve 3), $k = 0.4$ (Curve 4), and $k = 1.2$ (Curve 5). The domain in the (p, h) parameter plane where one can expect the dynamical stabilization of an unstable equilibrium in the wake of the travelling diffusive front is on the left-hand side of the dashed line and below the Hopf bifurcation Curve 2.

3. DYNAMICAL STABILIZATION IN TWO SPATIAL DIMENSIONS

The results of the previous section were obtained for a spatially one-dimensional diffusion-reaction system. However, the dynamics of natural systems is usually higher dimensional. In this connection, and also in order to show that the dynamical stabilization is not an exotic but a rather typical phenomenon in diffusion-reaction systems, it seems important to know whether the previous results can be extended to a more-dimensional case.

Note that, strictly speaking, this extension is not a formal routine and the results can hardly be foreseen. The matter is that the increase of the number of the spatial dimensions not only makes the dynamics of the system more complex, but may lead to suppression of the regimes which would be dominant in the system with fewer dimensions. This is just the situation described above: the increase of the number of spatial dimensions from 0 (cf. equations (3)) to 1 (equations (6),(7)) makes the unstable equilibrium dynamically stable.

Let us mention that, in some cases, the dimensionality of the system dynamics depends on the scale of the processes under consideration. For instance, the spatiotemporal functioning of a plankton community is three dimensional if considered on scale $L \leq L_0$ where L_0 is the thickness of the upper productive ocean layer, but becomes effectively two dimensional on scale $L \gg L_0$.

Taking into account that the 3-D case is much more complicated for computer simulations and for the visualization of the results, we restrict ourselves to the 2-D case here. Figure 4 shows the snapshots of the prey spatial distribution (the distribution of predator is qualitatively similar) obtained numerically for a prey-predator community described by equations (4),(5) where now $u = u(x, y, t)$, $v = v(x, y, t)$, and $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. The growing inner grey ring is easily recognized as the dynamically stabilized unstable coexistence region.

Thus, the phenomenon of the dynamical stabilization of an unstable equilibrium exists also in a 2-D diffusion-reaction system. As it was in 1-D case, the size of the unstable plateau grows with time. However, to the case of cylindrical diffusive fronts, neither condition (14) nor (15) is not directly applicable because they were obtained for the plane waves. Particularly, condition (14) now gives only rough estimates of the parameter values where the dynamical stabilization may be observed.

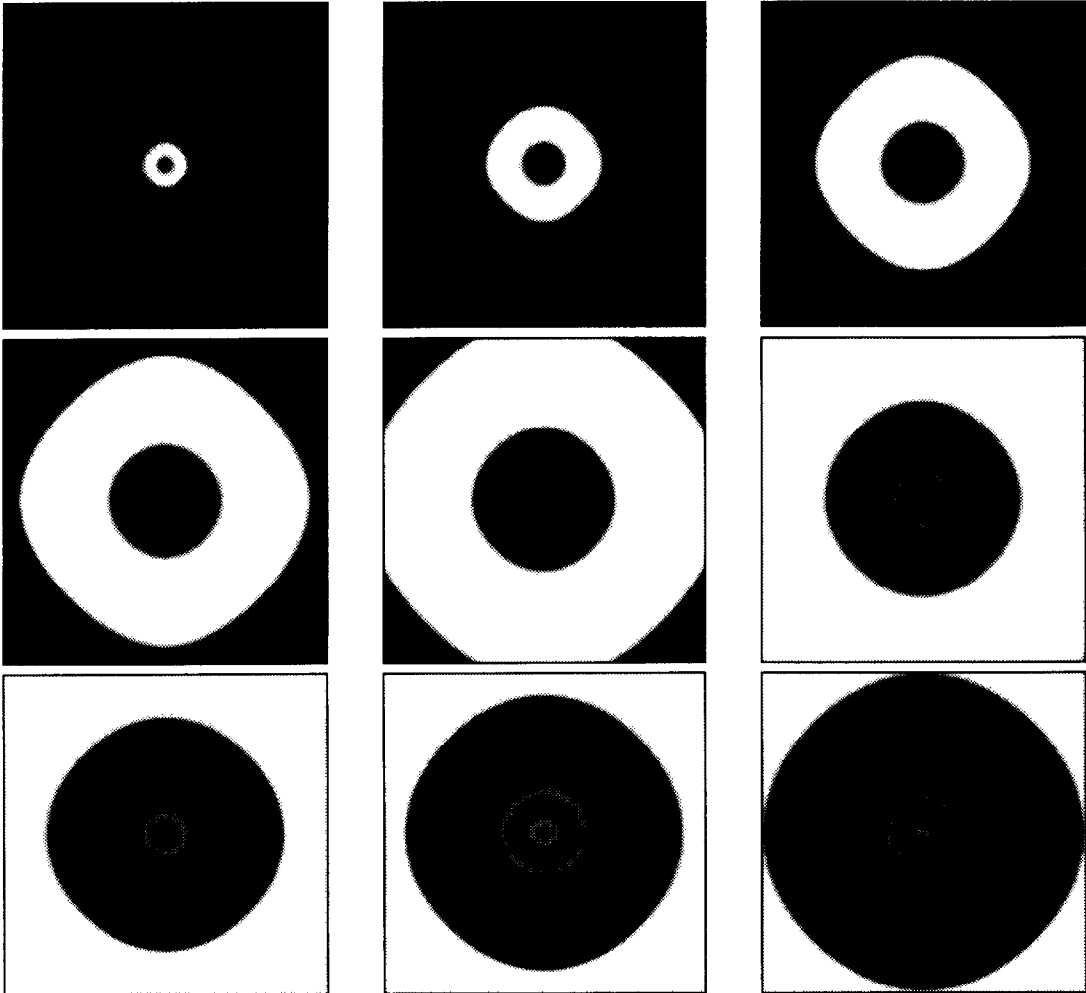


Figure 4. The 2-D spatial distribution of prey calculated at equidistant moments of time, parameter values are the same as in Figure 2. Black color corresponds to the no-species state, while in white areas the prey is at its carrying capacity. The grey color corresponds to the unstable coexistence state.

4. A CHEMICAL SYSTEM: THE GRAY-SCOTT MODEL

In the previous sections, the possibility of the dynamical stabilization of an unstable equilibrium has been shown by computer experiments. Furthermore, the restrictions on the parameter values necessary for the stabilization have been obtained analytically. It has been demonstrated that the phenomenon is not sensitive to the dimensionality of the system and can be observed in 1-D and 2-D cases.

A prey-predator system has been chosen as local kinetics to demonstrate the stabilization by computer experiments. Now, another question naturally arises: how strongly does the existence of this phenomenon depend on the type of the local kinetics? Although it is shown in [17,18] that the dynamical stabilization is robust with respect to some variations in the form of the nonlinearities in equations (1),(2) (cf. also [20]), a certain doubt may still exist. The matter is that any realistic parameterization of the prey-predator interactions, in spite of the details, should allow for, at least, two principal features:

- (i) $f(u = 0, v) = 0$, $g(u, v = 0) = 0$, and
- (ii) for large values of u the predation must show a tendency to saturation; cf. equations (4),(5).

These features impose certain constraints both on the structure of the phase plane of the reduced system (3), and on the spatiotemporal dynamics of the full system (1),(2).

To check the robustness of the results with respect to the type of local interactions, another field of application of equations (1),(2) is considered. A system of two chemical reactants is free from the above limitations and can possess quite different local kinetics. As a particular example, the well-known Gray-Scott model [1] is chosen, describing an autocatalytic reaction in an open 1-D flow reactor,

$$u_t = u_{xx} + F(1 - u) - uv^2, \quad (18)$$

$$v_t = v_{xx} + uv^2 - (F + k)v, \quad (19)$$

with accordingly chosen dimensionless variables; see [23] for details. Here $u(x, t)$ and $v(x, t)$ are the concentration of the substrate and the autocatalyst, respectively, F is the flow rate, and k is the effective rate constant for the decay of the autocatalyst. As in previous cases, we assume $D_u = D_v$ for simplicity.

Equations (18),(19) have been investigated in many papers; e.g., see [24,25]. A very brief summary of the results is only given here as far as they will be needed for the further considerations. One can easily see that under the limitation $d = 1 - 4(F + k)^2/F > 0$, there are three stationary points: “substrate only” $(1, 0)$ and two nontrivial coexistence states (u_s, v_s) (“substrate dominated”) and (u_a, v_a) (“autocatalyst dominated”) where

$$u_s = \frac{1 + \sqrt{d}}{2}, \quad v_s = \left(\frac{F}{F + k} \right) \frac{1 - \sqrt{d}}{2}, \quad (20)$$

$$u_a = \frac{1 - \sqrt{d}}{2}, \quad v_a = \left(\frac{F}{F + k} \right) \frac{1 + \sqrt{d}}{2}. \quad (21)$$

When crossing the critical curve $d = 0$ in the (k, F) -plane (Curve 1 in Figure 5) towards smaller values of k (i.e., from right to left), the two nontrivial states appear through a saddle-node bifurcation, the “autocatalyst dominated” state being an unstable node.

The “substrate only” state is always stable and the “substrate dominated” state is always unstable. A change in the local dynamics can be associated with the change of the type of the “autocatalyst dominated” state, first of all, with the change of its stability which takes place when

$$\frac{k - F}{k + F} = \sqrt{d}, \quad (22)$$

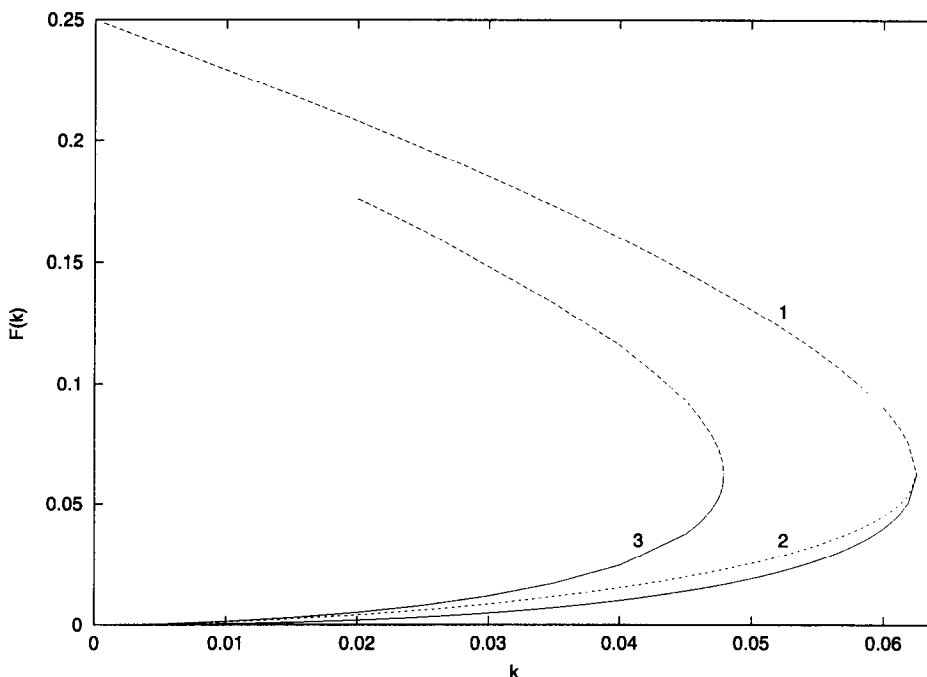


Figure 5. A map in the parametric (k, F) plane. Curve 1 bounds the domain where the two nontrivial states exist, Curve 2 is the Hopf bifurcation line, and Curve 3 corresponds to the critical relation for the dynamical stabilization.

i.e., when crossing Curve 2 in Figure 5. The Hopf bifurcation, which takes place when crossing the Curve (22), is supercritical for $k < k_{cr}$ (where k_{cr} is estimated as about 0.035, cf. [24]) and only in this case a stable limit cycle appears. Otherwise, i.e., for $k_{cr} < k < 0.625$, no limit cycle arises and any trajectory starting in the vicinity of the “autocatalyst dominated” state after a number of expanding convolutions is finally attracted to the “substrate only” state.

This brief consideration of the local dynamics of the Gray-Scott model provides “input” information for the investigation of the spatiotemporal dynamics of the distributed system (18),(19). Namely, accounting for the results of the previous sections, it is now obvious that, as far as one is concerned with the possibility to observe the dynamical stabilization, the values of F and k in equations (18),(19) should be chosen from the domain between the Hopf bifurcation curve and the saddle-node bifurcation curve where the “autocatalyst dominated” state is unstable, i.e., from the narrow strip between Curves 1 and 2 in Figure 5.

Thus, the structure of the local phase plane of the Gray-Scott model is essentially different from the one of the prey-predator system (4),(5). Quite naturally, it results in a significantly different behaviour of the diffusive fronts; cf. [17,18,23–25]. And this difference makes it probably even more remarkable that the phenomenon of dynamical stabilization also occurs in the system described by equations (18),(19). Figure 6 shows the profiles of the concentrations u and v at $t = 3200$ calculated for the initial conditions (8) with $\Delta = 400$ and $V_0 = 0.1$ for parameter values $F = 0.015$ and $k = 0.04$. Only half of the domain is shown.

Again, after promptly damping oscillations behind the leading edge, there comes a region where substrate and autocatalyst are distributed homogeneously at the level corresponding to the locally unstable state $u_a = 0.28$, $v_a = 0.196$. To stress the distinctions from the cases considered in Sections 2 and 3, it should be noted that the unstable focus (u_a, v_a) is not surrounded by a stable limit cycle now. Moreover, since the “substrate only” state is stable for all parameter values, the propagation of the diffusive front followed by the dynamical stabilization, with the length of the unstable plateau growing with time, means “switching” the system from the stable state to an unstable one, a situation that seems quite exotic if considered in a general physical context.

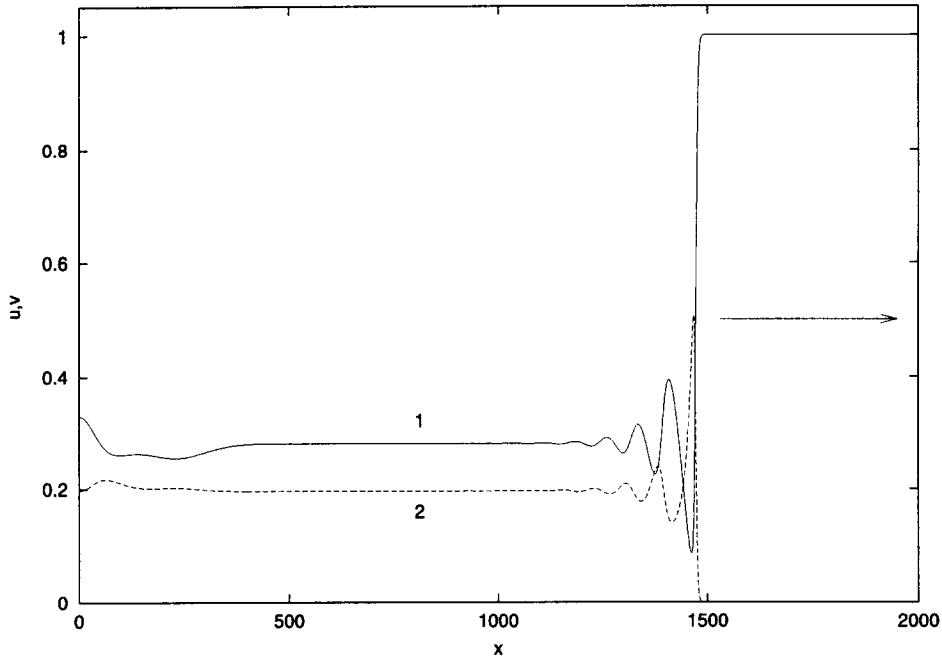


Figure 6. The profiles of the concentration of the substrate u (Curve 1) and the autocatalyst v (Curve 2) calculated at $t = 3200$ for parameters $F = 0.015$ and $k = 0.04$. The plateau behind the oscillating front appears as a result of the dynamical stabilization of the unstable “autocatalyst dominated” state.

The next step to be done is to examine the restrictions on the problem parameters F and k resulting from the conditions for dynamical stabilization; cf. equations (15) and (16). It is readily seen that for the Gray-Scott model (18),(19) $T_0 + \sqrt{\Delta_0} = -2F$, whereas $T_1 = (F + k) - Fu_a^{-1}$. Accounting for (21), equation (16) takes the following form:

$$\frac{F + k}{3F + k} = \sqrt{d}. \quad (23)$$

The critical relation (23) is shown in Figure 5 by Curve 3. Note that Curve 3 is situated in the domain where the “autocatalyst dominated” state is stable. Thus, in this case the necessary condition (15) of the dynamical stabilization is fulfilled for all parameter values from the domain where this state is unstable (cf. the strip between Curves 1 and 2) and does not bring any additional restraints on the values of the problem parameters.

5. DISCUSSION AND CONCLUSIONS

In this paper, a new facet of the dynamics of a two-component diffusion-reaction system has been demonstrated: the dynamical stabilization of an unstable equilibrium resulting in the formation of a homogeneous spatial distribution of the interacting components, a “plateau” at the level corresponding to a locally unstable steady state. It has been shown that this is not an exotic but a rather typical phenomenon occurring both in one and two spatial dimensions and in systems with essentially different local kinetics, i.e., with different structure of the local phase plane. The size of the dynamically stabilized homogeneous region is growing with time according to equation (14).

A simple necessary condition for the dynamical stabilization has been suggested, cf. equation (15), which imposes certain constraints on the parameter values. It must be mentioned, however, that equation (15) gives only a necessary condition. Particularly, concerning the Gray-Scott model, the absence of additional restrictions does not mean that the dynamical stabilization can be observed for all the parameter values where the “autocatalyst dominated” state is unstable. To obtain more detailed information about possible constraints, the semiempirical “physical”

approach considered in this paper should be complemented by the results of a strict bifurcation analysis; cf. [17]. Ideally, a rigorous mathematical investigation of the problem should also include an explicit analytical proof of the existence of a separatrix in the R^4 phase space of the system corresponding to the stationary travelling front. For a somewhat different problem, such consideration is done in [19,26], whereas it is still an open problem for equations (4),(5) or (18),(19).

Concerning a probable application of the results to natural systems, another open problem is how the phenomenon can be modified in the presence of noise. Our tentative numerical results show that for a periodically applied perturbation, the plateau can survive if the period is sufficiently large and the amplitude of the perturbation is sufficiently small. Under the influence of a perturbation, the length of the plateau may decrease, and in some cases it may be broken into a few parts separated by regions with irregular spatiotemporal oscillations. However, this problem needs more careful consideration and will be subject of a separate paper.

In conclusion, it should be noted that the existence of this phenomenon may shed a new light on some ecological problems. Particularly, it is shown in [27] that the temporary behaviour of the concentrations of key species in a biological community sometimes exhibits “intermittence”: an oscillatory behaviour gives way to a quasi-stationary state of the community which is followed again by the oscillations with a period of alternation much less than one year so that it can unlikely be related to the seasonal changes. Now, accounting for the results of this paper, one can consider the situation when an “observer” is taking measurements of the species concentrations in a fixed point in front of the population wave; cf. Figure 2. Then his account of the temporal dynamics of the community in a given point would be very similar to the one reported in [27]. The intermittent temporal behaviour of the community would arise as a result of a biological invasion combined with the dynamical stabilization behind the front. An indirect proof for this explanation can be also found in [28] where it is shown how a spatial structure of a community may result in complex temporal dynamics.

REFERENCES

1. P. Gray and S.K. Scott, *Chemical Oscillations and Instabilities*, Oxford University Press, Oxford, (1990).
2. J.D. Murray, *Mathematical Biology*, Springer-Verlag, Berlin, (1989).
3. A. Okubo, *Diffusion and Ecological Problems: Mathematical Models*, Springer-Verlag, Berlin, (1980).
4. N. Shigesada and K. Kawasaki, *Biological Invasions: Theory and Practice*, Oxford University Press, Oxford, (1997).
5. H. Haken, *Advanced Synergetics, Springer Series in Synergetics, Volume 20*, Springer-Verlag, Berlin, (1983).
6. G. Nicolis and I. Prigogine, *Self-Organization in Non-Equilibrium Systems*, Wiley, New York, (1977).
7. A.M. Turing, On the chemical basis of morphogenesis, *Phil. Trans. R. Soc. Lond. B* **237**, 37–72, (1952).
8. L.A. Segel and J.L. Jackson, Dissipative structure: An explanation and an ecological example, *J. Theor. Biol.* **37**, 545–559, (1972).
9. K. Kaneko, Editor, *Theory and Applications of Coupled Map Lattices*, Wiley & Sons, Chichester, (1993).
10. L. Schimansky-Geyer, M. Mieth, H. Rosé and H. Malchow, Structure formation by active Brownian particles, *Phys. Lett. A* **207**, 140–146, (1995).
11. F. Schweitzer, W. Ebeling and B. Tilch, Complex motion of Brownian particles with energy deposit, *Phys. Rev. Lett.* **80**, 5044–5047, (1998).
12. J.G. Skellam, Random dispersal in theoretical populations, *Biometrika* **38**, 196–218, (1951).
13. D. Dubois, A model of patchiness for prey-predator plankton populations, *Ecological Modelling* **1**, 67–80, (1975).
14. J.S. Wroblewski and J.J. O'Brien, A spatial model of plankton patchiness, *Marine Biology* **35**, 161–176, (1976).
15. H. Malchow, Spatio-temporal pattern formation in nonlinear non-equilibrium plankton dynamics, *Proc. R. Soc. Lond. B* **251**, 103–109, (1993).
16. S.V. Petrovskii and H. Malchow, A minimal model of pattern formation in a prey-predator system, *Mathl. Comput. Modelling* **29** (8), 49–63, (1999).
17. S.V. Petrovskii and H. Malchow, Critical phenomena in plankton communities: KISS model revisited, *Non-linear Analysis: Real World Applications* **1**, 37–51, (2000).
18. S.V. Petrovskii, M.E. Vinogradov and A.Yu. Morozov, Spatial-temporal dynamics of a localized populational “burst” in a distributed prey-predator system, *Oceanology* **38**, 881–890, (1998).
19. S.R. Dunbar, Travelling wave solutions of diffusive Lotka-Volterra equations, *J. Math. Biol.* **17**, 11–32, (1983).

20. J.A. Sherratt, Invadive wave fronts and their oscillatory wakes are linked by a modulated travelling phase resetting wave, *Physica D* **117**, 145–166, (1998).
21. J.A. Sherratt, M.A. Lewis and A.C. Fowler, Ecological chaos in the wake of invasion, *Proc. Natl. Acad. Sci. USA* **92**, 2524–2528, (1995).
22. S.V. Petrovskii and H. Malchow, Wave of chaos: A new mechanism of pattern formation in a prey-predator system, *Theoretical Population Biology* **59**, 157–174, (2001).
23. J.E. Pearson, Complex patterns in simple systems, *Science* **261**, 189–192, (1993).
24. K.E. Rasmussen, W. Mazin, E. Mosekilde, G. Dewel and P. Borckmans, Wave-splitting in the bistable Gray-Scott model, *Int. J. of Bifurcations and Chaos* **6**, 1077–1092, (1996).
25. F. Davidson, Chaotic wakes and other wave-induced behavior in a system of reaction-diffusion equations, *Int. J. of Bifurcation and Chaos* **8**, 1303–1313, (1998).
26. S.R. Dunbar, Travelling waves in diffusive predator-prey equations: Periodic orbits and point-to-periodic heteroclinic orbits, *SIAM J. Appl. Math.* **46**, 1057–1078, (1986).
27. W.W. Murdoch and E. McCauley, Three distinct types of dynamic behaviour shown by a single planktonic system, *Nature* **316**, 628–630, (1985).
28. E. Ranta, V. Kaitala and P. Lundberg, The spatial dimension in population fluctuations, *Science* **278**, 1621–1623, (1997).

80. Reinitz J, Mjolsness E, Sharp DH (1995) Model for cooperative control of positional information in *Drosophila* by bicoid and maternal hunchback. *J Exp Zool* 271:47–56
81. Rutherford SL, Lindquist S (1998) Hsp90 as a capacitor for morphological evolution. *Nature* 396:336–42
82. Sakuma R, Ohnishi Yi Y, Meno C et al (2002) Inhibition of Nodal signalling by Lefty mediated through interaction with common receptors and efficient diffusion. *Genes Cells* 7:401–12
83. Salazar-Ciudad I, Garcia-Fernandez J, Sole RV (2000) Gene networks capable of pattern formation: from induction to reaction-diffusion. *J Theor Biol* 205:587–603
84. Salazar-Ciudad I, Newman SA, Solé R (2001) Phenotypic and dynamical transitions in model genetic networks. I. Emergence of patterns and genotype-phenotype relationships. *Evol Dev* 3:84–94
85. Salazar-Ciudad I, Solé R, Newman SA (2001) Phenotypic and dynamical transitions in model genetic networks. II. Application to the evolution of segmentation mechanisms. *Evol Dev* 3:95–103
86. Schmalhausen II (1949) Factors of evolution. Blakiston, Philadelphia
87. Schulte-Merker S, Smith JC (1995) Mesoderm formation in response to Brachyury requires FGF signalling. *Curr Biol* 5:62–7
88. Small S, Blair A, Levine M (1992) Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J* 11:4047–4057
89. Small S, Kraut R, Hoey T et al (1991) Transcriptional regulation of a pair-rule stripe in *Drosophila*. *Genes Dev* 5:827–39
90. Solnica-Krezel L (2003) Vertebrate development: taming the nodal waves. *Curr Biol* 13:R7–9
91. Spemann H, Mangold H (1924) Über Induktion von Embryonalanlagen durch Implantation artfremder Organisatoren. *Wilhelm Roux' Arch Entw Mech Org* 100:599–638
92. St Johnston D, Nusslein-Volhard C (1992) The origin of pattern and polarity in the *Drosophila* embryo. *Cell* 68:201–19
93. Steinberg MS (1963) Reconstruction of tissues by dissociated cells. Some morphogenetic tissue movements and the sorting out of embryonic cells may have a common explanation. *Science* 141:401–8
94. Stern CD, Bellairs R (1984) Mitotic activity during somite segmentation in the early chick embryo. *Anat Embryol (Berl)* 169:97–102
95. Stollewerk A, Schoppmeier M, Damen WG (2003) Involvement of Notch and Delta genes in spider segmentation. *Nature* 423:863–5
96. Strogatz SH (1994) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. Perseus Pub, Cambridge
97. Sun B, Bush S, Collins-Racie L et al (1999) *derriere*: a TGF-beta family member required for posterior development in *Xenopus*. *Development* 126:1467–1482
98. Tsarfaty I, Resau JH, Rulong S, Keydar I, Faletto DL, Vande Woude GF (1992) The met proto-oncogene receptor and lumen formation. *Science* 257:1258–61
99. Turing AM (1952) The chemical basis of morphogenesis. *Phil Trans Royal Soc Lond B* 237:37–72
100. Van Obberghen-Schilling E, Roche NS, Flanders KC et al (1988) Transforming growth factor beta-1 positively regulates its own expression in normal and transformed cells. *J Biol Chem* 263:7741–7746
101. Waddington CH (1957) The Strategy of the Genes. Allen and Unwin, London
102. Winfree AT (1980) The geometry of biological time. Springer, New York
103. Wilkins AS (1997) Canalization: a molecular genetic perspective. *BioEssays* 19:257–262
104. Wolpert L (2002) Principles of development. Oxford University Press, Oxford New York

Books and Reviews

- Meinhardt H (1982) Models of biological pattern formation. Academic, New York
- Müller GB, Newman SA (2003) Origination of organismal form: beyond the gene in developmental and evolutionary biology. MIT Press, Cambridge, pp 221–239
- Newman SA, Comper WD (1990) 'Generic' physical mechanisms of morphogenesis and pattern formation. *Development* 110: 1–18

Biological Fluid Dynamics, Non-linear Partial Differential Equations

ANTONIO DESIMONE¹, FRANÇOIS ALOUGES²,
ALINE LEFEBVRE²

¹ SISSA-International School for Advanced Studies,
Trieste, Italy

² Laboratoire de Mathématiques, Université Paris-Sud,
Orsay cedex, France

Article Outline

Glossary
Definition of the Subject
Introduction
The Mathematics of Swimming
The Scallop Theorem Proved
Optimal Swimming
The Three-Sphere Swimmer
Future Directions
Bibliography

Glossary

Swimming The ability to advance in a fluid in the absence of external propulsive forces by performing cyclic shape changes.

Navier–Stokes equations A system of partial differential equations describing the motion of a simple viscous incompressible fluid (a Newtonian fluid)

$$\rho \left(\frac{\partial v}{\partial t} + (v \cdot \nabla) v \right) = -\nabla p + \eta \Delta v$$

$$\operatorname{div} v = 0$$

where v and p are the velocity and the pressure in the fluid, ρ is the fluid density, and η its viscosity. For simplicity external forces, such as gravity, have been dropped from the right hand side of the first equation, which expresses the balance between forces and rate of change of linear momentum. The second equation constrains the flow to be volume preserving, in view of incompressibility.

Reynolds number A dimensionless number arising naturally when writing Navier–Stokes equations in non-dimensional form. This is done by rescaling position and velocity with $x^* = x/L$ and $v^* = v/V$, where L and V are characteristic length scale and velocity associated with the flow. Reynolds number (Re) is defined by

$$\text{Re} = \frac{VL\rho}{\eta} = \frac{VL}{\nu}$$

where $\nu = \eta/\rho$ is the kinematic viscosity of the fluid, and it quantifies the relative importance of inertial versus viscous effects in the flow.

Steady Stokes equations A system of partial differential equations arising as a formal limit of Navier–Stokes equations when $\text{Re} \rightarrow 0$ and the rate of change of the data driving the flow (in the case of interest here, the velocity of the points on the outer surface of a swimmer) is slow

$$\begin{aligned} -\eta\Delta v + \nabla p &= 0 \\ \text{div } v &= 0. \end{aligned}$$

Flows governed by Stokes equations are also called creeping flows.

Microscopic swimmers Swimmers of size $L = 1\mu\text{m}$ moving in water ($\nu \sim 1\text{mm}^2/\text{s}$ at room temperature) at one body length per second give rise to $\text{Re} \sim 10^{-6}$. By contrast, a 1 m swimmer moving in water at $V = 1\text{m/s}$ gives rise to a Re of the order 10^6 .

Biological swimmers Bacteria or unicellular organisms are microscopic swimmers; hence their swimming strategies cannot rely on inertia. The devices used for swimming include rotating helical flagella, flexible tails traversed by flexural waves, and flexible cilia covering the outer surface of large cells, executing oar-like rowing motion, and beating in coordination. Self propulsion is achieved by cyclic shape changes described by time periodic functions (*swimming strokes*). A notable exception is given by the rotating flagella of bacteria, which rely on a submicron-size rotary motor capable of turning the axis of an helix without alternating between clockwise and anticlockwise directions.

Swimming microrobots Prototypes of artificial microswimmers have already been realized, and it is hoped that they can evolve into working tools in biomedicine. They should consist of minimally invasive, small-scale self-propelled devices engineered for drug delivery, diagnostic, or therapeutic purposes.

Definition of the Subject

Swimming, i. e., being able to advance in a fluid in the absence of external propulsive forces by performing cyclic shape changes, is particularly demanding at low Reynolds numbers (Re). This is the regime of interest for micro-organisms and micro-robots or nano-robots, where hydrodynamics is governed by Stokes equations. Thus, besides the rich mathematics it generates, low Re propulsion is of great interest in biology (How do microorganism swim? Are their strokes optimal and, if so, in which sense? Have these optimal swimming strategies been selected by evolutionary pressure?) and biomedicine (can small-scale self-propelled devices be engineered for drug delivery, diagnostic, or therapeutic purposes?).

For a microscopic swimmer, moving and changing shape at realistically low speeds, the effects of inertia are negligible. This is true for both the inertia of the fluid and the inertia of the swimmer. As pointed out by Taylor [10], this implies that the swimming strategies employed by bacteria and unicellular organism must be radically different from those adopted by macroscopic swimmers such as fish or humans. As a consequence, the design of artificial microswimmers can draw little inspiration from intuition based on our own daily experience.

Taylor's observation has deep implications. Based on a profound understanding of low Re hydrodynamics, and on a plausibility argument on which actuation mechanisms are physically realizable at small length scales, Berg postulated the existence of a sub-micron scale rotary motor propelling bacteria [5]. This was later confirmed by experiment.

Introduction

In his seminal paper *Life at low Reynolds numbers* [8], Purcell uses a very effective example to illustrate the subtleties involved in microswimming, as compared to the swimming strategies observable in our mundane experience. He argues that at low Re, any organism trying to swim adopting the reciprocal stroke of a scallop, which moves by opening and closing its valves, is condemned to the frustrating experience of not having advanced at all at the end of one cycle.

This observation, which became known as the *scallop theorem*, started a stream of research aiming at finding the simplest mechanism by which cyclic shape changes may lead to effective self propulsion at small length scales. Purcell's proposal was made of a chain of three rigid links moving in a plane; two adjacent links swivel around joints and are free to change the angle between them. Thus, shape is described by two scalar parameters (the angles between adjacent links), and one can show that, by changing them independently, it is possible to swim.

It turns out that the mechanics of swimming of Purcell's three-link creature are quite subtle, and a detailed understanding has started to emerge only recently [4,9]. In particular, the direction of the average motion of the center of mass depends on the geometry of both the swimmer and of the stroke, and it is hard to predict by simple inspection of the shape of the swimmer and of the sequence of movements composing the swimming stroke. A radical simplification is obtained by looking at axisymmetric swimmers which, when advancing, will do so by moving along the axis of symmetry. Two such examples are the three-sphere-swimmer in [7], and the push-me-pull-you in [3]. In fact, in the axisymmetric case, a simple and complete mathematical picture of low Re swimming is now available, see [1,2].

The Mathematics of Swimming

This article focuses, for simplicity, on swimmers having an axisymmetric shape Ω and swimming along the axis of symmetry, with unit vector \vec{l} . The configuration, or state s of the system is described by $N + 1$ scalar parameters: $s = \{x^{(1)}, \dots, x^{(N+1)}\}$. Alternatively, s can be specified by a position c (the coordinate of the center of mass along the symmetry axis) and by N shape parameters $\xi = \{\xi^{(1)}, \dots, \xi^{(N)}\}$. Since this change of coordinates is invertible, the generalized velocities $u^{(i)} := \dot{x}^{(i)}$ can be represented as linear functions of the time derivatives of position and shape:

$$(u^{(1)}, \dots, u^{(N+1)})^t = A(\xi^{(1)}, \dots, \xi^{(N)})(\dot{\xi}^{(1)}, \dots, \dot{\xi}^{(N)}, \dot{c})^t \quad (1)$$

where the entries of the $N + 1 \times N + 1$ matrix A are independent of c by translational invariance.

Swimming describes the ability to change position in the absence of external propulsive forces by executing a cyclic shape change. Since inertia is being neglected, the total drag force exerted by the fluid on the swimmer must also vanish. Thus, since all the components of the total force in directions perpendicular to \vec{l} vanish by symmetry,

self-propulsion is expressed by

$$0 = \int_{\partial\Omega} \sigma n \cdot \vec{l} \quad (2)$$

where σ is the stress in the fluid surrounding Ω , and n is the outward unit normal to $\partial\Omega$. The stress $\sigma = \eta(\nabla v + (\nabla v)^t) - p\text{Id}$ is obtained by solving Stokes equation outside Ω with prescribed boundary data $v = \bar{v}$ on $\partial\Omega$. In turn, \bar{v} is the velocity of the points on the boundary $\partial\Omega$ of the swimmer, which moves according to (1).

By linearity of Stokes equations, (2) can be written as

$$\begin{aligned} 0 &= \sum_{i=1}^{N+1} \varphi^{(i)}(\xi^{(1)}, \dots, \xi^{(N)}) u^{(i)} \\ &= A^t \Phi \cdot (\dot{\xi}^{(1)}, \dots, \dot{\xi}^{(N)}, \dot{c})^t \end{aligned} \quad (3)$$

where $\Phi = (\varphi^{(1)}, \dots, \varphi^{(N)})^t$, and we have used (1). Notice that the coefficients $\varphi^{(i)}$ relating drag force to velocities are independent of c because of translational invariance. The coefficient of \dot{c} in (3) represents the drag force corresponding to a rigid translation along the symmetry axis at unit speed, and it never vanishes. Thus (3) can be solved for \dot{c} , and we obtain

$$\dot{c} = \sum_{i=1}^N V_i(\xi^{(1)}, \dots, \xi^{(N)}) \dot{\xi}^{(i)} = V(\xi) \cdot \dot{\xi}. \quad (4)$$

Equation (4) links positional changes to shape changes through shape-dependent coefficients. These coefficients encode all hydrodynamic interactions between Ω and the surrounding fluid due to shape changes with rates $\dot{\xi}^{(1)}, \dots, \dot{\xi}^{(N)}$.

A stroke is a closed path γ in the space S of admissible shapes given by $[0, T] \ni t \mapsto (\xi^{(1)}, \dots, \xi^{(N-1)})$. Swimming requires that

$$0 \neq \Delta c = \int_0^T \sum_{i=1}^N V_i \dot{\xi}^{(i)} dt \quad (5)$$

i. e., that the differential form $\sum_{i=1}^N V_i d\xi^{(i)}$ is not exact.

The Scallop Theorem Proved

Consider a swimmer whose motion is described by a parametrized curve in two dimensions ($N = 1$), so that (4) becomes

$$\dot{c}(t) = V(\xi(t)) \dot{\xi}(t), \quad t \in \mathbb{R}, \quad (6)$$

and assume that $V \in L^1(S)$ is an integrable function in the space of admissible shapes and $\xi \in W^{1,\infty}(\mathbb{R}; S)$ is a Lipschitz-continuous and T -periodic function for some $T > 0$, with values in S .

Figure 1 is a sketch representing concrete examples compatible with these hypotheses. The axisymmetric case consists of a three-dimensional cone with axis along \vec{i} and opening angle $\xi \in [0, 2\pi]$ (an *axisymmetric octopus*). A non-axisymmetric example is also allowed in this discussion, consisting of two rigid parts (*valves*), always maintaining mirror symmetry with respect to a plane (containing \vec{i} and perpendicular to it) while swiveling around a joint contained in the symmetry plane and perpendicular to \vec{i} (a *mirror-symmetric scallop*), and swimming parallel to \vec{i} .

Among the systems that are *not* compatible with the assumptions above are those containing helical elements with axis of rotation \vec{i} , and capable of rotating around \vec{i} always in the same direction (call θ the rotation angle). Indeed, a monotone function $t \mapsto \theta(t)$ is not periodic.

The celebrated “scallop theorem” [8] states that, for a system like the one depicted in Fig. 1, the net displacement of the center of mass at the end of a periodic stroke will always vanish. This is due to the linearity of Stokes equation (which leads to symmetry under time reversals), and to the low dimensionality of the system (a one-dimensional periodic stroke is necessarily reciprocal). Thus, whatever forward motion is achieved by the scallop by closing its valves, it will be exactly compensated by a backward motion upon reopening them. Since the low Re world is unaware of inertia, it will not help to close the valves quickly and reopen them slowly. A precise statement and a rigorous short proof of the scallop theorem are given below.

Theorem 1 Consider a swimmer whose motion is described by

$$\dot{c}(t) = V(\xi(t))\dot{\xi}(t), \quad t \in \mathbb{R}, \quad (7)$$

with $V \in L^1(S)$. Then for every T -periodic stroke $\xi \in W^{1,\infty}(\mathbb{R}; S)$, one has

$$\Delta c = \int_0^T \dot{c}(t) dt = 0. \quad (8)$$

Proof Define the primitive of V by

$$\Psi(s) = \int_0^s V(\sigma) d\sigma \quad (9)$$

so that $\Psi'(\xi) = V(\xi)$. Then, using (7),

$$\begin{aligned} \Delta c &= \int_0^T V(\xi(t))\dot{\xi}(t) dt \\ &= \int_0^T \frac{d}{dt} \Psi(\xi(t)) dt \\ &= \Psi(\xi(T)) - \Psi(\xi(0)) = 0 \end{aligned}$$

by the T -periodicity of $t \mapsto \xi(t)$.

Optimal Swimming

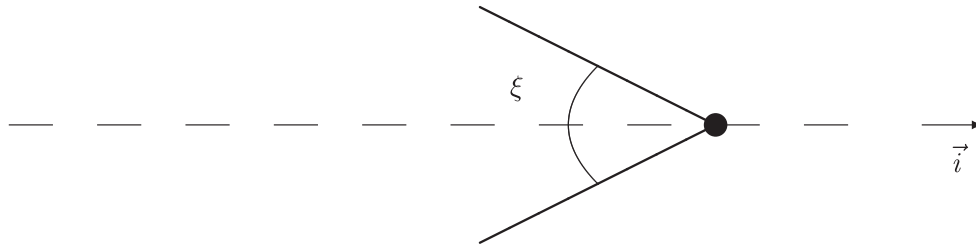
A classical notion of swimming efficiency is due to Lighthill [6]. It is defined as the inverse of the ratio between the average power expended by the swimmer during a stroke starting and ending at the shape $\xi_0 = (\xi_0^{(1)}, \dots, \xi_0^{(N)})$ and the power that an external force would spend to translate the system rigidly at the same average speed $\bar{c} = \Delta c/T$:

$$\text{Eff}^{-1} = \frac{\frac{1}{T} \int_0^T \int_{\partial\Omega} \sigma n \cdot v}{6\pi\eta L \bar{c}^2} = \frac{\int_0^1 \int_{\partial\Omega} \sigma n \cdot v}{6\pi\eta L (\Delta c)^2} \quad (10)$$

where η is the viscosity of the fluid, $L = L(\xi_0)$ is the effective radius of the swimmer, and time has been rescaled to a unit interval to obtain the second identity. The expression in the denominator in (10) comes from a generalized version of Stokes formula giving the drag on a sphere of radius L moving at velocity \bar{c} as $6\pi\eta L \bar{c}$.

Let $DN: H^{1/2}(\partial\Omega) \rightarrow H^{-1/2}(\partial\Omega)$ be the Dirichlet to Neumann map of the outer Stokes problem, i. e., the map such that $\sigma n = DNv$, where σ is the stress in the fluid, evaluated on $\partial\Omega$, arising in response to the prescribed velocity v on $\partial\Omega$, and obtained by solving the Stokes problem outside Ω . The expended power in (10) can be written as

$$\int_{\partial\Omega} \sigma n \cdot v = \int_{\partial\Omega} DN(v) \cdot v. \quad (11)$$



Biological Fluid Dynamics, Non-linear Partial Differential Equations, Figure 1
A mirror-symmetric scallop or an axisymmetric octopus

At a point $p \in \partial\Omega$, the velocity $v(p)$ accompanying a change of state of the swimmer can be written as a linear combination of the $u^{(i)}$

$$v(p) = \sum_{i=1}^{N+1} \mathcal{V}_i(p, \xi) u^{(i)} \quad (12)$$

$$= \sum_{i=1}^N \mathcal{W}_i(p, \xi) \dot{\xi}^{(i)}. \quad (13)$$

Indeed, the functions \mathcal{V}_i are independent of c by translational invariance, and (4) has been used to get (13) from the line above.

Substituting (13) in (11), the expended power becomes a quadratic form in $\dot{\xi}$

$$\int_{\partial\Omega} \sigma n \cdot v = (G(\xi) \dot{\xi}, \dot{\xi}) \quad (14)$$

where the symmetric and positive definite matrix $G(\xi)$ is given by

$$G_{ij}(\xi) = \int_{\partial\Omega} DN(\mathcal{W}_i(p, \xi)) \cdot \mathcal{W}_j(p, \xi) dp. \quad (15)$$

Strokes of maximal efficiency may be defined as those producing a given displacement Δc of the center of mass with minimal expended power. Thus, from (10), maximal efficiency is obtained by minimizing

$$\int_0^1 \int_{\partial\Omega} \sigma n \cdot v = \int_0^1 (G(\xi) \dot{\xi}, \dot{\xi}) \quad (16)$$

subject to the constraint

$$\Delta c = \int_0^1 V(\xi) \cdot \dot{\xi} \quad (17)$$

among all closed curves $\xi: [0, 1] \rightarrow S$ in the set S of admissible shapes such that $\xi(0) = \xi(1) = \xi_0$.

The Euler–Lagrange equations for this optimization problem are

$$-\frac{d}{dt}(G\dot{\xi}) + \frac{1}{2} \begin{pmatrix} \left(\frac{\partial G}{\partial \xi^{(1)}} \dot{\xi}, \dot{\xi} \right) \\ \vdots \\ \left(\frac{\partial G}{\partial \xi^{(N)}} \dot{\xi}, \dot{\xi} \right) \end{pmatrix} + \lambda \left(\nabla_{\xi} V - \nabla_{\xi}^t V \right) \dot{\xi} = 0 \quad (18)$$

where $\nabla_{\xi} V$ is the matrix $(\nabla_{\xi} V)_{ij} = \partial V_i / \partial \xi_j$, $\nabla_{\xi}^t V$ is its transpose, and λ is the Lagrange multiplier associated with the constraint (17).

Given an initial shape ξ_0 and an initial position c_0 , the solutions of (18) are in fact sub-Riemannian geodesics joining the states parametrized by (ξ_0, c_0) and $(\xi_0, c_0 + \Delta c)$ in the space of admissible states \mathcal{X} , see [1]. It is well known, and easy to prove using (18), that along such geodesics $(G(\gamma) \dot{\gamma}, \dot{\gamma})$ is constant. This has interesting consequences, because swimming strokes are often divided into a power phase, where $|G(\gamma)|$ is large, and a recovery phase, where $|G(\gamma)|$ is smaller. Thus, along optimal strokes, the recovery phase is executed quickly while the power phase is executed slowly.

The Three-Sphere Swimmer

For the three-sphere-swimmer of Najafi and Golestanian [7], see Fig. 2, Ω is the union of three rigid disjoint balls $B^{(i)}$ of radius a , shape is described by the distances x and y , the space of admissible shapes is $S = (2a, +\infty)^2$, and the kinematic relation (1) takes the form

$$\begin{aligned} u^{(1)} &= \dot{c} - \frac{1}{3}(2\dot{x} + \dot{y}) \\ u^{(2)} &= \dot{c} + \frac{1}{3}(\dot{x} - \dot{y}) \\ u^{(3)} &= \dot{c} + \frac{1}{3}(2\dot{y} + \dot{x}). \end{aligned} \quad (19)$$

Consider, for definiteness, a system with $a = 0.05$ mm, swimming in water. Calling $f^{(i)}$ the total propulsive force on ball $B^{(i)}$, we find that the following relation among forces and ball velocities holds

$$\begin{pmatrix} f^{(1)} \\ f^{(2)} \\ f^{(3)} \end{pmatrix} = R(x, y) \begin{pmatrix} u^{(1)} \\ u^{(2)} \\ u^{(3)} \end{pmatrix} \quad (20)$$

where the symmetric and positive definite matrix R is known as the resistance matrix. From this last equation, using also (19), the condition for self-propulsion $f^{(1)} + f^{(2)} + f^{(3)} = 0$ is equivalent to

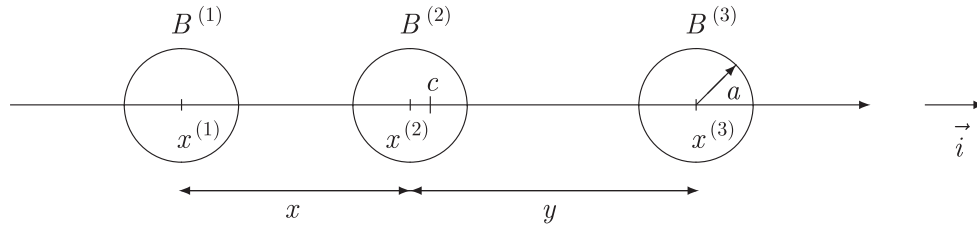
$$\dot{c} = V_x(x, y)\dot{x} + V_y(x, y)\dot{y}, \quad (21)$$

where

$$V_x(x, y) = \frac{Re_c \cdot (e_c \times e_y)}{Re_c \cdot (e_x \times e_y)} \quad (22)$$

$$V_y(x, y) = -\frac{Re_c \cdot (e_c \times e_x)}{Re_c \cdot (e_x \times e_y)}. \quad (23)$$

Moreover, $e_x = (-1, 1, 0)^t$, $e_y = (0, -1, 1)^t$, $e_c = (1/3, 1/3, 1/3)^t$.



Biological Fluid Dynamics, Non-linear Partial Differential Equations, Figure 2
Swimmer's geometry and notation

Biological Fluid Dynamics, Non-linear Partial Differential Equations, Table 1

Energy consumption (10^{-12} J) for the three strokes of Fig. 3 inducing the same displacement $\Delta c = 0.01$ mm in $T = 1$ s

Optimal stroke	Small square stroke	Large square stroke
0.229	0.278	0.914

Given a stroke $\gamma = \partial\omega$ in the space of admissible shapes, condition (5) for swimming reads

$$0 \neq \Delta c = \int_0^T (V_x \dot{x} + V_y \dot{y}) dt = \int_\omega \text{curl} V(x, y) dx dy \quad (24)$$

which is guaranteed, in particular, if $\text{curl} V$ is bounded away from zero. Strokes of maximal efficiency for a given initial shape (x_0, y_0) and given displacement Δc are obtained by solving Eq. (18). For $N = 2$, this becomes

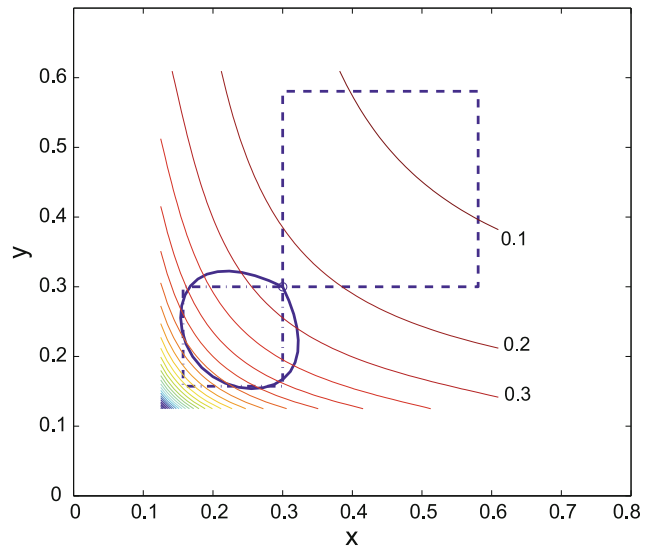
$$-\frac{d}{dt}(G\dot{\gamma}) + \frac{1}{2} \left(\begin{pmatrix} \partial_x G \dot{\gamma} \\ \partial_y G \dot{\gamma} \end{pmatrix} \right) + \lambda \text{curl} V(\gamma) \dot{\gamma}^\perp = 0 \quad (25)$$

where $\partial_x G$ and $\partial_y G$ stand for the x and y derivatives of the 2×2 matrix $G(x, y)$.

It is important to observe that, for the three-sphere swimmer, all hydrodynamic interactions are encoded in the shape dependent functions $V(x, y)$ and $G(x, y)$. These can be found by solving a two-parameter family of outer Stokes problems, where the parameters are the distances x and y between the three spheres. In [1], this has been done numerically via the finite element method: a representative example of an optimal stroke, compared to two more naive proposals, is shown in Fig. 3.

Future Directions

The techniques discussed in this article provide a head start for the mathematical modeling of microscopic swimmers, and for the quantitative optimization of their strokes. A complete theory for axisymmetric swimmers is already available, see [2], and further generalizations to



Biological Fluid Dynamics, Non-linear Partial Differential Equations, Figure 3

Optimal stroke and square strokes which induce the same displacement $\Delta c = 0.01$ mm in $T = 1$ s, and equally spaced level curves of $\text{curl} V$. The small circle locates the initial shape $\xi_0 = (0.3 \text{ mm}, 0.3 \text{ mm})$

arbitrary shapes are relatively straightforward. The combination of numerical simulations with the use of tools from sub-Riemannian geometry proposed here may prove extremely valuable for both the question of adjusting the stroke to global optimality criteria, and of optimizing the stroke of complex swimmers. Useful inspiration can come from the sizable literature on the related field dealing with control of swimmers in a perfect fluid.

Bibliography

Primary Literature

1. Alouges F, DeSimone A, Lefebvre A (2008) Optimal strokes for low Reynolds number swimmers: an example. *J Nonlinear Sci* 18:277–302
2. Alouges F, DeSimone A, Lefebvre A (2008) Optimal strokes for low Reynolds number axisymmetric swimmers. Preprint SISSA 61/2008/M
3. Avron JE, Kenneth O, Oakmin DH (2005) Pushmepullyou: an efficient micro-swimmer. *New J Phys* 7:234–1–8

4. Becker LE, Koehler SA, Stone HA (2003) On self-propulsion of micro-machines at low Reynolds numbers: Purcell's three-link swimmer. *J Fluid Mechanics* 490:15–35
5. Berg HC, Anderson R (1973) Bacteria swim by rotating their flagellar filaments. *Nature* 245:380–382
6. Lighthill MJ (1952) On the Squirming Motion of Nearly Spherical Deformable Bodies through Liquids at Very Small Reynolds Numbers. *Comm Pure Appl Math* 5:109–118
7. Najafi A, Golestanian R (2004) Simple swimmer at low Reynolds numbers: Three linked spheres. *Phys Rev E* 69:062901-1–4
8. Purcell EM (1977) Life at low Reynolds numbers. *Am J Phys* 45:3–11
9. Tan D, Hosoi AE (2007) Optimal stroke patterns for Purcell's three-link swimmer. *Phys Rev Lett* 98:068105-1–4
10. Taylor GI (1951) Analysis of the swimming of microscopic organisms. *Proc Roy Soc Lond A* 209:447–461

Books and Reviews

- Agrachev A, Sachkov Y (2004) Control Theory from the Geometric Viewpoint. In: *Encyclopaedia of Mathematical Sciences*, vol 87, Control Theory and Optimization. Springer, Berlin
- Childress S (1981) Mechanics of swimming and flying. Cambridge University Press, Cambridge
- Happel J, Brenner H (1983) Low Reynolds number hydrodynamics. Nijhoff, The Hague
- Kanso E, Marsden JE, Rowley CW, Melli-Huber JB (2005) Locomotion of Articulated Bodies in a Perfect Fluid. *J Nonlinear Sci* 15: 255–289
- Koiller J, Ehlers K, Montgomery R (1996) Problems and Progress in Microswimming. *J Nonlinear Sci* 6:507–541
- Montgomery R (2002) A Tour of Subriemannian Geometries, Their Geodesics and Applications. AMS Mathematical Surveys and Monographs, vol 91. American Mathematical Society, Providence

Biological Models of Molecular Network Dynamics

HERBERT M. SAURO

Department of Bioengineering,
University of Washington, Seattle, USA

Article Outline

[Glossary](#)
[Definition of the Subject](#)
[Introduction](#)
[Modeling Approaches](#)
[Deterministic Modeling](#)
[Stoichiometry Matrix](#)
[System Equation](#)
[Theoretical Approaches to Modeling](#)
[Negative Feedback Systems](#)
[FeedForward Systems](#)
[Positive Feedback](#)

Future Prospects

Bibliography

Glossary

Deterministic continuous model A mathematical model where the variables of the model can take any real value and where the time evolution of the model is set by the initial conditions.

Stochastic discrete model A mathematical model where the variables of the model take on discrete values and where the time evolution of the model is described by a set of probability distributions.

Definition of the Subject

Understanding the operation cellular networks is probably one of the most challenging and intellectually exciting scientific fields today. With the availability of new experimental and theoretical techniques our understanding of the operation of cellular networks has made great strides in the last few decades. An important outcome of this work is the development of predictive quantitative models. Such models of cellular function will have a profound impact on our ability of manipulate living systems which will lead to new opportunities for generating energy, mitigating our impact on the biosphere and last but not least, opening up new approaches and understanding of important disease states such as cancer and aging.

Introduction

Cellular networks are some of the most complex natural systems we know. Even in a “simple” organism such as *E. coli*, there are at least four thousand genes with many thousands of interactions between molecules of many different sizes [11]. In a human cell the number of interactions is probably orders of magnitude larger. Why all this complexity? Presumably the earliest living organisms were much simpler than what we find today but competition for resources and the need to adapt in unfavorable conditions must have led to the development of sensory and decision-making capabilities above and beyond the basic requirements for life. What we see today in almost all living organisms are complex signaling and genetic networks whose complexity and subtlety is beyond most man-made technological systems [87].

Over the last sixty or so years, biochemists and molecular biologists have identified many of the components in living cells and have traced out many of the interactions that delineate cellular networks. What emerges is a picture that would be familiar to many control engineers

Model-based design in synthetic biology, linking continuous differential equations with discrete logical network dynamics

Computing Science

School of Computing Science, Newcastle University

School of Computer Engineering, NTU, Singapore

Supervisory Team

- Dr. Jason Steggles: <http://tinyurl.com/hyu69u8>
<http://www.ncl.ac.uk/computing/people/profile/jason.steggles>
- Professor Anil Wipat: <http://tinyurl.com/ppqsl35>
- Dr. Goksel Misirli: <http://homepages.cs.ncl.ac.uk/goksel.misirli/>
- Assistant Professor Jie Zheng, School of Computer Engineering, NTU: <http://www.ntu.edu.sg/home/zhengjie>
- Assistant Professor Chueh Loo Poh, School of Chemical and Biomedical Engineering, NTU: <http://www3.ntu.edu.sg/home/clpoh/>
- The Lead Supervisor is Early Career or newly hired Staff ☐

Key Words

Mathematical modelling, synthetic biology, differential equations, logical modelling, network dynamics

Overview

Computational cell modelling has been successfully used to study the complexity of biological networks that consist of multiple feedback regulations¹. Particularly, the use of continuous differential equations as a well-established method to formulate the molecular interactions in biological networks is crucial for building quantitative models. However, the use of differential equations often involves many unknown biochemical mechanisms and parameters that have to be estimated, and is hard to scale to large networks. In order to overcome the issues of parameter values and lack of scalability, qualitative and logical modelling (such as Boolean networks) has been increasingly used to model large biological networks².

Both differential equations and logical modelling play important roles in understanding a cell's molecular systems³. A few studies have linked the continuous differential equations with discrete logical models^{4,5}. Recently, logical modelling with model reduction has been shown to be a powerful technique for studying large networks⁶. This project aims to conduct a study to use logical (Boolean or multilevel) network or

extended logical model (e.g. Fuzzy logic) to scale the dynamic modelling of ODEs to relatively large regulatory networks (say, of size > 100 genes). The aim of the project is to research less granular approaches to simulation and model-based design in synthetic biology⁷ and will investigate the use of modular, composable, qualitative Boolean models. The system will be tested in the model-based design and construction of regulatory networks in *Escherichia coli* and *Bacillus subtilis*.

Methodology

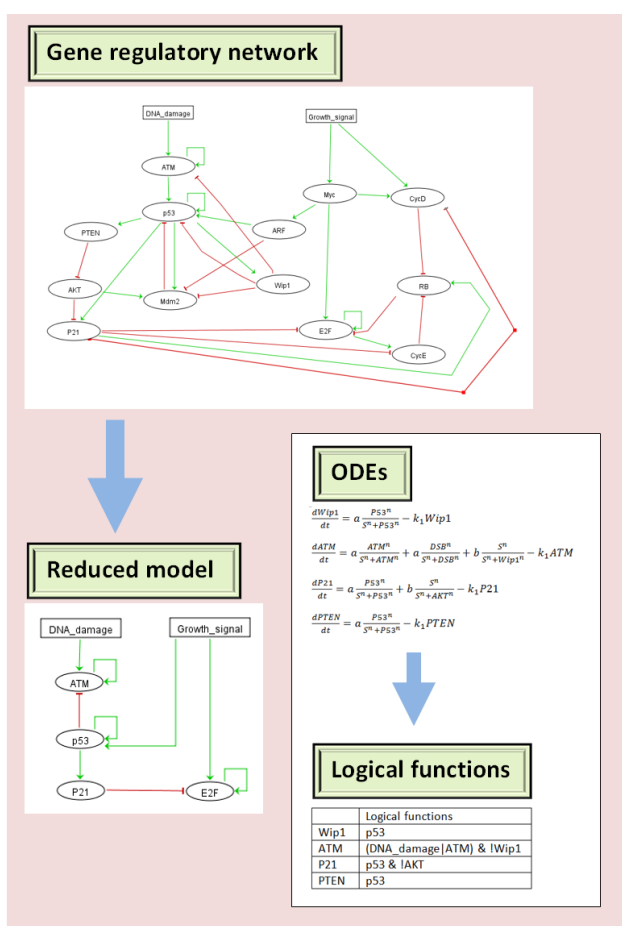
Designing large/complex biological systems requires *in silico* modelling to guide more predictive design. Traditional dynamic modelling based on differential equations requires the details for the design specification and does not scale well for large systems. Hence, we need an intermediate form of modelling that is able to capture the required dynamic/functional behaviour but is more abstract and therefore more scalable. The methods used in this project will involve non-linear ordinary differential equations and logical (Boolean or multilevel) modelling. Existing computer software for ODE modelling (e.g. XPPAUT, CellDesigner) and logical modelling software (e.g.

GINsim) will be used. In qualitative Boolean models, molecular interactions (activation or inhibition) are represented as logical functions and implemented with two types of updating schemes: synchronous and asynchronous. The synchronous mode assumes that all processes happen at the same time and asynchronous update assumes that processes happen at different times. State transition graphs will be used to visualize the activities of gene expression and signaling pathways. The recent technique of model reduction^{6,8} in asynchronous update will be used to map the large biological system to a reduced model that still contains the essential dynamics of the system such as stable states and conserved attractors. Furthermore, high-performance computing software (e.g. cloud computing, GPGPU) suitable for application to systems and synthetic biology, as well as biomedicine, will be developed. Synthetic biology

plan. 2nd Year: Construct the method/software for logical modelling in synthetic biology. Mathematical models developed. Model composition explored. Construct and test the synthetic regulatory networks
3rd Year: Documentation and testing of the method/software or models. Conference paper written. Model analysis and interpretation of the results. 4th Year: Thesis preparation and writing scientific publications. Journal paper produced.

Training & Skills

The student will receive training in computational and mathematical modelling, model analysis and interpretation in a biological context. Training in conceptual model construction or computational cell biology and molecular biology will also be provided. The Ph.D. project will primarily be based at School of Computing Science, Newcastle University where computational and experimental work will be carried out. The student will attend regular ICOS and ASL meetings. Research will be communicated in research seminars at the both Universities and at international conferences. Dr. Zheng and Dr. Poh (NTU) will provide additional complementary training in Boolean and advanced biological systems modelling. The student will be supported to travel to Singapore for about a year as part of the exchange program, and thereby gain international experience of learning and living. Through this project, the student will become an independent researcher in solving problems and perform interdisciplinary research.



requires the design and composition of models from parts and so theories for composing logical models will be investigated and developed to provide a basis for model construction and analysis.

Timeline

1st Year: Literature review of mathematical modelling of cell and synthetic biology in continuous differential equations and discrete logical modelling methods. Technical background or knowledge in computational cell modelling and dynamical system theory. Project

References & Further Reading

- Goldbeter, A. Computational approaches to cellular rhythms. *Nature* 420, 238-245 (2002).
- Naldi, A. et al. Cooperative development of logical modelling standards and tools with CoLoMoTo. *Bioinformatics* 31, 1154-1159 (2015).
- Le Novère, N. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics* (2015).
- Abou-Jaoudé, W., Ouattara, D.A. & Kaufman, M. From structure to dynamics: frequency tuning in the p53–mdm2 network: I. logical approach. *Journal of theoretical biology* 258, 561-577 (2009).
- Snoussi, E.H. Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach. *Dynamics and stability of Systems* 4, 565-583 (1989).
- Bérenguier, D. et al. Dynamical modeling and analysis of large cellular regulatory networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 23, 025114 (2013).
- Crook, N. & Alper, H.S. Model-based design of synthetic, biological systems. *Chemical Engineering Science* 103, 2-11 (2013).
- Steggles LJ, Banks R, Shaw O, Wipat A. Qualitatively modelling and analysing genetic regulatory networks: A Petri net approach. *Bioinformatics* 2007, 23(3), 336-343

Further Information

Professor Anil Wipat: anil.wipat@ncl.ac.uk

Assistant Professor Jie Zheng: ZhengJie@ntu.edu.sg

Dynamical robustness of biological networks with hierarchical distribution of time scales

A.N. Gorban and O. Radulescu

Abstract: Concepts of distributed robustness and r -robustness proposed by biologists to explain a variety of stability phenomena in molecular biology are analysed. Then, the robustness of the relaxation time using a chemical reaction description of genetic and signalling networks is discussed. First, the following result for linear networks is obtained: for large multiscale systems with hierarchical distribution of time scales, the variance of the inverse relaxation time (as well as the variance of the stationary rate) is much lower than the variance of the separate constants. Moreover, it can tend to 0 faster than $1/n$, where n is the number of reactions. Similar phenomena are valid in the nonlinear case as well. As a numerical illustration, a model of signalling network is used for the important transcription factor NF κ B.

1 Introduction

Robustness, defined as stability against external perturbations and internal variability, represents a common feature of living systems. The fittest organisms are those that resist to diseases, to imperfections or damages of regulatory mechanisms, and that can function reliably in various conditions. There are many theories that describe, quantify and explain robustness. Waddington's canalisation [1] was formalised by Thom [2] as structural stability of attractors under perturbations. Many useful ideas on robustness have been imported from the theory of control of dynamical systems and of automata [3, 4]. The new field of systems biology places robustness in a central position among the living systems organising principles, identifying redundancy, modularity and negative feedback as sources of robustness [5–7].

In this paper, we provide some justification to a different, less understood source of robustness.

Early insights into this problem can be found in the von Neumann's discussion of robust coupling schemes of automata [8]. von Neumann noticed the intrinsic relation between randomness and robustness. Quoting him 'without randomness, situations may arise where errors tend to be amplified instead of cancelled out; for example it is possible that the machine remembers its mistakes, and thereafter perpetuates them'. To cope with this, von Neumann introduces multiplexing and random perturbations in the design of robust automata.

Related to this is Wagner's concept of distributed robustness that 'emerges from the distributed nature of many biological systems, where many (and different) parts contribute to system functions' [9, 10]. To a certain extent, distributed robustness and control are antithetical. In a robust system, any localised perturbation should have only small effects. Robust properties should not depend on only one, but on

many components and parameters of the system. A weaker version of distributed robustness is the r -robustness, when r or less changes have small effect on the functioning of the system [11].

Molecular biology offers numerous examples of distributed robustness and of r -robustness. Single knockouts of developmental genes in the fruit fly have localised effects and do not lead to instabilities [12]. Complex diseases are the result of deregulation of many genetic pathways [13]. Transcriptional control of metazoa is based on promoter and enhancer regulating DNA regions that collect influences from many proteins [14]. Networks of regulating micro-RNA could be key players in canalising genetic developmental programmes [15]. Interestingly, computer models of gene regulation networks [16] have distributed robustness with respect to variations of their parameters. Flux balance analysis in-silico studies of the effects of multiple knockouts in *Saccharomyces cerevisiae* showed that yeast metabolism is less robust to multiple attacks than to single attacks [11].

Let us formulate the problem mathematically. A property M of the biological system is a function of several parameters of the system, $M = f(K_1, K_2, \dots, K_n)$. Let us assume that the parameters (K_1, K_2, \dots, K_n) are independent, random variables. There are various causes of variability: mutations, across individuals variability, changes in the functional context, and so on. For different causes, the distribution of parameters may be significantly different. For example, if parameters change because of random deletion of some reactions, then the appropriate model is $K_i = K_i^0$, with probability $1 - p$, and $K_i = 0$, with probability p . On the other hand, the fluctuation of enzyme activity can be formalised as a distribution of K_i with continuous density.

Considering independent and identical distributions of K_i , we can give two basic examples of functions $M = f(K_1, K_2, \dots, K_n)$ that have much less variability than individual K_i . The first example considers the average value of K_i , that is, $M = \sum_i K_i/n$: $\text{Var}(M) = \frac{1}{n} \sum_i \text{Var}(K_i)$. If all $\text{Var}(K_i) = \text{Var}(K)$, then $\text{Var}(M) = \text{Var}(K)/n$. The second example considers the order statistics [17]: $M = K_{(l)}$ or $M = K_{(n-l)}$, where K_l is the l st parameter in the order $K_{(1)} \geq K_{(2)} \geq \dots K_{(n)}$. When l does not depend on n (or is

uniformly bounded), $\text{Var}(M)$ goes to 0 when $n \rightarrow \infty$ as $1/n^2$. This is faster than for the average.

Following these examples, for definitions of robustness we can start from the inequality: $\text{Var}(M) \ll \text{Var}(K)$, where $\text{Var}(K_i) = \text{Var}(K)$ for all $i = 1, \dots, n$.

To avoid the problem of units and supposing that $M, K_i > 0$, we can use logarithmic scale.

Definition 1: M is robust with respect to distributed variations if the log-variance of M is much smaller than the log-variance of any of the parameters. Let $\text{Var}(\log K_i) = \text{Var}(\log K)$ for all $i = 1, \dots, n$. Then

$$\text{Var}(\log M) \ll \text{Var}(\log K) \quad (1)$$

Let us consider r -index subsets $I_r = \{i_1, i_2, \dots, i_r\} \subset \{1, 2, \dots, n\}$ for given r . Let $K_i^0, i = 1, \dots, n$, be the central values of the parameters. For given I_r , the perturbed values K_i are obtained by multiplying r -selected central values by independent random scales $s_i > 0, i = 1, \dots, r, K_i = K_i^0 s_i, i \in I_r, \text{Var}(\log s_i) = \text{Var}(\log s)$ for all $i \in I_r$, and $\text{Var}(K_j) = 0$ for all $j \notin I_r$.

Definition 2: M is robust with respect to r variations or r -robust if for any I_r

$$\text{Var}(\log M) \ll \text{Var}(\log s) \quad (2)$$

r -robustness holds if (2) is valid for any deterministic choice of r targets. If the target set I_r is randomly chosen, we shall speak of weak r -robustness. We call *robustness index* the maximal value of r such that the system is r -robust.

The above definitions are inspired from biological ideas. Our first definition corresponds to Wagner's distributed robustness [10]. It expresses the fact that M is not sensitive to random variations of the parameters. r -robustness has been defined in [11] as resistance with respect to multiple mutations. r -robustness can also be interpreted as functional redundancy (this is different from the structural redundancy of Wagner [10], meaning that many genes code for the same protein) meaning that the property M is collectively controlled by more than r parameters, and cannot be considerably influenced by changing a number of parameters $\leq r$. One should also notice the introduction of a new concept. Even if there are r critical targets (for instance genes whose mutations lead to large effects), the probability of hitting these r targets randomly could be small. We have introduced the weak r -robustness to describe this situation.

Robustness with respect to distributed variations can be a consequence of the Gromov–Talagrand concentration of measure in high dimensional metric-measure spaces [18, 19]. In Gromov's theory, the concentration has a geometrical significance: objects in very high-dimension look very small when they are observed via the values of real functions with bounded rate of change (1-Lipschitzian functions: $|f(x) - f(y)| \leq \|x - y\|$). This represents an important generalization of the law of large numbers and has many applications in mathematics. In this paper, we shall discuss two types of concentration effects: cube concentration that applies to sums or averages and the faster simplex concentration that applies to order statistics (see above).

In both definitions, we propose a robustness criterion. There are two difficulties in relation to this. First, it is difficult to impose an objective criterion for what ' \ll ' means in (1) and (2). In the sense of asymptotic behaviour, it is clear that $\text{Var}(\log M)/\text{Var}(\log K) \rightarrow 0$ when $n \rightarrow \infty$. When concentration phenomena are present, the ratio $\rho = \text{Var}(\log M)/\text{Var}(\log K)$ should scale like $1/n$ or even like

$1/n^2$, where n is the number of independent variable parameters. In practice, we always consider finite number of parameters. In this case, ρ is finite and robustness means that the ratio is smaller than some threshold, $\rho < \theta$. Obviously, when $\text{Var}(\log M)/\text{Var}(\log K) \geq 1$, the system is not robust, hence $\theta < 1$. In general, we should study dependence $\text{Var}(\log M)$ on $\text{Var}(\log K)$ and n (or r – for r -robustness). An example of such study for nonlinear signalling network is presented below. The dependence $\text{Var}(\log M)$ on $\text{Var}(\log K)$ may be nonlinear, but often remains close to a piecewise linear function. In that case, the slopes $d\text{Var}(\log M)/d\text{Var}(\log K)$ are more informative than the ratios $\text{Var}(\log M)/\text{Var}(\log K)$. One can reformulate definitions of robustness and r -robustness using these slopes.

Second, some homogeneity of the parameters is implicit. For instance, in this paper, K_i are kinetic parameters. Because of the exponential Arrhenius law, log-variances of the kinetic parameters can be arbitrarily large with respect to log-variances of the activation energies. A robust property with respect to the kinetic parameters may be artificially declared non-robust with respect to activation energies. Furthermore, we want to exclude trivial cases when M does not depend on K_i . To avoid problems, we can consider only positively homogeneous functions of degree one: $f(\alpha K_1, \alpha K_2, \dots, \alpha K_n) = \alpha f(K_1, K_2, \dots, K_n)$ for positive α . If K_i are, for example, matrix elements of a matrix \mathbf{K} , then eigenvalues λ_i of \mathbf{K} are homogeneous functions of K_i of degree one. If for all λ_i , the real part is non-positive, $\text{Re} \lambda_i \leq 0$, and non-zero purely imaginary eigenvalues do not exist, then inverse relaxation time $1/\tau = \min\{-\text{Re} \lambda_i | \lambda_i \neq 0\}$ is positively homogeneous function of K_i of degree one. If right-hand side of a system of differential equations is a homogeneous linear function of K_i , $\dot{x} = \sum K_i \phi_i(x)$, then eigenvalues of Jacobian matrices at any point, inverse periods of limit cycles, and inverse relaxation times are positively homogeneous functions of K_i of degree one. In logarithmic scale, variance of $\log M$ is the same as of $\log(M^{-1})$. Hence, we can consider in Definitions 1, 2 positively homogeneous functions of degrees 1 and -1 together. This is enough for our purposes in this paper.

In this paper, we choose a signalling module example as an illustration of the various concepts of robustness. The robust property that we study here is the relaxation time of a biological molecular system modelled as a network of chemical reactions. Relaxation time is an important issue in chemical kinetics, but there exists biological specifics. A biological system is a hierarchically structured open system. Any biological model is necessarily a submodel of a bigger one. After a change of the external conditions, a cascade of relaxations takes place and the spatial extension of a minimal model describing this cascade depends on time. Timescales are important in signalling between cells and between different parts of an organism. It is therefore important to know how the relaxation time depends on the size and the topology of a network and how robust is this time against variations of the kinetic constants.

In this paper, first, we extend the classical results on limiting steps of stationary states of one-route cyclic linear networks onto dynamic of relaxation of any linear network. This allows us to relate the relaxation time of a linear network with hierarchical distribution of time scales to low-order statistics of the network constants and to prove the distributed robustness of this relaxation time. Last, using a model of the NF κ B signalling module as an example, we show that similar results apply to nonlinear networks. For this nonlinear network, the robustness of another characteristic time, the period of its oscillations is studied as well.

2 Limitation of relaxation in linear reaction networks

First, we consider a linear network of chemical reactions. In a linear network, all the reactions are of the type $A_i \rightarrow A_j$, and the reaction rates r_{ji} are proportional to the reagents A_i concentration: $r_{ji} = k_{ji}c_i$.

The dynamics of the network is described by

$$\dot{c}_i = \sum_{j, j \neq i} (k_{ij}c_j - k_{ji}c_i) \text{ or } \dot{c} = Kc \quad (3)$$

where $K = (K_{ij})$, for $i \neq j$, K_{ij} is the reaction rate constant k_{ij} of the reaction producing A_i and consuming A_j (this is zero if no such reaction exists), and $K_{ii} = -\sum_{j, j \neq i} k_{ji}$.

For the analysis of kinetic systems, linear conservation laws and positively invariant polyhedra are important. A linear conservation law is a linear function defined on the concentrations $b(c) = \sum_{i=1}^q b_i c_i$ (q is the number of reagents), whose value is preserved by the dynamics (3). The conservation laws coefficient vectors b_i are left eigenvectors of the matrix K corresponding to the zero eigenvalue. For any kinetic system, $b^0 = \sum_{i=1}^q c_i$ is the conservation law. A set E is positively invariant with respect to kinetic equations (3), if any solution $c(t)$ that starts in E at time t_0 [$c(t_0) \in E$] belongs to E for $t > t_0$ [$c(t) \in E$ if $t > t_0$]. It is straightforward to check that the standard simplex $\Sigma = \{c | c_i \geq 0, \sum_i c_i = 1\}$ is positively invariant set for kinetic equation (3): just check that if $c_i = 0$ for some i , and all $c_j \geq 0$ then $\dot{c}_i \geq 0$. This simple fact immediately implies the following properties of K :

- all eigenvalues λ of K have non-positive real parts, $\text{Re} \lambda \leq 0$, because solutions cannot leave Σ in positive time;
- If $\text{Re} \lambda = 0$, then $\lambda = 0$, because intersection of Σ with any plane is a polygon, and a polygon cannot be invariant with respect to rotations of sufficiently small angles;
- The Jordan cell of K that corresponds to zero eigenvalue is diagonal, because all solutions should be bounded in Σ for positive time.
- The shift in time operator $\exp(Kt)$ is a contraction in the l_1 norm for $t > 0$: for positive t and any two solutions of (3) $c(t)$, $c'(t) \in \Sigma$

$$\sum_i |c_i(t) - c'_i(t)| \leq \sum_i |c_i(0) - c'_i(0)|$$

Vertices of Σ correspond to components A_i (in each vertex only one $c_i \neq 0$). For any initial state, $c(0) \in \Sigma$; there exists a limit state $\lim_{t \rightarrow \infty} \exp(Kt)c(0)$. We call a linear network weakly ergodic, if these limits coincide for all $c(0) \in \Sigma$. This is equivalent to uniqueness of steady state in Σ . The steady-state $c^* \in \Sigma$ for weakly ergodic network is not obligatory strictly positive, some of c_i^* could be zero. This is the difference from ergodic networks that have strictly positive steady state.

The ergodicity of the network follows from its topological properties. A non-empty subset V of the reaction digraph vertices forms a sink, if there are no oriented edges from $A_i \in V$ to any $A_j \notin V$. For example, in the reaction digraph $A_1 \leftarrow A_2 \rightarrow A_3$, the one-vertex sets $\{A_1\}$ and $\{A_3\}$ are sinks. A sink is minimal if it does not contain a strictly smaller sink. In the previous example, $\{A_1\}$ and $\{A_3\}$ are minimal sinks. Minimal sinks are also called ergodic components.

The following properties are equivalent:

1. the network is weakly ergodic.
2. for each two vertices A_i, A_j ($i \neq j$) we can find such a vertex A_k that an oriented paths exist from A_i to A_k and

from A_j to A_k . One of these paths can be degenerated: it might be $i = k$ or $j = k$.

3. the network has only one minimal sink (one ergodic component).
4. there is an unique linear conservation law, namely $b^0(c) = \sum_{i=1}^q c_i$; in other words, the zero eigenvalue of the matrix K is not degenerate.

Hence, the number of independent linear conservation laws is equal to the maximal number of disjoint ergodic components.

These properties of weakly ergodic reaction networks are well known in chemical kinetics [20]. They can be also extracted from the theory of Markov chains [21].

In the proof of this statement, the following transformation plays central role. Let $b^0(c), b^1(c), \dots, b^l(c)$ be independent linear conservation laws and $b^0(c) = \sum_i c_i$. The map $c \mapsto [b^1(c), \dots, b^l(c)]$ projects the simplex Σ onto the l -dimensional polyhedron B . Preimage of each point of B is a positively invariant polyhedron in Σ , and preimage of a vertex is a positively invariant face of Σ . The vertices of such a face form a sink (we identify components and vertices of Σ). The number of vertices in l -dimensional polyhedron B cannot be smaller than $l + 1$. So, if there are $l + 1$ independent, linear conservation laws, then there exist $l + 1$ disjoint sinks in reaction graph. Let us assume inverse: there exist l sinks, S_1, \dots, S_l . For each $c \in \Sigma$, the limit exists $c^*(c) = \lim_{t \rightarrow \infty} \exp(Kt)c$. The independent conservation laws b^j are $b^j(c) = \sum_{i \in S_j} c_i^*(c)$.

Now, let us suppose that the kinetic parameters are well separated and let us sort them in decreasing order: $k_{(1)} \gg k_{(2)} \gg \dots \gg k_{(n)}$. Let us also suppose that the network has only one ergodic component (when there are several ergodic components, each one has its longest relaxation time that can be found independently). We say that $k_{(r)}$, $1 \leq r \leq n$ is the *ergodicity boundary* if the network of reactions with parameters k_1, k_2, \dots, k_r is weakly ergodic, but the network with parameters k_1, k_2, \dots, k_{r-1} it is not. In other words, when eliminating reactions in decreasing order of their characteristic times, starting with the slowest one, the ergodicity boundary is the constant of the first reaction whose elimination breaks the ergodicity of the reaction digraph.

Relaxation to equilibrium of the network is multi-exponential, but the longest relaxation time is given by

$$\tau = \frac{1}{\min\{-\text{Re} \lambda_i | \lambda_i \neq 0\}} \quad (4)$$

An estimate of the longest relaxation time can be obtained by applying the perturbation theory for linear operators to the degenerated case of the zero eigenvalue of the matrix K . We have $K = K_{<r}(k_1, k_2, \dots, k_{r-1}) + k_r Q + o(k_r)$, where $K_{<r}$ is obtained from K by letting $k_r = k_{r+1} = \dots = k_n = 0$, Q is a constant matrix and $o(k_r)$ includes terms that are negligible relative to k_r . From equivalence of the properties (1)–(4), it follows that the zero eigenvalue is twice degenerate in $K_{<r}$ and not degenerate in $K_{<r} + k_r Q$. One gets the following estimate

$$\bar{a} \frac{1}{k_{(r)}} \geq \tau \geq \underline{a} \frac{1}{k_{(r)}} \quad (5)$$

where $\bar{a}, \underline{a} > 0$ are some positive functions of k_1, k_2, \dots, k_{r-1} (and of the reaction graph topology).

Two simplest examples give us the structure of the perturbation theory terms for $\min_{\lambda \neq 0} \{-\text{Re} \lambda\}$.

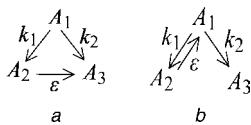


Fig. 1 Two basic examples of ergodicity boundary reaction

a Connection between ergodic components

b Connection from one ergodic component to element that is connected to the both ergodic components by oriented paths. In both cases, for $\varepsilon = 0$, the ergodic components are $\{A_2\}$ and $\{A_3\}$

1. For the reaction mechanism shown in Fig. 1a, $\min_{\lambda \neq 0} \{-\text{Re}\lambda\} = \varepsilon$, if $\varepsilon < k_1 + k_2$.
2. For the reaction mechanism shown in Fig. 1b, $\min_{\lambda \neq 0} \{-\text{Re}\lambda\} = \varepsilon k_2 / (k_1 + k_2) + o(\varepsilon)$, if $\varepsilon < k_1 + k_2$. For well-separated parameters, there exists a *trigger alternative*: if $k_1 \ll k_2$, then $\min_{\lambda \neq 0} \{-\text{Re}\lambda\} \simeq \varepsilon$; if, inverse, $k_1 \gg k_2$, then $\min_{\lambda \neq 0} \{-\text{Re}\lambda\} = o(\varepsilon)$.

More generally

$$\tau \simeq \frac{1}{ak_{(r)}} \quad (6)$$

with $a \lesssim 1$. This means that $1/k_{(r)}$ gives the lower estimate of the relaxation time, but τ could be larger. The detailed analysis of multiscale networks [22] shows that there is a trigger alternative too: if the constants are well separated, then either $a \simeq 1$ or $a \ll 1$.

Thus, the well-known concept of stationary reaction rates limitation by ‘narrow places’ or ‘limiting steps’ (slowest reaction) should be complemented by the *ergodicity* boundary limitation of relaxation time. It should be stressed that the relaxation process is limited not by the classical limiting steps (narrow places), but by the reactions that may be absolutely different. The simplest example of this kind is an irreversible catalytic cycle: the stationary rate is limited by the slowest reaction (the smallest constant), but the relaxation time is limited by the reaction constant with the second lowest value (in order to break the weak ergodicity of a cycle two reactions must be eliminated).

3 Robustness of relaxation time in linear systems

In general, for large multiscale systems, we observe concentration effects: the log-variance of the relaxation time is much lower than that of the separate constants. For linear networks, this follows from well-known properties of the order statistics [17]. For instance, if k_i are independent, log-uniform random variables, we have $\text{Var}[\log(k_{(r)})] \sim 1/n^2$. Here, we meet a “simplex-type” concentration ([19] pp. 234–236) and the log-variance of the relaxation time can tend to 0 faster than $1/n$, where n is the number of reactions.

For parameters whose logarithm is uniformly distributed in the interval $[0, 1]$, $k_{(r)}$ has a log-beta distribution $\log(k_{(r)}) \sim \mathcal{B}(r, n+1-r)$, i.e. for any $0 \leq a \leq b \leq 1$, $\mathbb{P}[a < \log(k_{(r)}) < b] = 1/B(r, n+1-r) \int_a^b x^{r-1} (1-x)^{n-r} dx$, where $B(r, n+1-r) = \int_0^1 x^{r-1} (1-x)^{n-r} dx$.

The above estimates for the variance of the order statistics, hence of relaxation time of linear networks, are based on identical distributions of the kinetic constants. A more realistic approach is to consider non-identical distributions with different means. Let δ be the average separation between mean parameters, in logarithmic scale (this separation is zero for identical distributions) and let $\Delta = n\delta$ be the spread of the means. Let us suppose that

all the parameters have the same variance $\text{Var}(\log k_i) = \text{Var}(\log k)$. When $\text{Var}(\log k) < \delta^2$, the overlap of distributions of successive parameters is improbable and one has $\text{Var}(\log k_{(r)}) = \text{Var}(\log k)$. When $\delta^2 < \text{Var}(\log k) < \Delta^2$, there is overlap and the variance of $\log k_{(r)}$ is limited by the distance δ , one has saturation: $\text{Var}(\log k_{(r)}) = \delta^2$. Finally, when $\Delta^2 < \text{Var}(\log k)$, we recover the case of identical distributions and one has simplex concentration $\text{Var}(\log k_{(r)}) = \text{Var}(\log k)/n^2$. The three regimes can be observed even for relaxation times of nonlinear models as will be discussed in Section 4.4.

Let us now discuss some design principles for robust networks. Suppose we have to construct a linear chemical reaction network. How to increase robustness of the largest relaxation times for this network? To be more realistic, let us take into account two types of network perturbation:

1. random noise in constants;
2. elimination of a link or of a node in reaction network.

Long routes are more robust for the perturbations of the first kind. So, the first recipe is simple: let us create long cycles! But longer cycles are destroyed by link or node elimination with higher probability. So, the second recipe is also simple: let us create a system with many alternative routes!

Finally, the resources are expensive, and we should create a network of minimal size.

Hence, we come to a new combinatorial problem. How to create a minimal network that satisfies the following restrictions

1. the length of each route is $> L$;
2. after destruction of arbitrarily chosen D_{links} and D_{nodes} , there remains at least one long route in the network.

To obtain the minimal network that fulfills the above constraints, we should include bridges between cycles, but the density of these bridges should be sufficiently low in order not to affect the length of the cycles significantly.

Additional restrictions could be involved. For example, we can discuss not all the routes, but productive routes only (that obligatory include some of the reaction steps).

For acyclic networks, we obtain similar recipes: long chains should be combined with bridges. A compromise between the chain length and number of bridges is needed.

4 Robustness of characteristic times in nonlinear systems: an example

4.1 Model

Our example is one of the most documented transcriptional regulation systems in eukaryote organisms: the signalling module of NF κ B. The response of this factor to a signal has been modelled by several authors [23–26].

The transcription factor NF κ B is a protein (actually a heterodimer made of two smaller molecules p50 and p65) that regulates the activity of more than one hundred genes and other transcription factors that are involved in the immune and stress response, apoptosis, and so on. NF κ B is thus the principal mediator of the response to cellular aggression and is activated by more than 150 different stimuli: bacteria, viral and bacterial products, mitogen agents and stress factors (radiations, ischemia, hypoxia, hepatic regeneration and drugs among which some anticancer drugs). NF κ B has complex regulation, including inhibitor degradation and production, translocation between nucleus and cytoplasm,

negative and positive feed-back. Under normal conditions, NF κ B is trapped in the cytoplasm where it forms a molecular complex with its inhibitor I κ B. Under this form, NF κ B cannot perform its regulatory function, the complex cannot penetrate the nucleus. A signal that can be modelled by a kinase (IKK) frees NF κ B by degrading its inhibitor. Free NF κ B enters the nucleus and regulates the transcription of many genes, among which the gene of its inhibitor I κ B and the gene of a protein A20 that inactivates the kinase.

Here, we would like to study the robustness of the characteristic times of a nonlinear molecular system. In particular, the double negative feed-back (via I κ B and A20) is responsible for oscillations of NF κ B activity under persistent stimulation [23–25]. We are thus interested in three characteristic times of the NF κ B model: the period and the damping time of the oscillations and the largest relaxation time (the damping of the oscillations is not necessarily the only relaxation process, therefore the damping time is not necessarily equal to the largest relaxation time).

We use the model introduced in [24] for the response of NF κ B module to a signal. This model is represented in Fig. 2. The first reaction of the model is the activation of the kinase. In the absence of a signal, the kinetic constant of the activation reaction is zero $k_1 = 0$, meaning that the kinase IKK remains inactive. The presence of a signal is

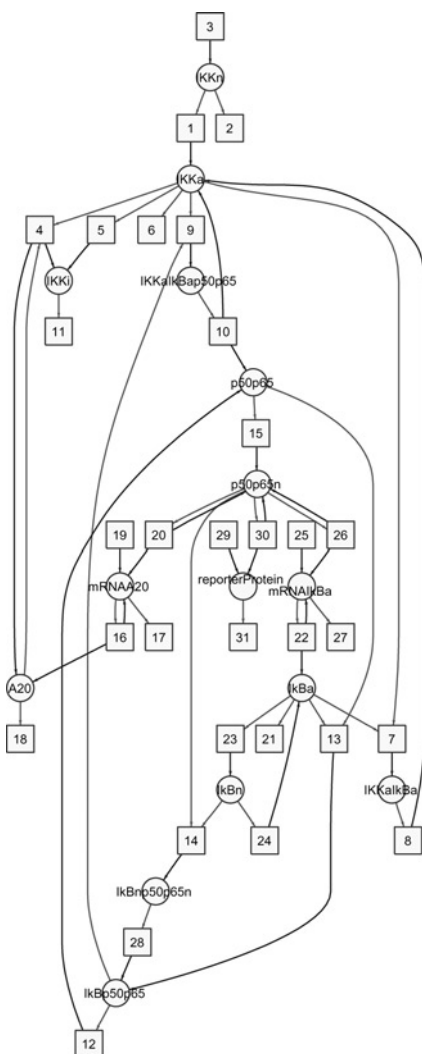


Fig. 2 Model of NF κ B signalling

Non linear reaction mechanism is represented as bipartite graph. There are 15 chemical species and 31 reactions

modelled by a non-zero activation constant $k_1 > 0$, meaning that the kinase is activated.

We have numerically studied the dependence of these time scales on the parameters of the model, which are the kinetic constants of the reactions. The damping time τ_d and the largest relaxation time τ_{\max} were computed by linearising the dynamical equations at steady state. The period of the oscillation has a rigorous meaning only for a limit cycle, when the oscillations are sustained. At a Hopf bifurcation and close to it, the inverted imaginary part of the conjugated eigenvalues crossing the imaginary axis provide good estimate for the period. Another method for computing the period is the direct determination of the timing between successive peaks. We have noticed that in logarithmic scale (throughout this paper, we use natural logarithm), the differences between the periods computed by the two methods were small, therefore we have decided to use the first method, which is more rapid. A criterion for the existence (observability) of the oscillations is the damping time to period ratio. This ratio is infinite for self-sustained oscillations, big for observable oscillations (when at least two peaks are visible). A low ratio means over-damped oscillations. We call the period an observable one, if the above ratio is larger than one.

4.2 *r*-robustness of the period

First, we have tested the 1-robustness of the characteristic times. Each parameter has been multiplied by a variable, positive scale factor (changing from 0.001 to 1000), all the other parameters being kept fixed. The result can be seen in Fig. 3.

Large plateaus over which characteristic times are practically constant correspond to robustness. The period of the oscillations is particularly robust. For the damping time

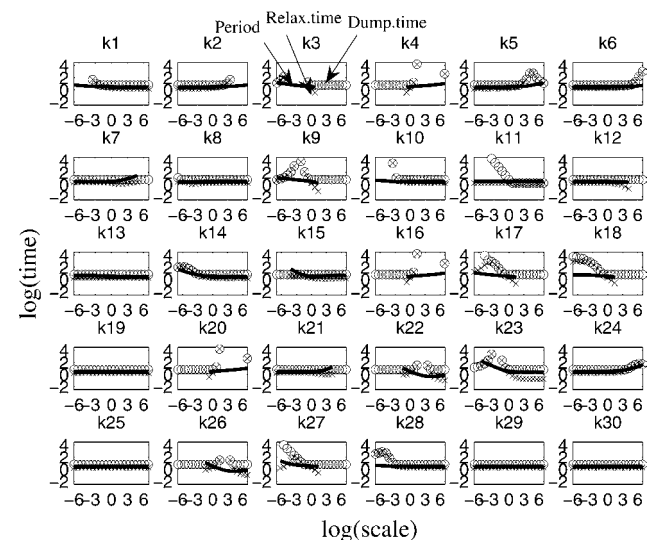


Fig. 3 Log–log dependence of the characteristic times (circles: largest relaxation time, x marks: dumping time, solid line: period) on the scale factor that multiplies the value of one parameter, while all the other parameters are fixed

Scale factor varies from 0.001 to 1000 (from -6.9 to 6.9 in logarithm). Oscillations have limited existence regions (outside these regions they are overdamped; our subjective criterion for overdamping is a damping time over period ratio < 1.75). There are also regions where oscillations are self-sustained. The limits of these regions are Hopf bifurcation points, where the damping time and the largest relaxation time diverge. Inside these regions, the damping time and the largest relaxation time are infinite and not represented.

and the largest relaxation time, we have domains of substantial variation. There are two types of such domains:

1. domains where $d \log \tau / d \log k \simeq -1$.
2. domains where $|d \log \tau / d \log k| > 1$

where k is the variable parameter.

The first type of behaviour is the same as the one of linear networks. For linear networks, when one acts on the ergodicity boundary $k_{(r)}$, the longest relaxation time changes inversely proportional to k (this corresponds to $k = k_{(r)}$). When parameters change, they in turn become the ergodicity boundary. Acting on a parameter, which is not the ergodicity boundary, has no effect; this means a plateau in the graph.

The second type of behaviour exists only for nonlinear networks and is related to bifurcations. The variation of one parameter can bring the system close to a bifurcation (for the NF κ B model, this is a Hopf bifurcation) where the relaxation time diverges.

The in silico experiment shows that the largest relaxation time is not 1-robust; this time can be significantly changed by modifying a single parameter, for instance k_9 . The damping time has similar behaviour being even less robust (some plateaus of the largest relaxation time are higher than the damping time, which continues to decrease; consider for instance the effect of k_9 in Fig. 3).

As also noticed by the biologists [25], the period of the oscillations is 1-robust. We do not have a rigorous explanation of this property. An heuristic explanation is the following. Close to the Hopf bifurcation, two conjugated eigenvalues $\lambda \pm i\mu$ of the Jacobian cross the imaginary axis of the complex plane; λ vanishes that explains the divergence of the relaxation time, while μ , whose inverse is the period, does not change much. However, this is not a full explanation because it does not say what happens far from the Hopf bifurcation point.

4.3 Parameter sensitivity

Not all the parameters have the same influence on the characteristic times. This can already be seen in Fig. 3. To quantify these differences, we have computed the distributions of the characteristic times when one parameter is multiplied by a log-uniform random scale, all the other parameters being fixed. This computation, whose results are represented in Fig. 4, is also a first step towards testing weak r -robustness.

Although rather robust, the period is not constant. Several parameters induce relatively significant changes of this quantity. In the order of increasing strength of their effect on the period, these parameters are: k_7 , k_9 , k_{15} , k_{23} , k_{22} and k_{26} . Among these, k_{22} , k_{26} , and k_9 expressing the transcription rates of mRNA-I κ B, the translation rates of I κ B and the binding rate of the kinase to the NF κ B-I κ B complex are particularly interesting because by changing them, one can increase and also decrease the period. These results confirm and complete the findings of [26]. The parameters that have the greatest influence on the period are the kinetic constants of the production module of I κ B: k_{22} and k_{26} . The strong influence of NF κ B translocation constant k_{15} on the period, missed in [26], is present here. Interestingly, the delay produced in the transcription/translation module of A20 have smaller effect on the period than the delay produced by the I κ B production module. Less obvious is the effect of k_7 and k_9 (binding of IKK to I κ B or to the complex) on the period, detected as important both here and in [26].

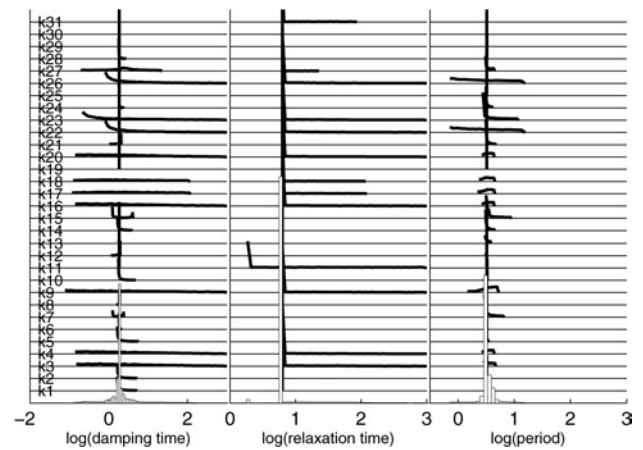


Fig. 4 Parameter sensitivity study; distributions of the characteristic times when different parameters are, in turn, multiplied by a log-uniform (between 0.1 and 10), random scale factor, while all the other parameters are fixed

Distributions corresponding to various parameters are spread out vertically. The lower-most, bar-plotted distribution is the average of all the distributions and corresponds to choosing randomly the parameter to be modified. 1-robustness means that all distributions are concentrated (their spread in log-scale is small). Weak 1-robustness means that only the average distribution is concentrated

The damping time to period ratio represents a criterion for observability of the oscillations. To increase the number of visible peaks, one should increase the above ratio. Because the period is robust, this is equivalent to increasing the damping time. Figs. 3 and 4 show that this is possible in many ways by changing only one parameter (decrease in k_3 , k_9 , k_{17} , k_{18} , k_{23} and k_{27} or increase in k_4 , k_{16} , k_{20} , k_{22} and k_{26}).

4.4 Weak r -robustness of all the characteristic times

The divergence of the relaxation time close to a bifurcation does not necessarily imply the absence of weak r -robustness or of distributed robustness. The set of bifurcation points forms a manifold in the space of parameters, of codimension equal to the codimension of the bifurcation; in general, this set has zero measure (stochastic cellular automata provide an interesting counter-example: the NEC automaton of Andrei Toom [27]). The probability of being by chance close to a bifurcation is generally small.

We have tested the weak r -robustness of the characteristic times, by using independent, log-uniform distributions of the parameters over 2 decades interval. All the three characteristic times are weakly r -robust when r is small (see Figs. 4 and 5a and b). Thus, although controllable (there are critical parameters), the system is weakly robust. Only a directed choice of the right targets has an effect, random choice of a small number of targets is inefficient.

For further study of the r -robustness, we have plotted in Fig. 6 the dependence of the log-variance of the characteristic times on the number of the perturbed parameters r ($1 \leq r \leq n$).

The dependence of the variance of the characteristic times on the number r of perturbed parameters can easily be predicted for a linear network. Let us present simple estimates for only one critical parameter, the ergodicity boundary. Suppose that perturbation of parameters is sufficiently small, $\text{Var}(\log k_i) = \text{Var}(\log k) < \delta^2$ (see Section 3 for the definitions and notations). If the chosen target is the ergodicity boundary, then for the log-variance of the

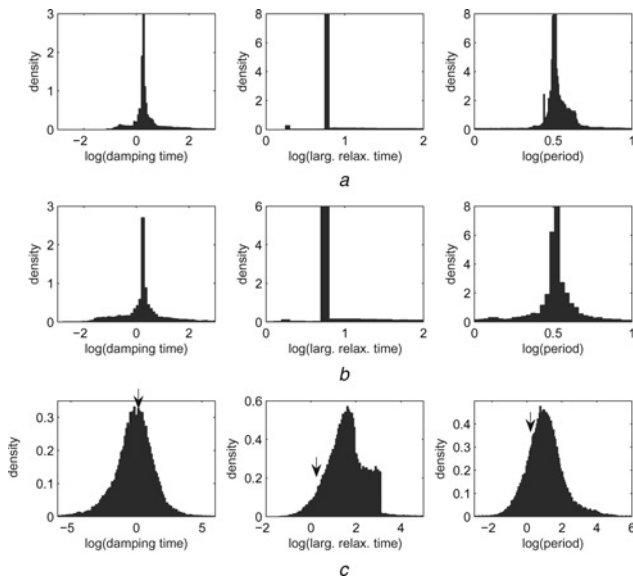


Fig. 5 Distributions of characteristic times for log-uniform (between 0.1 and 10), independent random scales multiplying the kinetic parameters

a one parameter, randomly chosen
b changes in two parameters, randomly chosen
c all the parameters

Unperturbed values of the characteristic times are indicated with arrows. The concentration of the distributions at *a* and *b* shows that the period and the relaxation time are weakly 1- and 2-robust. The variation in all parameters produce long tailed distributions (that can be fitted by log-generalised logistic distributions) of the period and of the damping time, slightly biased relative to the unperturbed values (the bias of the period is positive, suggesting that it is easier to increase, than to decrease the period by random perturbations). The distribution of the relaxation time can be described as a mixture of a log-generalised logistic, and of a log-beta distribution. Let us remind that order statistics for log-uniform, independent variables follow log-beta distributions

relaxation time τ we have $\text{Var}(\log \tau) \sim \text{Var}(\log k)$. The probability to pick the ergodicity boundary is $1 - (1 - 1/n)^r \simeq 1 - \exp(-r/n)$ (for sufficiently big r), so $\text{Var}(\log \tau)/\text{Var}(\log k_i) \simeq 1 - (1 - 1/n)^r \simeq 1 - \exp(-r/n)$. This result can be extended to the case when one has r_0 critical targets. In this case $\text{Var}(\log \tau)/\text{Var}(\log k_i) \simeq C^2[1 - (1 - r_0/n)^r] \simeq C^2[1 - \exp(-rr_0/n)]$, where $C > 0$ is a sensitivity. In our case, we know the number of critical targets from the sensitivity studies $r_0 \simeq 10$ (see Fig. 3). The theoretical curve with $C = 1$, $r_0 = 10$ fits well with the calculated log-variance of the damping time for small values of r , see Fig. 6a). There are differences at larger r that should be explained by the nonlinear interference between the variations of the parameters. For quantities that follow cube concentration the log-variance is just proportional to r ; it is the case of the period, see Fig. 6a). To conclude, the plot of $\text{Var}(\log \tau)$ against r can be used to distinguish between cube concentration and presence of critical targets, and in the latter case to estimate the number of critical targets.

We have also used a protocol for testing distributed robustness. This corresponds to changing all the parameters ($r = n = 31$ in Fig. 6b). Distributed robustness protocol can be used to distinguishing between cube concentration, simplex concentration and the cases with slightly interfering critical targets. It is then useful to plot the log-variance of the characteristic time against the log-variance of the parameters. In the case of cube concentration, one just has the proportionality. For simplex concentration, the discussion from Section 3 applies. There are three regimes: first, proportionality for log-variances up to δ^2 , then saturation for log-variances up to Δ^2 and again proportionality with

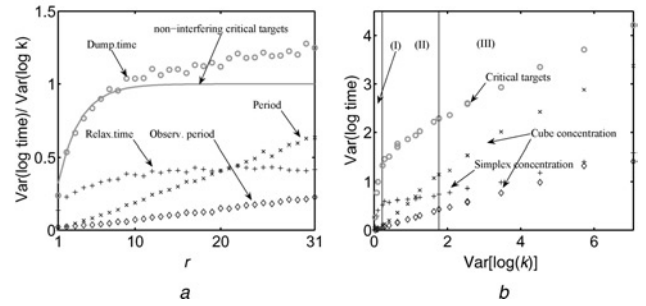


Fig. 6 Relaxation times of nonlinear regimes

a Log-variance of the characteristic times against r , the number of perturbed parameters. The choice of the r parameters is random (uniform) and the values of the random scales are independent, log-uniform (between 0.1 and 10). Some statistical samples correspond to overdamped oscillations (damping time/period ratio < 1); these samples were rejected when computing the log-variance of the observable period. The log-variance of the dumping time is compared with the theoretical curve for $r_0 = 10$ non-interfering critical targets
b Log-variance of the characteristic times against the log-variance of the parameters, for $r = 31$. Relaxation time shows typically simplex concentration behaviour, with a saturation regime (II) between two proportionality regimes (I) and (III)

a smaller slope. The first regime applies with no modifications to the case with critical targets, but if there is no interference between targets, no saturation is observed. Fig. 6b suggests that the behaviours of the relaxation time, of the period and of the dumping time are examples of simplex concentration, of cube concentration and of weakly interfering critical targets, respectively.

We may also want to know the distributions of the characteristic times for a distributed robustness protocol. When all the parameters take independent log-uniform values, the distributions of characteristic times are much broader than the ones induced by changing a small number of parameters (compare Fig. 5a with c). Neither the longest relaxation time nor the damping time has distributed robustness (quantitatively, this follows from Fig. 6a: for $r = 31$ the variance ratios are larger than one). However, Figs. 6a and 5 clearly show that the period is more robust than the other characteristic times. In logarithmic scale, the distributions of the dumping time and of the period have tails with different exponential decay rates towards ∞ and $-\infty$. These distributions (a possible fit is by log-generalized logistic distributions) have longer tails in log scale (exponential, compared to gaussian) than log-normal distributions that are sometimes observed in biology [28–32]. The tails are also longer than the ones of the Tracy–Widom distribution characterising largest eigenvalues of certain classes of random matrices [33, 34]. These long tails are related to the critical retardation phenomena [35] close to the Hopf bifurcation (see also Fig. 3). The distribution of the relaxation time can be seen as a mixture between a log-beta (sharply limited by a maximal time) and a log-generalised logistic distribution (accounting for critical retardation).

5 Discussion and conclusions

We demonstrated the possibility of a new kind of robustness of biological systems. This type of robustness has geometrical origin, being related to the high dimension in which variability sources act. There are two basic types of such geometrical effects: cube-type and simplex-type concentrations.

The classical example of the cube concentration gives the central limit theorem, when the robust property is the sum of

many (n), independent contributions. For concentration of this type, the relative standard deviation decreases as $1/\sqrt{n}$. The classical example of the simplex concentration, is the situation when the robust property depends on the k th order effect (parameter) in a collection of many (n) effects (parameters), for example, the relaxation time of a system with limiting step. For concentration of this type, the relative standard deviation decreases much faster, as $1/n$.

We have also defined the concepts of distributed robustness, r -robustness that occur naturally in molecular biology. We have introduced a new notion: weak r -robustness means that the system is robust with respect to blind attacks (the targets are randomly chosen).

Both distributed and r -robustness imply low sensitivity. Thus, sensitivity studies can be useful for the analysis of robustness, but this may be not enough for proving robustness. Indeed, changing many parameters could have an effect even when there are no critical parameters (parameters with respect to which sensitivity is high). Conversely, it may be sometimes difficult to distinguish between a system with critical parameters and a system with limiting steps (simplex concentration). We showed that the log-variance of the output of the system should have a saturation plateau in the first case and not in the latter, as a function of the log-variance of the parameters. For the nonlinear model of NF κ B signalling, we have distinguished among three types of phenomena: cube concentration for the period, simplex concentration for the relaxation time and critical parameters for the damping time. We have also shown that weak r -robustness protocols can be used to identify the number of critical parameters, when these exist.

For linear networks, we relate the largest relaxation time to the ergodicity boundary (a topological concept). The notion of ergodicity boundary could not be applied directly to nonlinear systems. Nevertheless, direct computation demonstrates that a nonlinear signalling network also has robust relaxation characteristics, and concentration effects for relaxation time seems similar to linear systems (with some additional long-tail effect related to critical retardation).

In our discussion of robust design of linear networks (Section 3), we considered two types of noise: random noise in constants and destruction of links. The necessity of robustness to both types leads to a new combinatorial problem. How to create a minimal network that has sufficiently long routes (the length of each route is $>L$) and, at the same time, sufficiently many routes; after destruction of D_{links} links and D_{nodes} nodes, there remains at least one long route in the network.

In a recent work, Rand *et al.* [36] introduces the flexibility dimension that quantifies the range of evolution of clocks. This notion applies to multitask evolution, simultaneously fulfilling several objectives. By using linear response theory, the authors propose a method to compute the directions in the characteristic space that are not robust to changes of the parameters: the flexibility dimension is the largest linear space of characteristics that contains non-robust directions. Our notion of robustness index is different, because it does not follow from linear response and more importantly it applies to parameters and not to characteristics. We can explain the sense of robustness index r as follows: for significant change of characteristics by random perturbation, one needs to perturb $>r$ parameters. Nevertheless, the flexibility dimension and the robustness index have properties in common: they are both small for simple networks and tend to be increased by the loop complexity and by the unevenness of the lifetimes of various species.

Concerning the analysed example, several conclusions are important. NF κ B dynamics belong to the category of ultradian oscillators. As for circadian oscillators [36], the period of the oscillations is a relatively robust property. Even if the biological role of these oscillations has not yet been proved (for some conjectures the reader can refer to [25]), it is important to know that the robustness applies to different timescales. A specificity of the NF κ B system is the proximity to a Hopf bifurcation. Two nonlinear phenomena could be relevant for the behaviour of the signalling system: the critical retardation and the excitability. The first property would produce long-tail distributions of the damping time of the oscillations. Thus, there are critical parameters for the damping time, which is less robust than the period of the oscillations. The second property could raise the efficiency of the regulatory role of NF κ B by increasing the amplitude of its response to signals.

The robustness of a system could be related to its complexity. To test the concentration rigorously, from high-dimension, one needs to build an hierarchy of models obtained from another model by model reduction. Parameters of simpler models in the hierarchy are functions of packages of parameters ('atoms') of more complex models. Independent perturbations of the atoms produce less variability than overall perturbation of packages. Another source of complexity is dynamics itself. It is necessary to take into account dynamical complexity as well as complexity of hierarchical organisation. These ideas have been briefly discussed in [37] and will be presented in detail in a future work.

6 References

- 1 Waddington, C.H.: 'The strategy of genes' (Allen and Unwin, London, 1957)
- 2 Thom, R.: 'Structural stability and morphogenesis' (Benjamin, New York, 1975)
- 3 Fates, N.: 'Robustesse de la dynamique des systèmes discrets: le cas de l'asynchronisme dans les automates cellulaires', PhD Dissertation, ENS Lyon, 2004
- 4 Chaves, J.M., Albert, R., and Sontag, E.D.: 'Robustness and fragility of Boolean models for genetic regulatory', *J. Theor. Biol.*, 2005, **235**, pp. 431–449
- 5 Morohashi, M., Winn, A., Borisuk, M.T., Bolouri, H., Doyle, J., and Kitano, H.: 'Robustness as a measure of plausibility in models of biochemical networks', *J. Theor. Biol.*, 2002, **216**, pp. 19–30
- 6 Kitano, H., Oda, K., Kimura, T., Matsuoka, Y., Csete, M., and Doyle, J., *et al.*: 'Metabolic syndrome and robustness tradeoffs', *Diabetes*, 2004, **53**, S6–S15
- 7 Kitano, H.: 'Biological robustness', *Nat. Rev.*, 2004, **5**, pp. 826–837
- 8 von Neumann, J.: 'Probabilistic logics and the synthesis of reliable organisms from unreliable components', von Neumann, J. Collected Works' (Pergamon Press, Oxford, 1963, vol. 5)
- 9 Wagner, A.: 'Robustness and evolvability in living systems'. (Oxford, Princeton University Press, Princeton, 2005)
- 10 Wagner, A.: 'Distributed robustness versus redundancy as causes of mutational robustness', *BioEssays*, 2005, **27**, pp. 176–188
- 11 Deutscher, D., Meilijson, I., Kupiec, M., and Ruppin, E.: 'Multiple knockout analysis of genetic robustness in the yeast metabolic network', *Nat. Genetics*, 2006, **9**, pp. 993–998
- 12 Houchmanzadeh, B., Wieschaus, E., and Leibler, S.: 'Establishment of developmental precision and proportions in the early Drosophila embryo', *Nature*, 2002, **415**, pp. 798–802
- 13 Watkins, H., and Farrall, M.: 'Genetic susceptibility to coronary artery disease: from promise to progress', *Nat. Rev. Genetics*, 2006, **7**, pp. 163–173
- 14 Ptashne, M., and Gann, A.: 'Genes and Signals' (CSHL Press, Cold Spring Harbor, 2002)
- 15 Hornstein, E., and Shomron, N.: 'Canalization of development by microRNAs', *Nat. Genetics*, 2006, **38**, S20–S24
- 16 von Dassow, G., Meir, E., Munro, E.M., and Odell, G.M.: 'The segment polarity network is a robust developmental module', *Nature*, 2000, **406**, pp. 188–192
- 17 Lehmann, E.L.: 'Nonparametrics' (Holden-Day, San Francisco, 1975)

- 18 Talagrand, M.: 'Concentration of measure and isoperimetric inequalities in product spaces', *Ins. Hautes. Etudes. Sci. Publ. Math.*, 1995, **81**, pp. 73–205
- 19 Gromov, M.: 'Metric structures for Riemannian and non-Riemannian spaces, Progr. Math. 152' (Birkhauser, Boston, 1999)
- 20 Gorban, A.N., Bykov, V.I., and Yablonskii, G.S.: 'Essays on chemical relaxation' (Nauka, Novosibirsk, 1986)
- 21 Iosifescu, M.: 'Finite Markov processes and their applications' (Wiley and Sons, New York, 1980)
- 22 Gorban, A.N., and Radulescu, O.: 'Dynamic and static limitation in reaction networks', revisited, arXiv:physics/0703278
- 23 Hoffmann, A., Levchenko, A., Scott, M.L., and Baltimore, D.: 'The I κ B–NF– κ B signalling module: temporal control and selective gene activation', *Science*, 2002, **298**, pp. 1241–1245
- 24 Lipniacki, T., Paszek, P., Brasier, A.R., Luxon, B., and Kimmel, M.: 'Mathematical model of NF– κ B regulatory module', *J. Theor. Biol.*, 2004, **228**, pp. 195–215
- 25 Nelson, D.E., Ihekweba, A.E.C., Elliot, M., Johnson, J.R., Gibney, C.A., Foreman, B.E., *et al.*: 'Oscillations in NF– κ B signalling control the dynamics of gene expression', *Science*, 2004, **306**, pp. 704–708
- 26 Ihekweba, A.E.C., Broomhead, D.S., Grimley, R.L., Benson, N., and Kell, D.B.: 'Sensitivity analysis of parameters controlling oscillatory signaling in the NF– κ B pathway: the roles of IKK and I κ B α ', *Syst. Biol.*, 2004, **1**, pp. 93–102
- 27 Grinstein, G.: 'Can complex structures be generically stable in a noisy world?', *IBM J. Res. Dev.*, 2004, **48**, pp. 5–12
- 28 Liu, D., Umbach, D.M., Pedada, S.D., Li, L., Crockett, P.W., and Weinberg, C.R.: 'A random-periods model for expression of cell-cycle genes', *Proc. Natl. Acad. Sci., USA*, 2004, **101**, pp. 7240–7245
- 29 Konishi, T.: 'Three-parameter lognormal distribution ubiquitously found in cDNA microarray data and its application to parametric data treatment', *BMC Bioinform.*, 2004, **5**
- 30 Begtson, M., Stahlberg, A., Rorsman, P., and Kubista, M.: 'Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels', *Genome Res.*, 2005, **15**, pp. 1388–1392
- 31 Furusawa, C., Suzuki, T., Kashiwagi, A., Yomo, T., and Kaneko, K.: 'Ubiquity of log-normal distributions in intra-cellular reaction dynamics', *Biophysics*, 2005, **1**, pp. 25–31
- 32 Limpert, E., Stahel, W.A., and Abbt, M.: 'Log-normal distributions across the sciences: keys and clues', *BioScience*, 2001, **51**, pp. 341–352
- 33 Tracy, C.A., and Widom, H.: 'On orthogonal and symplectic matrix ensembles', *Commun. Math. Phys.*, 1996, **177**, pp. 727–754
- 34 Soshnikov, A.: 'A note on universality of the distribution of the largest eigenvalues', *J. Stat. Phys.*, 2002, **108**, pp. 1033–1056
- 35 Gorban, A.N.: 'Singularities of transition processes in dynamical systems', *Electronic J. Differential Equations*, 2004; Monograph 05
- 36 Rand, D.A., Shulgin, B.V., Salazar, J.D., and Millar, A.J.: 'Uncovering the design principles of circadian clocks: mathematical analysis of flexibility and evolutionary goals', *J. Theor. Biol.*, 2006, **238**, pp. 616–635
- 37 Radulescu, O., Gorban, A., Vakulenko, S., and Zinovyev, A.: 'Hierarchies and modules in complex biological systems'. Proc. ECCS'06, 2006

MATHEMATICAL BIOLOGY AND ECOLOGY LECTURE NOTES



DR RUTH E. BAKER
MICHAELMAS TERM 2011

Contents

1	Introduction	5
1.1	References	6
2	Spatially independent models for a single species	7
2.1	Continuous population models for single species	7
2.1.1	Investigating the dynamics	8
2.1.2	Linearising about a stationary point	11
2.1.3	Insect outbreak model	12
2.1.4	Harvesting a single natural population	15
2.2	Discrete population models for a single species	18
2.2.1	Linear stability	20
2.2.2	Further investigation	20
2.2.3	The wider context	25
3	Continuous population models: interacting species	27
3.1	Predator-prey models	27
3.1.1	Finite predation	29
3.2	A look at global behaviour	30
3.2.1	Nullclines	31
3.2.2	The Poincaré-Bendixson Theorem	31
3.3	Competitive exclusion	32
3.4	Mutualism (symbiosis)	35
3.5	Interacting discrete models	35
4	Enzyme kinetics	36
4.1	The Law of Mass Action	36
4.2	Michaelis-Menten kinetics	37
4.2.1	Non-dimensionalisation	38
4.2.2	Singular perturbation investigation	38
4.3	More complex systems	40
4.3.1	Several enzyme reactions and the pseudo-steady state hypothesis . .	40
4.3.2	Allosteric enzymes	41
4.3.3	Autocatalysis and activator-inhibitor systems	41

5	Introduction to spatial variation	43
5.1	Derivation of the reaction-diffusion equations	44
5.2	Chemotaxis	46
6	Travelling waves	48
6.1	Fisher's equation: an investigation	48
6.1.1	Key points	48
6.1.2	Existence and the phase plane	50
6.1.3	Relation between the travelling wave speed and initial conditions . .	53
6.2	Models of epidemics	54
6.2.1	The SIR model	55
6.2.2	An SIR model with spatial heterogeneity	56
7	Pattern formation	59
7.1	Minimum domains for spatial structure	59
7.1.1	Domain size	60
7.2	Diffusion-driven instability	61
7.2.1	Linear analysis	62
7.3	Detailed study of the conditions for a Turing instability	65
7.3.1	Stability without diffusion	65
7.3.2	Instability with diffusion	66
7.3.3	Summary	67
7.3.4	The threshold of a Turing instability.	68
7.4	Extended example 1	68
7.4.1	The influence of domain size	69
7.5	Extended example 2	69
8	Excitable systems: nerve pulses	72
8.1	Background	72
8.1.1	Resistance	73
8.1.2	Capacitance	73
8.2	Deducing the Fitzhugh Nagumo equations	74
8.2.1	Space-clamped axon	74
8.3	A brief look at the Fitzhugh Nagumo equations	76
8.3.1	The (n, v) phase plane	76
8.4	Modelling the propagation of nerve signals	78
8.4.1	The cable model	78
A	The phase plane	81
A.1	Properties of the phase plane portrait	82
A.2	Equilibrium points	82
A.2.1	Equilibrium points: further properties	83
A.3	Summary	84
A.4	Investigating solutions of the linearised equations	84
A.4.1	Case I	85

A.4.2	Case II	87
A.4.3	Case III	87
A.5	Linear stability	90
A.5.1	Technical point	90
A.6	Summary	91

Chapter 1

Introduction

An outline for this course.

- We will observe that many phenomena in ecology, biology and biochemistry can be modelled mathematically.
- We will initially focus on systems where the spatial variation is not present or, at least, not important. Therefore only the temporal evolution needs to be captured in equations and this typically (but not exclusively) leads to difference equations and/or ordinary differential equations.
- We are inevitably confronted with systems of non-linear difference or ordinary differential equations, and thus we will study analytical techniques for extracting information from such equations.
- We will proceed to consider systems where there is explicit spatial variation. Then models of the system must additionally incorporate spatial effects.
- In ecological and biological contexts the main physical phenomenon governing the spatial variation is typically, but again not exclusively, diffusion. Thus we are invariably required to consider parabolic partial differential equations. Mathematical techniques will be developed to study such systems.
- These studies will be in the context of ecological, biological and biochemical applications. In particular we will draw examples from:
 - enzyme dynamics and other biochemical reactions;
 - epidemics;
 - interaction ecological populations, such as predator-prey models;
 - biological pattern formation mechanisms;
 - chemotaxis;
 - the propagation of an advantageous gene through a population;
 - nerve pulses and their propagation.

1.1 References

The main references for this lecture course will be:

- J. D. Murray, Mathematical Biology, 3rd edition, Volume I [8].
- J. D. Murray, Mathematical Biology, 3rd edition, Volume II [9].

Other useful references include (but are no means compulsory):

- J. P. Keener and J. Sneyd, Mathematical Physiology [7].
- L. Edelstein-Keshet, Mathematical Models in Biology [2].
- N. F. Britton, Essential Mathematical Biology [1].

Chapter 2

Spatially independent models for a single species

In this chapter we consider modelling a single species in cases where spatial variation is not present or is not important. In this case we can simply examine the temporal evolution of the system.

References.

- J. D. Murray, Mathematical Biology, 3rd edition, Volume I, Chapter 1 and Chapter 2 [8].
- L. Edelstein-Keshet, Mathematical Models in Biology, Chapter 1, Chapter 2 and Chapter 6 [2].
- N. F. Britton, Essential Mathematical Biology, Chapter 1 [1].

2.1 Continuous population models for single species

A core feature of population dynamics models is the conservation of population number, *i.e.*

$$\begin{aligned} \text{rate of increase of population} &= \text{birth rate} - \text{death rate} \\ &+ \text{rate of immigration} - \text{rate of emigration.} \end{aligned} \tag{2.1}$$

We will make the assumption the system is closed and thus there is no immigration or emigration.

Let $N(t)$ denote the population at time t . Equation (2.1) becomes

$$\frac{dN}{dt} = f(N) = Ng(N), \tag{2.2}$$

where $g(N)$ is defined to be the intrinsic growth rate. Examples include:

The Malthus model. This model can be written as:

$$g(N) = b - d \stackrel{\text{def}}{=} r, \quad (2.3)$$

where b and d are constant birth and death rates. Thus

$$\frac{dN}{dt} = rN, \quad (2.4)$$

and hence

$$N(t) = N_0 e^{rt}. \quad (2.5)$$

The Verhulst model. This model is also known as the logistic growth model:

$$g(N) = r \left(1 - \frac{N}{K} \right). \quad (2.6)$$

Definition. In the logistic growth equation, r is defined to be the *linear birth rate* and K is defined to be the *carrying capacity*.

For $N \ll K$, we have

$$\frac{dN}{dt} \simeq rN \Rightarrow N \simeq N_0 e^{rt}. \quad (2.7)$$

However, as N tends towards K ,

$$\frac{dN}{dt} \rightarrow 0, \quad (2.8)$$

the growth rate tends to zero.

We have

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right), \quad (2.9)$$

and hence

$$N(t) = \frac{N_0 K e^{rt}}{K + N_0 (e^{rt} - 1)} \rightarrow K \quad \text{as } t \rightarrow \infty. \quad (2.10)$$

Sketching $N(t)$ against time yields solution as plotted in Figure 2.1: we see that solutions always monotonically relaxes to K as $t \rightarrow \infty$.

Aside. The logistic growth model has been observed to give very good fits to population data in numerous, disparate, scenarios ranging from bacteria and yeast to rats and sheep [8].

2.1.1 Investigating the dynamics

There are two techniques we can use to investigate the model

$$\frac{dN}{dt} = f(N) = Ng(N). \quad (2.11)$$

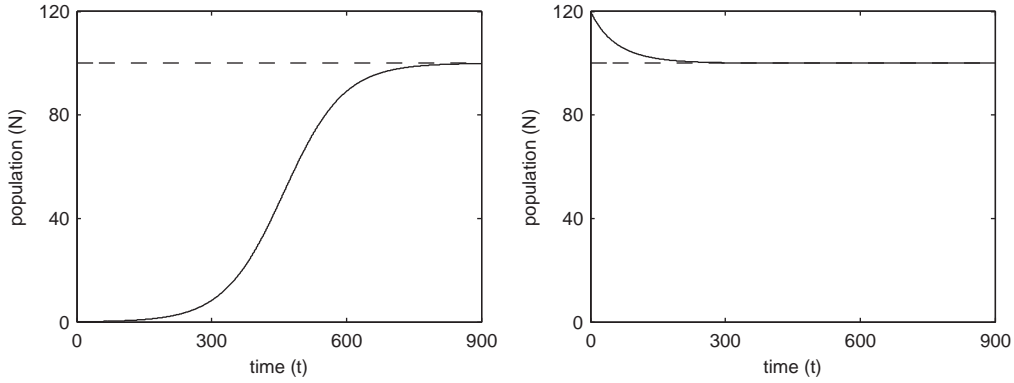


Figure 2.1: Logistic growth for $N_0 < K$ (left-hand) and $N_0 > K$ (right-hand). Parameters are as follows: $r = 0.015$ and $K = 100$.

Method (i): analytical solution

For the initial conditions $N(t = 0) = N_0$, with N_0 fixed, we can formally integrate equation (2.2) to give $N(t) = N^*(t)$, where $N^*(\cdot)$ is the inverse of the function $F(\cdot)$ defined by

$$F(x) = \int_{N_0}^x \frac{1}{f(s)} ds. \quad (2.12)$$

However, unless integrating and finding the inverse function is straightforward, there is an easier way to determine the dynamics of the system.

Method (ii): plot the graph

Plot $dN/dt = f(N) = Ng(N)$ as a function of N . For example, with

$$f(N) = Ng(N) = N(6N^2 - N^3 - 11N + 6) = N(N - 1)(N - 2)(3 - N), \quad (2.13)$$

we have the plot shown in Figure 2.2.

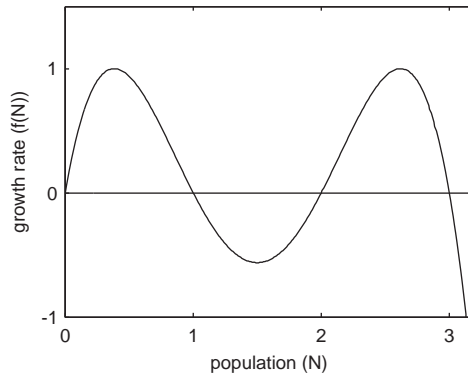


Figure 2.2: Growth according to the dynamics $f(N) = N(N - 1)(N - 2)(3 - N)$.

Note 1. For a given initial condition, N_0 , the system will tend to the nearest root of $f(N) = Ng(N)$ in the direction of $f(N_0)$. The value of $|N(t)|$ will tend to infinity with large time if no such root exists.

For $f(N) = Ng(N) = N(N-1)(N-2)(3-N)$, we have:

- when $N_0 \in (0, 2]$ the large time asymptote is $N(\infty) = 1$;
- for $N_0 > 2$ the large time asymptote is $N(\infty) = 3$;
- $N(t) = 0 \forall t$ if $N(0) = 0$.

Note 2. On more than one occasion we will have a choice between using a graphical method and an analytical method, as seen above. The most appropriate method to use is highly dependent on context. The graphical method, Method (ii), *quickly and simply* gives the large time behaviour of the system and stability information (see below). The analytical method, Method (i), is often significantly more cumbersome, but yields *all* information, at a detailed quantitative level, about the system.

Definition. A *stationary point*, also known as an *equilibrium point*, is a point where the dynamics does not change in time. Thus in our specific context of $dN/dt = f(N) = Ng(N)$, the stationary points are the roots of $f(N) = 0$.

Example. For $dN/dt = f(N) = Ng(N) = N(N-1)(N-2)(3-N)$, the stationary points are

$$N = 0, 1, 2, 3. \quad (2.14)$$

Definition. A stationary point is *stable* if a solution starting sufficiently close to the stationary point remains close to the stationary point.

Non-examinable. A rigorous definition is as follows. Let $N_{N_0}(t)$ denote the solution to $dN/dt = f(N) = Ng(N)$ with initial condition $N(t=0) = N_0$. A stationary point, N_s , is stable if, and only if, for all $\epsilon > 0$ there exists a δ such that if $|N_s - N_0| < \delta$ then $|N_{N_0}(t) - N_s| < \epsilon$.

Exercise. Use Figure 2.2 to deduce which of stationary points of the system

$$\frac{dN}{dt} = f(N) = Ng(N) = N(N-1)(N-2)(3-N), \quad (2.15)$$

are stable.

Solution. Figure 2.2 shows that both $N_s = 1$ and $N_s = 3$ are stable.

2.1.2 Linearising about a stationary point

Suppose N_s is a stationary point of $dN/dt = f(N)$ and make a small perturbation about N_s :

$$N(t) = N_s + n(t), \quad n(t) \ll N_s. \quad (2.16)$$

We have, by using a Taylor expansion of $f(N)$ and denoting $' = d/dN$, that

$$f(N(t)) = f(N_s + n(t)) = f(N_s) + n(t)f'(N_s) + \frac{1}{2}n(t)^2 f''(N_s) + \dots, \quad (2.17)$$

and hence

$$\frac{dn}{dt} = \frac{dN}{dt} = f(N(t)) = f(N_s) + n(t)f'(N_s) + \frac{1}{2}n(t)^2 f''(N_s) + \dots \quad (2.18)$$

The linearisation of $dN/dt = f(N)$ about the stationary point N_s is given by neglecting higher order (and thus smaller) terms to give

$$\frac{dn}{dt} = f'(N_s)n(t).$$

The solution to this *linear system* is simply

$$n(t) = n(t=0) \exp \left[t \frac{df}{dN}(N_s) \right]. \quad (2.19)$$

Definition. Let N_s denote a stationary point of $dN/dt = f(N)$, and let

$$n(t) = n(t=0) \exp \left[t \frac{df}{dN}(N_s) \right], \quad (2.20)$$

be the solution of the linearisation about N_s . Then N_s is *linearly stable* if $n(t) \rightarrow 0$ as $t \rightarrow \infty$. In other words, N_s is linearly stable if

$$\frac{df}{dN}(N_s) < 0. \quad (2.21)$$

Exercise. By algebraic means, deduce which stationary points of the system

$$\frac{dN}{dt} = f(N) = Ng(N) = N(N-1)(N-2)(3-N), \quad (2.22)$$

are linearly stable. Can your answer be deduced graphically?

Solution. Differentiating $f(N)$ with respect to N gives

$$f'(N) = 2 - 22N + 18N^2 - 4N^3, \quad (2.23)$$

and hence $f'(0) = 6$ (unstable), $f'(1) = -8$ (stable) *etc.*

Consider the graph of $f(N)$ to deduce stability graphically—steady states with negative gradient are linearly stable *c.f.* Figure 2.2.

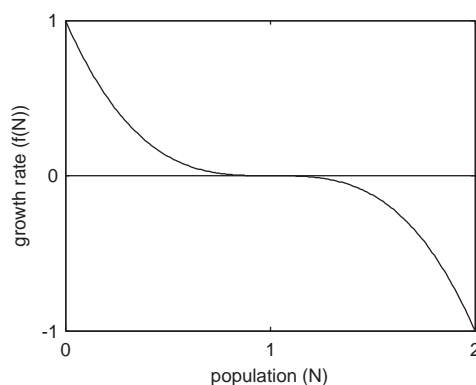


Figure 2.3: Growth according to the dynamics $f(N) = (1 - N)^3$.

Exercise. Find a function $f(N)$ such that $dN/dt = f(N)$ has a stationary point which is stable and *not* linearly stable.

Solution. The function

$$f(N) = (1 - N)^3, \quad (2.24)$$

gives $f'(1) = 0$ and is therefore not linearly stable (see Figure 2.3).

2.1.3 Insect outbreak model

First introduced by Ludwig in 1978, the model supposes budworm population dynamics to be modelled by the following equation:

$$\frac{dN}{dt} = r_B N \left(1 - \frac{N}{K_B} \right) - p(N), \quad p(N) \stackrel{\text{def}}{=} \frac{BN^2}{A^2 + N^2}. \quad (2.25)$$

The function $p(N)$ is taken to represent the effect upon the population of predation by birds. Plotting $p(N)$ as a function of N gives the graph shown in Figure 2.4.

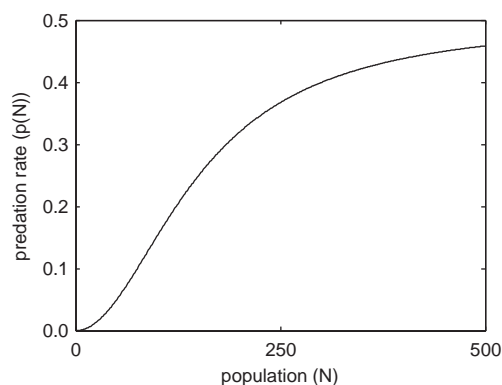


Figure 2.4: Predation, $p(N)$, in the insect outbreak model. Parameters are as follows: $A = 150$, $B = 0.5$.

Non-dimensionsionalisation

Let

$$N = N^*u, \quad t = T\tau, \quad (2.26)$$

where N^* , N have units of biomass, and t , T have units of time, with N^* , T constant. Then

$$\frac{N^*}{T} \frac{du}{d\tau} = r_B N^* u \left(1 - \frac{N^* u}{K_B} \right) - \frac{B(N^*)^2 u^2}{A^2 + (N^*)^2 u^2}, \quad (2.27)$$

$$\Rightarrow \frac{du}{d\tau} = r_B T u \left(1 - \frac{N^* u}{K_B} \right) - \frac{B T N^* u^2}{A^2 + (N^*)^2 u^2}. \quad (2.28)$$

Hence with

$$N^* = A, \quad T = \frac{A}{B}, \quad r = r_B T = \frac{r_B A}{B}, \quad q = \frac{K_B}{N^*} = \frac{K_B}{A}, \quad (2.29)$$

we have

$$\frac{du}{d\tau} = r u \left(1 - \frac{u}{q} \right) - \frac{u^2}{1 + u^2} \stackrel{\text{def}}{=} f(u; r, q). \quad (2.30)$$

Thus we have reduced the number of parameters in our model from four to two, which substantially simplifies our subsequent study.

Steady states

The steady states are given by the solutions of

$$r u \left(1 - \frac{u}{q} \right) - \frac{u^2}{1 + u^2} = 0. \quad (2.31)$$

Clearly $u = 0$ is a steady state. We proceed graphically to consider the other steady states which are given by the intersection of the graphs

$$f_1(u) = r \left(1 - \frac{u}{q} \right) \quad \text{and} \quad f_2(u) = \frac{u}{1 + u^2}. \quad (2.32)$$

The top left plot of Figure 2.5 shows plots of $f_1(u)$ and $f_2(u)$ for different values of r and q . We see that, depending on the values of r and q , we have either one or three non-zero steady states. Noting that

$$\left. \frac{df(u; r, q)}{du} \right|_{u=0} = r > 0, \quad (2.33)$$

typical plots of $du/d\tau$ vs. u are shown in Figure 2.5 for a range of values of r and q .

Definition. A system displaying *hysteresis* exhibits a response to the increase of a driving variable which is not precisely reversed as the driving variable is decreased.

Remark. Hysteresis is remarkably common. Examples include ferromagnetism and elasticity, amongst others. See <http://en.wikipedia.org/wiki/Hysteresis> for more details.

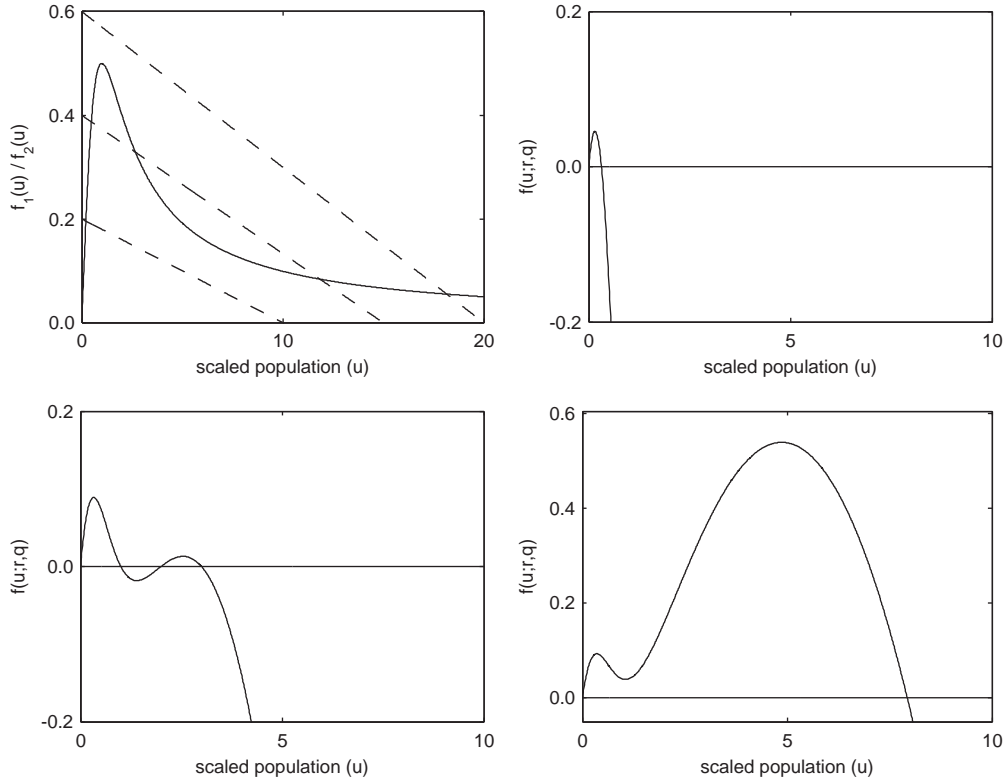


Figure 2.5: Dynamics of the non-dimensional insect outbreak model. Top left: plots of the functions $f_1(u)$ (dashed line) and $f_2(u)$ (solid line) with parameters $r = 0.2, 0.4, 0.6$, $q = 10, 15, 20$, respectively. Top right: plot of $f(u; r, q)$ with parameters $r = 0.6$, $q = 0.6$. Bottom left: plot of $f(u; r, q)$ with parameters $r = 0.6$, $q = 6$. Bottom right: plot of $f(u; r, q)$ with parameters $r = 0.6$, $q = 10$.

Extended Exercise

- Fix $r = 0.6$. Explain how the large time asymptote of u , and hence N , changes as one slowly increases q from $q \ll 1$ to $q \gg 1$ and then one decreases q from $q \gg 1$ to $q \ll 1$. In particular, show that hysteresis is present. Note for this value of r , there are three non-zero stationary points for $q \in (q_1, q_2)$ with $1 < q_1 < q_2 < 10$.

Solution. For small values of q there is only one non-zero steady state, S_1 . As q is increased past q_1 , three non-zero steady states exist, S_1, S_2, S_3 , but the system stays at S_1 . As q is increased further, past q_2 , the upper steady state S_3 is all that remains and hence the system moves to S_3 . If q is now decreased past q_2 , three non-zero steady states (S_1, S_2, S_3) exist but the system remains at S_3 until q is decreased past q_1 .

Figure 2.6 shows $f(u; r, q)$ for different values of q . The dashed line shows a plot for $q = q_1$ whilst the dash-dotted line shows a plot for $q = q_2$.

- What is the biological interpretation of the presence of hysteresis in this model?

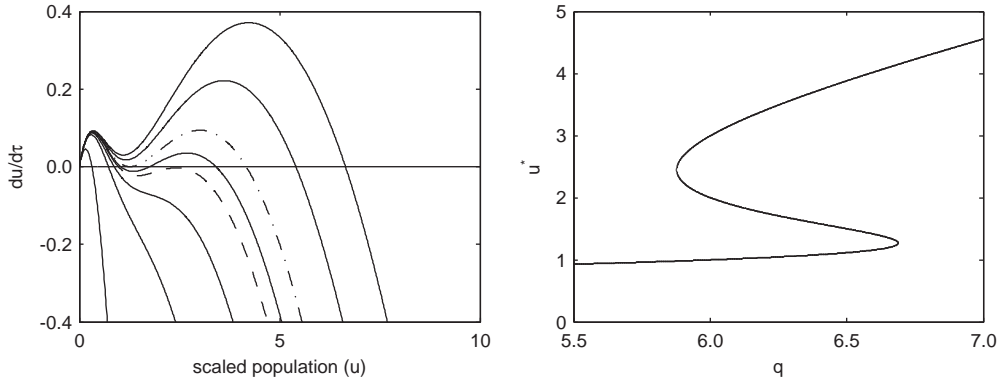


Figure 2.6: Left-hand plot: $du/d\tau = f(u; r, q)$ in the non-dimensional insect outbreak model as q is varied. For small q there is one, small, steady state, for $q \in (q_1, q_2)$ there are three non-zero steady states and for large q there is one, large, steady state. Right-hand plot: the steady states plotted as a function of the parameter q reveals the hysteresis loop.

Solution. If the carrying capacity, q , is accidentally manipulated such that an outbreak occurs ($S_1 \rightarrow S_3$) then reversing this change is not sufficient to reverse the outbreak.

2.1.4 Harvesting a single natural population

We wish to consider a simple model for the maximum sustainable yield. Suppose, in the absence of harvesting, we have

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right). \quad (2.34)$$

We consider a perturbation from the non-zero steady state, $N = K$. Thus we write $N = K + n$, and find, on linearising,

$$\frac{dn}{dt} = -rn \quad \Rightarrow \quad n = n_0 e^{-rt}. \quad (2.35)$$

Hence the system returns to equilibrium on a timescale of $T_R(0) = \mathcal{O}(1/r)$.

We consider two cases for harvesting:

- constant yield, Y ;
- constant effort, E .

Constant yield

For a constant yield, $Y = Y_0$, our equations are

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) - Y_0 \stackrel{def}{=} f(N; Y_0). \quad (2.36)$$

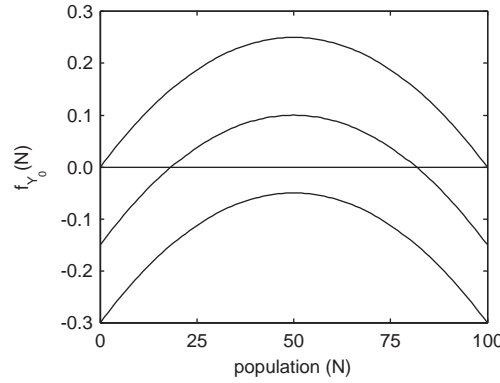


Figure 2.7: Dynamics of the constant yield model for $Y_0 = 0.00, 0.15, 0.30$. As Y_0 is increased beyond a critical value the steady states disappear and $N \rightarrow 0$ in finite time. Parameters are as follows: $K = 100$ and $r = 0.01$.

Plotting dN/dt as a function of N reveals (see Figure 2.7) that the steady states disappear as Y_0 is increased beyond a critical value, and then $N \rightarrow 0$ in finite time.

The steady states are given by the solutions of

$$rN^* - \frac{rN^{*2}}{K} - Y_0 = 0 \quad \Rightarrow \quad N^* = \frac{r \pm \sqrt{r^2 - 4rY_0/K}}{2r/K}. \quad (2.37)$$

Therefore extinction will occur once

$$Y_0 > \frac{rK}{4}. \quad (2.38)$$

Constant effort

For harvesting at constant effort our equations are

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right) - EN \stackrel{\text{def}}{=} f(N; E) = N(r - E) - \frac{rN^2}{K}, \quad (2.39)$$

where the yield is $Y(E) = EN$. The question is: how do we maximise $Y(E)$ such that the stationary state still recovers?

The steady states, N^* , are such that $f(N^*; E) = 0$ (see Figure 2.8). Thus

$$N^*(E) = \frac{(r - E)K}{r} = \left(1 - \frac{E}{r}\right) K, \quad (2.40)$$

and hence

$$Y^*(E) = EN^*(E) = \left(1 - \frac{E}{r}\right) KE. \quad (2.41)$$

Thus the maximum yield, and corresponding value of N^* , are given by the value of E such that

$$\frac{\partial Y^*}{\partial E} = 0 \quad \Rightarrow \quad E = \frac{r}{2}, \quad Y_{max}^* = \frac{rK}{4}, \quad N_{max}^* = \frac{K}{2}. \quad (2.42)$$

Linearising about the stationary state $N^*(E)$ we have $N = N^*(E) + n$ with

$$\frac{dn}{dt} \simeq f_E(N^*) + \left. \frac{df(N; E)}{dN} \right|_{N=N^*} n + \dots = -(r - E)n + \dots, \quad (2.43)$$

and hence the recovery time is given by

$$T_R(E) \simeq \mathcal{O}\left(\frac{1}{r - E}\right). \quad (2.44)$$

Defining the recovery time to be the time for a perturbation to decrease by a factor of e according to the linearised equations about the non-zero steady state, then

$$T_R(0) = \frac{1}{r}, \quad T_R(E) = \frac{1}{r - E}. \quad (2.45)$$

Hence, at the maximum yield state,

$$T_R(E) = \frac{2}{r} \quad \text{since} \quad E = \frac{r}{2} \quad \text{at maximum yield.} \quad (2.46)$$

As we measure Y it is useful to rewrite E in terms of Y to give the ratio of recovery times in terms of the yield $Y(E)$ and the maximum yield Y_M :

$$\frac{T_R(Y)}{T_R(0)} = \frac{2}{1 \pm \sqrt{1 - \frac{Y}{Y_M}}}. \quad (2.47)$$

Derivation. At steady state, we have

$$\frac{K}{r}E^2 - KE + Y^* = 0 \quad \text{as} \quad Y^* = EN^* = KE \left(1 - \frac{E}{r}\right). \quad (2.48)$$

This gives

$$E = \frac{r \pm r\sqrt{1 - 4Y^*/Kr}}{2} \quad \Rightarrow \quad r - E = \frac{r}{2} \left[1 \mp \sqrt{1 - \frac{Y^*}{Y_M^*}}\right]. \quad (2.49)$$

Substituting into equation (2.45) gives the required result.

Plotting $T_R(Y)/T_R(0)$ as a function of Y/Y_M yields some interesting observations, as shown in Figure 2.8.

Note. As T_R increases the population recovers less quickly, and therefore spends more time away from the steady state, N^* . The biological implication is that, in order to maintain a constant yield, E must be increased. This, in turn, implies T_R increases, resulting in a positive feedback loop that can have disastrous consequences upon the population.

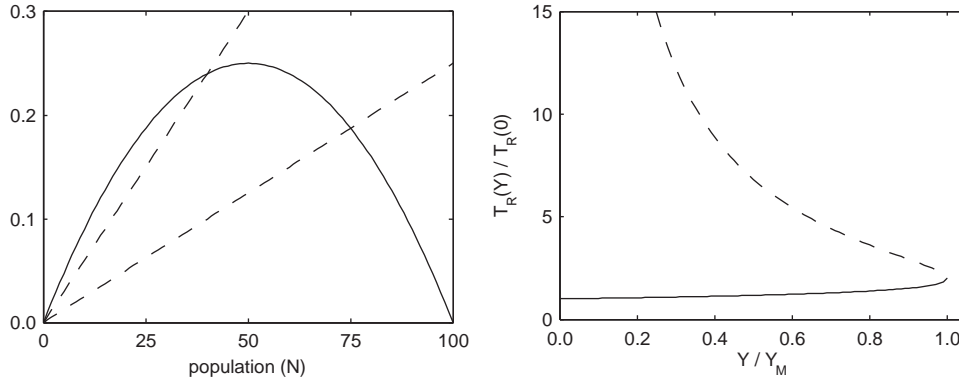


Figure 2.8: Dynamics of the constant effort model. The left-hand plot shows the logistic growth curve (solid line) and the yield, $Y = EN$ (dashed lines), for two values of E . The right-hand plot shows the ratio of recovery times, $T_R(Y)/T_R(0)$, with the negative root plotted as a dashed line and the positive root as a solid line. Parameters are as follows: $K = 100$ and $r = 0.01$.

2.2 Discrete population models for a single species

When there is no overlap in population numbers between each generation, we have a discrete model:

$$N_{t+1} = N_t f(N_t) = H(N_t). \quad (2.50)$$

A simple example is

$$N_{t+1} = r N_t, \quad (2.51)$$

which implies

$$N_t = r^t N_0 \rightarrow \begin{cases} \infty & r > 1 \\ N_0 & r = 1 \\ 0 & r < 1 \end{cases}. \quad (2.52)$$

Definition. An *equilibrium point*, N^* , for a discrete population model satisfies

$$N^* = N^* f(N^*) = H(N^*). \quad (2.53)$$

Such a point is often known as a *fixed point*.

An extension of the simple model, equation (2.51), called the Ricker model includes a reduction of the growth rate for large N_t :

$$N_{t+1} = N_t \exp \left[r \left(1 - \frac{N_t}{K} \right) \right], \quad r > 0 \quad K > 0, \quad (2.54)$$

or, in non-dimensionalised form,

$$u_{t+1} = u_t \exp [r (1 - u_t)] \stackrel{\text{def}}{=} H(u_t). \quad (2.55)$$

We can start developing an idea of how this system evolves in time via *cobwebbing*, a graphical technique, as shown in Figure 2.9.

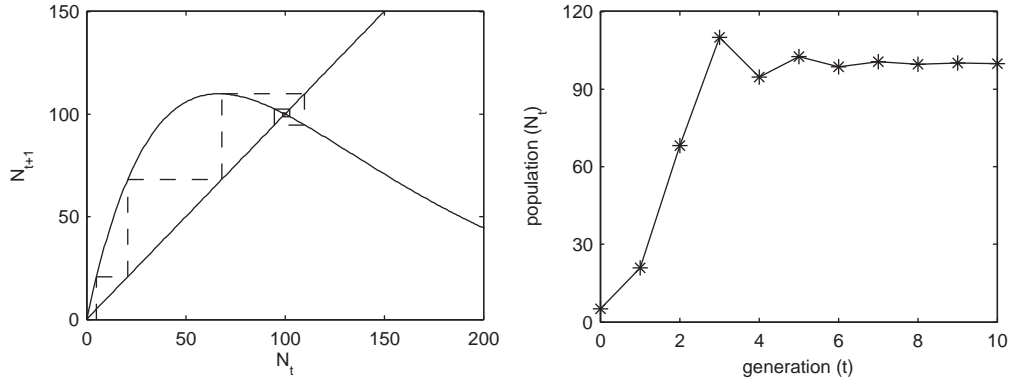
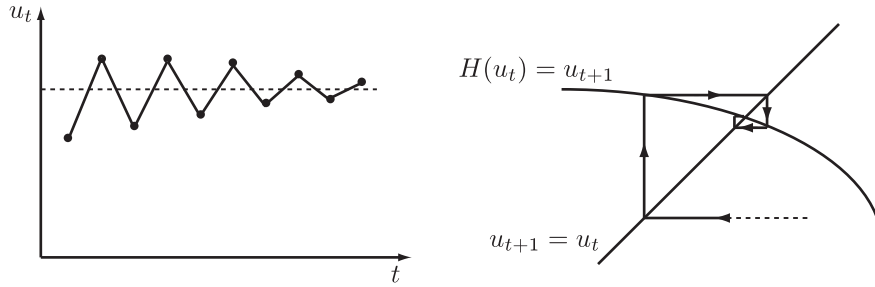


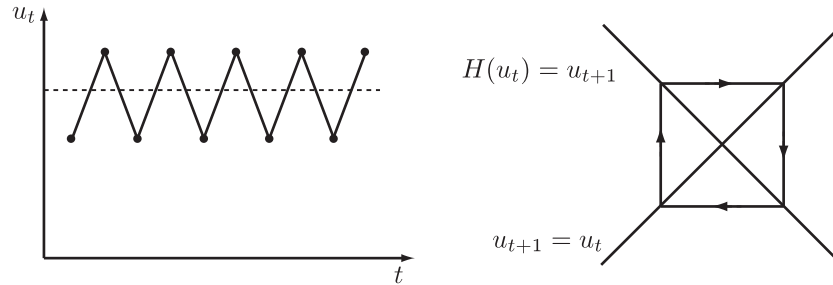
Figure 2.9: Dynamics of the Ricker model. The left-hand plot shows a plot of $N_{t+1} = N_t \exp[r(1 - N_t/K)]$ alongside $N_{t+1} = N_t$ with the cobwebbing technique shown. The right-hand plot shows N_t for successive generation times $t = 1, 2, \dots, 10$. Parameters are as follows: $N_0 = 5$, $r = 1.5$ and $K = 100$.

In particular, it is clear that the behaviour sufficiently close to a fixed point, u^* , depends on the value of $H'(u^*)$. For example:

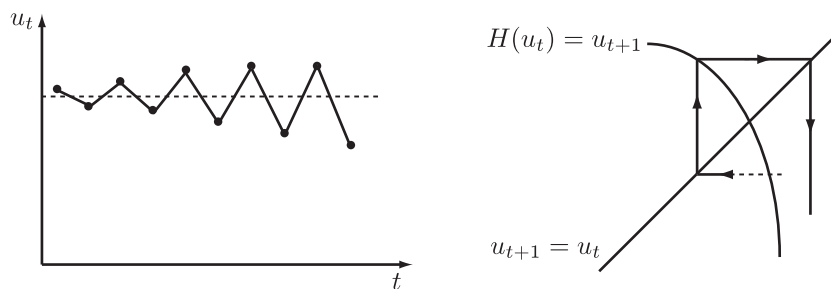
- $-1 < H'(u^*) < 0$



- $H'(u^*) = -1$



- $H'(u^*) < -1$



2.2.1 Linear stability

More generally, to consider the stability of an equilibrium point algebraically, rather than graphically, we write

$$u_t = u^* + v_t, \quad (2.56)$$

where u^* is an equilibrium value. Note that u^* is time-independent and satisfies $u^* = H(u^*)$. Hence

$$u_{t+1} = u^* + v_{t+1} = H(u^* + v_t) = H(u^*) + v_t H'(u^*) + o(v_t^2). \quad (2.57)$$

Consequently, we have

$$v_{t+1} = H'(u^*)v_t \quad \text{where } H'(u^*) \text{ is a constant, independent of } t, \quad (2.58)$$

and thus

$$v_t = [H'(u^*)]^t v_0. \quad (2.59)$$

This in turn enforces stability if $|H'(u^*)| < 1$ and instability if $|H'(u^*)| > 1$.

Definition. A discrete population model is *linearly stable* if $|H'(u^*)| < 1$.

2.2.2 Further investigation

The equations are not as simple as they seem. For example, from what we have seen thus far, the discrete time logistic model seems innocuous enough.

$$N_{t+1} = rN_t \left(1 - \frac{N_t}{K}\right), \quad r > 0 \quad K > 0. \quad (2.60)$$

If we put in enough effort, one could be forgiven for thinking that the use of cobwebbing will give a simple representation of solutions of this equation. However, the effects of increasing r are stunning. Figure 2.10 shows examples of cobwebbing when $r = 1.5$ and $r = 4.0$.

It should now be clear that even this simple equation does *not* always yield a simple solution! How do we investigate such a complicated system in more detail?

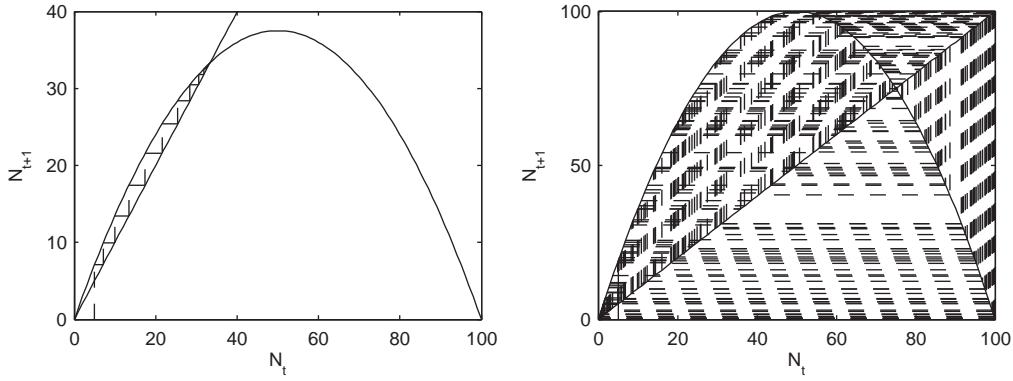


Figure 2.10: Dynamics of the discrete logistic model. The left-hand plot shows results for $r = 1.5$ whilst the right-hand plot shows results for $r = 4.0$. Other parameters are as follows: $N_0 = 5$ and $K = 100$.

Definition. A *bifurcation point* is, in the current context, a point in parameter space where the number of equilibrium points, or their stability properties, or both, change.

We proceed to take a closer look at the non-dimensional discrete logistic growth model:

$$u_{t+1} = ru_t(1 - u_t) = H(u_t), \quad (2.61)$$

for different values of the parameter r , and, in particular, we seek the values where the number or stability nature of the equilibrium points change. Note that we have equilibrium points at $u^* = 0$ and $u^* = (r - 1)/r$, and that $H'(u) = r - 2ru$.

For $0 < r < 1$, we have:

- $u^* = 0$ is a stable steady state since $|H'(0)| = |r| < 1$;
- the equilibrium point at $u^* = (r - 1)/r$ is unstable. It is also unreachable, and thus irrelevant, for physical initial conditions with $u_0 \geq 0$.

For $1 < r < 3$ we have:

- $u^* = 0$ is an unstable steady state since $|H'(0)| = |r| > 1$;
- $u^* = (r - 1)/r$ is an stable steady state since $|H'((r - 1)/r)| = |2 - r| < 1$.

In Figure 2.11 we plot this on a diagram of steady states, as a function of r , with stable steady states indicated by solid lines and unstable steady states by dashed lines.

When $r = 1$ we have $(r - 1)/r = 0$, so both equilibrium points are at $u^* = 0$, with $H'(u^* = 0) = 1$. Clearly we have a switch in the stability properties of the equilibrium points, and thus $r = 1$ is a bifurcation point.

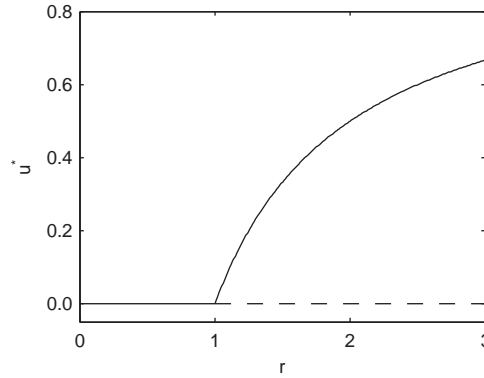


Figure 2.11: Bifurcation diagram for the non-dimensional discrete logistic model. The non-zero steady state is given, for $r > 1$, by $N^* = (r - 1)/r$.

What happens for $r > 3$? We have equilibrium points at $u^* = 0$, $u = (r - 1)/r$ and $H'(u^* = (r - 1)/r) < -1$; both equilibrium points are unstable. Hence if the system approaches one of these equilibrium points the approach is only transient; it quickly moves away. We have a switch in the stability properties of the equilibrium points, and thus $r = 3$ is a bifurcation point.

To consider the dynamics of this system once $r > 3$ we consider

$$u_{t+2} = H(u_{t+1}) = H[H(u_t)] \stackrel{\text{def}}{=} H_2(u_t) = r[ru_t(1 - u_t)][1 - ru_t(1 - u_t)]. \quad (2.62)$$

Figure 2.12 shows $H_2(u_t)$ for $r = 2.5$ and $r = 3.5$ and demonstrates the additional steady states that arise as r is increased past $r = 3$.

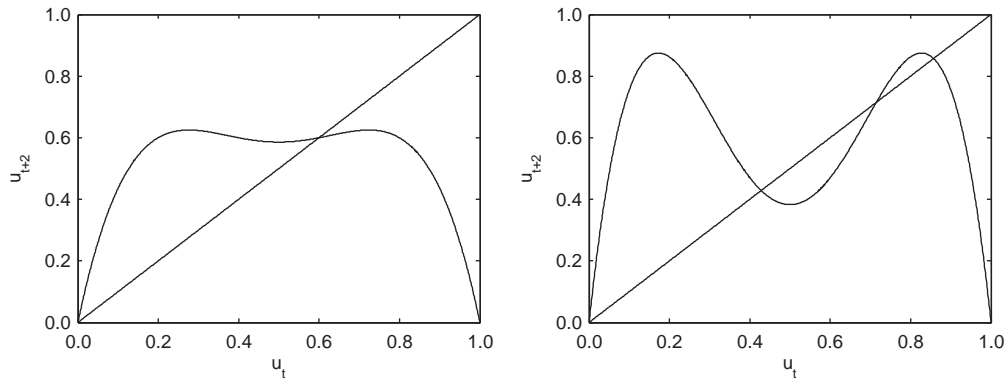


Figure 2.12: Dynamics of the non-dimensional discrete logistic model in terms of every second iteration. The left-hand plot shows results for $r = 2.5$ whilst the right-hand plot shows results for $r = 3.5$.

Note. The fixed points of H_2 satisfy $u_2^* = H_2(u_2^*)$, which is a quartic equation in u_2^* . However, we know two solutions, the fixed points $H(\cdot)$, *i.e.* 0 and $(r-1)/r$. Using standard

techniques we can reduce the quartic to a quadratic, which can be solved to reveal the further fixed points of H_2 , namely

$$u_2^* = \frac{r+1}{2r} \pm \frac{1}{2r} [(r-1)^2 - 4]^{1/2}. \quad (2.63)$$

These roots exist if $(r-1)^2 > 4$, *i.e.* $r > 3$.

Definition. The m^{th} composition of the function H is given by

$$H_m(u) \stackrel{\text{def}}{=} \underbrace{[H \cdot H \dots H \cdot H]}_{m \text{ times}}(u). \quad (2.64)$$

Definition. A point u is *periodic of period m* for the function H if

$$H_m(u) = u, \quad H_i(u) \neq u, \quad i \in \{1, 2, \dots, m-1\}. \quad (2.65)$$

Thus the points

$$u_2^* = \frac{r+1}{2r} \pm \frac{1}{2r} [(r-1)^2 - 4]^{1/2}, \quad (2.66)$$

are points of period 2 for the function H .

Problem. Show that the u_2^* are stable with respect to the function H_2 for $r > 3$, $(r-3) \ll 1$.

Let

$$u_0 \stackrel{\text{def}}{=} \frac{r+1}{2r} \pm \frac{1}{2r} [(r-1)^2 - 4]^{1/2}, \quad u_1 = H(u_0), \quad u_2 = H_2(u_0), \quad (2.67)$$

and let

$$\lambda = \frac{\partial}{\partial u} [H_2(u)]|_{u=u_0}. \quad (2.68)$$

Then

$$\lambda = \frac{\partial}{\partial u} [H \cdot H(u)]|_{u=u_0} = H'(u_0)H'(u_1). \quad (2.69)$$

Thus for stability we require $|H'(u_0)H'(u_1)| < 1$.

Exercise. Finish the problem: show that the steady states

$$u_2^* = \frac{r+1}{2r} \pm \frac{1}{2r} [(r-1)^2 - 4]^{1/2}, \quad (2.70)$$

are stable for the dynamical system $u_{t+1} = H_2(u_t)$, with $r > 3$, $(r-3) \ll 1$.

Exercise. Suppose u_0 is an equilibrium point of period m for the function H . Show that u_0 is stable if

$$\prod_{i=0}^{m-1} [H'(u_i)] < 1, \quad (2.71)$$

where $u_i = H_i(u_0)$ for $i \in \{1, 2, \dots, m-1\}$.

Solution. Defining λ in a similar manner as before, we have

$$\lambda = \left. \frac{\partial}{\partial u} H_m(u) \right|_{u=u_0}, \quad (2.72)$$

$$= \left. \frac{\partial}{\partial u} [H(Q(u))] \right|_{u=u_0}, \quad \text{where } Q(u) = H_{m-1}(u), \quad (2.73)$$

$$= H'(Q(u)) \left. \frac{\partial Q}{\partial u} \right|_{u=u_0}, \quad (2.74)$$

$$= H'(u_{m-1}) \left. \frac{\partial}{\partial u} H_{m-1} \right|_{u=u_0}. \quad (2.75)$$

Hence, by iteration, we have the result.

We plot the fixed points of H_2 , which we now know to be stable, in addition to the fixed points of H_1 , in Figure 2.13. The upper branch, u_{2U}^* , is given by the positive root of equation (2.70) whilst the lower branch, u_{2L}^* , is given by the negative root. We have $u_{2L}^* = H(u_{2U}^*)$, $u_{2U}^* = H(u_{2L}^*)$. Thus a stable, period 2, oscillation is present, at least for $(r-3) \ll 1$. Any solution which gets sufficiently close to either u_{2U}^* or u_{2L}^* stays close.

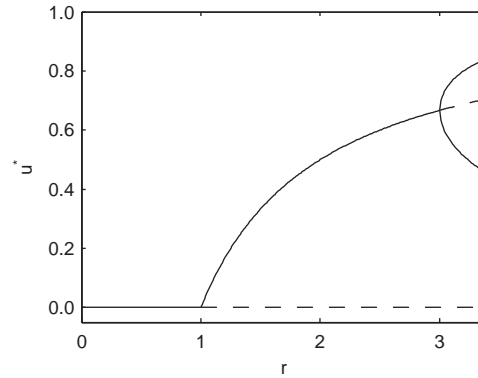


Figure 2.13: Bifurcation diagram for the non-dimensional discrete logistic model with inclusion of the period 2 solutions.

For higher values of r , there is a bifurcation point for H_2 ; we can then find a stable fixed point for $H_4(u) :: H_2[H_2(u)]$ in a similar manner. Increasing r further there is a bifurcation point for $H_4(u)$. Again, we are encountering a level of complexity which is too much to deal with our current method.

To bring further understanding to this complex system, we note the following definition.

Definition. An *orbit* generated by the point u_0 are the points $\{u_0, u_1, u_2, u_3, \dots\}$ where $u_i = H_i(u) = H(u_{i-1})$.

We are primarily interested in the large time behaviour of these systems in the context of biological applications. Thus, for a fixed value of r , we start with a reasonable initial seed, say $u^* = 0.5$, and plot the large time asymptote of the orbit of u^* , *ie.* the points $H_i(u^*)$ once i is sufficiently large for there to be no transients. This gives an intriguing plot; see Figure 2.14.

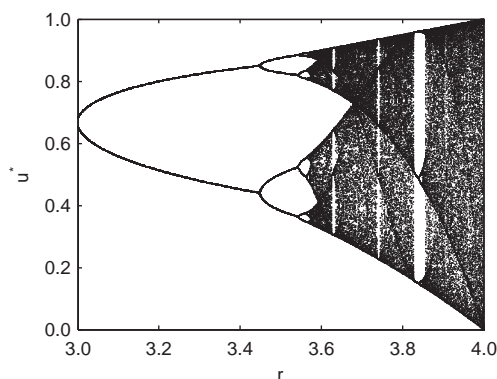


Figure 2.14: The orbit diagram of the logistic map. For each value of $r \in [3, 4]$ along horizontal axis, points on the large time orbits of the logistic map are plotted.

In particular, we have regions where, for r fixed, there are three points along the vertical corresponding to period 3 oscillations. This means any period of oscillation exists and we have a chaotic system. This can be proved using Sharkovskii's theorem. See P. Glendinning, *Stability, Instability and Chaos* [4] for more details on chaos and chaotic systems.

Note. A common discrete population model in mathematical biology is

$$N_{t+1} = \frac{rN_t}{1 + aN_t^b}. \quad (2.76)$$

Models of this form for the Colorado beetle are within the periodic regimes, while Nicholson's blowfly model is in the chaotic regime [8].

2.2.3 The wider context

In investigating the system

$$u_{t+1} = ru_t(1 - u_t) \stackrel{\text{def}}{=} H(u_t), \quad (2.77)$$

we have explored a very simple equation which, in general, exhibits greatly different behaviours with only a small change in initial conditions or parameters (*ie.* linear growth rate, r). Such sensitivity is a hallmark feature of chaotic dynamics. In particular, it makes

prediction very difficult. There will always be errors in a model's formalism, initial conditions and parameters and, in general, there is no readily discernible pattern in the way the model behaves. Thus, assuming the real system behaviour is also chaotic, using statistical techniques to extract a pattern of behaviour to thus enable an extrapolation to predict future behaviour is also fraught with difficulty. Attempting to make accurate predictions with models containing chaos is an active area of research, as is developing techniques to analyse seemingly random data to see if such data can be explained by a simple chaotic dynamical system.

Chapter 3

Continuous population models: interacting species

In this chapter we consider interacting populations, but again in the case where spatial variation is not important. Appendix A contains relevant information for phase plane analysis that may be useful.

References.

- J. D. Murray, Mathematical Biology, 3rd edition, Volume I, Chapter 3 [8].
- L. Edelstein-Keshet, Mathematical Models in Biology, Chapter 6 [2].
- N. F. Britton, Essential Mathematical Biology, Chapter 2 [1].

There are three main forms of interaction:

Predator-prey An upsurge in population I (prey) induces a growth in population II (predator). An upsurge in population II (predator) induces a decline in population I (prey).

Competition An upsurge in either population induces a decline in the other.

Symbiosis An upsurge in either population induces an increase in the other.

Of course, there are other possible interactions, such as cannibalism, especially with the “adult” of a species preying on the young, and parasitism.

3.1 Predator-prey models

The most common predator-prey model is the Lotka-Volterra model. With N the number of prey and P the number of predators, this model can be written

$$\frac{dN}{dt} = aN - bNP, \tag{3.1}$$

$$\frac{dP}{dt} = cNP - dP, \tag{3.2}$$

with a, b, c, d positive parameters and $c < b$. Non-dimensionalising with $u = (c/d)N$, $v = (b/a)P$, $\tau = at$ and $\alpha = d/a$, we have

$$\frac{1}{1/a} \frac{d}{d\tau} \frac{du}{c} = \frac{ad}{c} u - \frac{bd}{c} \frac{a}{b} uv \quad \Rightarrow \quad \frac{du}{d\tau} = u - uv = u(1 - v) \equiv f(u, v), \quad (3.3)$$

$$\frac{1}{1/a} \frac{a}{b} \frac{dv}{d\tau} = c \frac{d}{c} \frac{a}{b} uv - d \frac{a}{b} v, \quad \Rightarrow \quad \frac{dv}{d\tau} = \alpha(uv - v) = \alpha v(u - 1) \equiv g(u, v), \quad (3.4)$$

There are stationary points at $(u, v) = (0, 0)$ and $(u, v) = (1, 1)$.

Exercise. Find the stability of the stationary points $(u, v) = (0, 0)$ and $(u, v) = (1, 1)$.

The Jacobian, \mathbf{J} , is given by

$$\mathbf{J} = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix} = \begin{pmatrix} 1 - v & -u \\ \alpha v & \alpha(u - 1) \end{pmatrix}. \quad (3.5)$$

At $(0, 0)$ we have

$$\mathbf{J} = \begin{pmatrix} 1 & 0 \\ 0 & -\alpha \end{pmatrix}, \quad (3.6)$$

with eigenvalues $1, -\alpha$. Therefore the steady state $(0, 0)$ is an unstable saddle.

At $(1, 1)$ we have

$$\mathbf{J} = \begin{pmatrix} 0 & -1 \\ \alpha & 0 \end{pmatrix}, \quad (3.7)$$

with eigenvalues $\pm i\sqrt{\alpha}$. Therefore the steady state $(1, 1)$ is a centre (not linearly stable).

These equations are special; we can integrate them once, as follows, to find a conserved constant:

$$\frac{du}{dv} = \frac{u(1 - v)}{\alpha(u - 1)v} \quad \Rightarrow \quad \int \frac{u - 1}{u} du = \int \frac{1 - v}{\alpha v}. \quad (3.8)$$

Hence

$$H = \text{const} = \alpha u + v - \alpha \ln u - \ln v. \quad (3.9)$$

This can be rewritten as

$$\left(\frac{e^v}{v} \right) \left(\frac{e^u}{u} \right)^\alpha = e^H, \quad (3.10)$$

from which we can rapidly deduce that the trajectories in the (u, v) plane take the form shown in Figure 3.1. Thus u and v oscillate in time, though not in phase, and hence we have a prediction; predators and prey population numbers oscillate out of phase. There are often observations of this *e.g.* hare-lynx interactions.

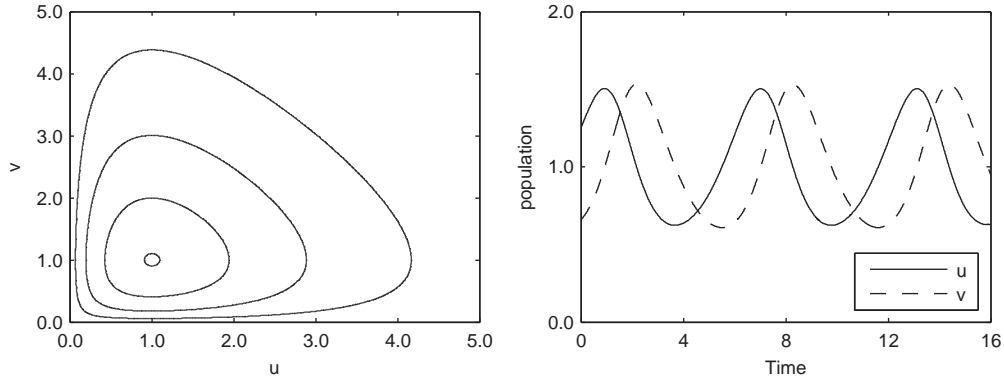


Figure 3.1: Dynamics of the non-dimensional Lotka-Volterra system for $\alpha = 1.095$ and $H = 2.1, 2.4, 3.0, 4.0$. The left-hand plot shows the dynamics in the (u, v) phase plane whilst the right-hand plot shows the temporal evolution of u and v .

3.1.1 Finite predation

The common predator-prey model assumes that as $N \rightarrow \infty$ the rate of predation per predator becomes unbounded, as does the rate of increase of the predator's population. However, with an abundance of food, these quantities will saturate rather than become unbounded. Thus, a more realistic incorporation of an abundance of prey requires a refinement of the Lotka-Volterra model. A suitable, simple, model for predator-prey interactions under such circumstances would be (after a non-dimensionalisation)

$$\frac{du}{d\tau} = f(u, v) = u(1 - u) - \frac{auv}{d + u}, \quad (3.11)$$

$$\frac{dv}{d\tau} = g(u, v) = bv \left(1 - \frac{v}{u}\right), \quad (3.12)$$

where a, b, d are positive constants. Note the effect of predation per predator saturates at high levels of u whereas the predator levels are finite at large levels of prey and drop exceedingly rapidly in the absence of prey.

There is one non-trivial equilibrium point, (u^*, v^*) , satisfying

$$v^* = u^* \quad \text{where} \quad (1 - u^*) = \frac{au^*}{d + u^*}, \quad (3.13)$$

and hence

$$u^* = \frac{1}{2} \left[-(a + d - 1) + \sqrt{(a + d - 1)^2 + 4d} \right]. \quad (3.14)$$

The Jacobian at (u^*, v^*) is

$$\mathbf{J} = \begin{pmatrix} f_u & f_v \\ g_u & g_v \end{pmatrix} \bigg|_{(u^*, v^*)}, \quad (3.15)$$

where

$$f_u(u^*, v^*) = 1 - 2u^* - \frac{au^*}{d + u^*} + \frac{au^*v^*}{(d + u^*)^2} = -u^* + \frac{a(u^*)^2}{(d + u^*)^2}. \quad (3.16)$$

$$f_v(u^*, v^*) = -\frac{au^*}{d + u^*}, \quad (3.17)$$

$$g_u(u^*, v^*) = \frac{b(v^*)^2}{(u^*)^2} = b, \quad (3.18)$$

$$g_v(u^*, v^*) = b \left(1 - 2\frac{v^*}{u^*} \right) = -b. \quad (3.19)$$

The eigenvalues satisfy

$$(\lambda - f_u)(\lambda - g_v) - f_v g_u = 0 \quad \Rightarrow \quad \lambda^2 - (f_u + g_v)\lambda + (f_u g_v - f_v g_u) = 0, \quad (3.20)$$

and hence

$$\lambda^2 - \alpha\lambda + \beta = 0 \quad \Rightarrow \quad \lambda = \frac{\alpha \pm \sqrt{\alpha^2 - 4\beta}}{2}, \quad (3.21)$$

where

$$\alpha = -u^* + \frac{a(u^*)^2}{(u^* + d)^2} - b, \quad \beta = b \left(u^* - \frac{a(u^*)^2}{(u^* + d)^2} - (u^* - 1) \right). \quad (3.22)$$

Note that

$$\beta = 1 - \frac{a(u^*)^2}{(u^* + d)^2} = 1 - \frac{u^*(1 - u^*)}{(u^* + d)} = \frac{(u^* + d) - u^* + (u^*)^2}{u^* + d} = \frac{d + (u^*)^2}{d + u^*} > 0. \quad (3.23)$$

Thus, if $\alpha < 0$ the eigenvalues λ are such that we have either:

- a stable node ($\alpha^2 - 4\beta > 0$);
- stable focus ($\alpha^2 - 4\beta < 0$);

at the equilibrium point (u^*, v^*) .

If $\alpha > 0$ we have an unstable equilibrium point at (u^*, v^*) .

3.2 A look at global behaviour

This previous section illustrated local dynamics: we have conditions for when the dynamics will stably remain close to the non-trivial equilibrium point. One is also often interested in the global dynamics. However, determining the global dynamics of a system, away from its equilibrium points, is a much more difficult problem compared to ascertaining the local dynamics, sufficiently close to the equilibrium points. For specific parameter values, one can readily solve the ordinary differential equations to consider the behaviour of the system. One is also interested in the general properties of the global behaviour. This is more difficult, and we will consider one possible approach below.

There are many potential tools available: nullcline analysis, the Poincaré-Bendixson Theorem, the Poincaré Index and the Bendixson-Dulac Criterion. The Poincaré-Bendixson Theorem is a useful tool for proving that limit cycles must exist, while Poincaré indices and the Bendixson-Dulac Criterion are useful tools for proving a limit cycle cannot exist.

We will *briefly* consider nullclines and the Poincaré-Bendixson Theorem in detail. Please refer to P. Glendinning, *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations* [4], or D. W. Jordan and P. Smith, *Mathematical Techniques: An Introduction for Engineering, Mathematical and Physical Sciences* [6], for further details than considered here.

3.2.1 Nullclines

Definition. Consider the equations

$$\frac{du}{dt} = f(u, v), \quad \frac{dv}{dt} = g(u, v). \quad (3.24)$$

The *nullclines* are the curves in the phase plane where $f(u, v) = 0$ and $g(u, v) = 0$.

Reconsider

$$\frac{du}{d\tau} = f(u, v) = u(1 - u) - \frac{auv}{d + u}, \quad (3.25)$$

$$\frac{dv}{d\tau} = g(u, v) = bv \left(1 - \frac{v}{u}\right). \quad (3.26)$$

The u nullclines are given by

$$f(u, v) \equiv 0 \quad \Rightarrow \quad u \equiv 0 \quad \text{and} \quad v = \frac{1}{a}(1 - u)(u + d). \quad (3.27)$$

The v nullclines are given by

$$g(u, v) \equiv 0 \quad \Rightarrow \quad v \equiv 0 \quad \text{and} \quad v = u. \quad (3.28)$$

A sketch of the nullclines and the behaviour of the phase plane trajectories is shown in Figure 3.2.

3.2.2 The Poincaré-Bendixson Theorem

For a system of two first order ordinary differential equations, consider a closed bounded region D . Suppose a positive half path, H , lies entirely within D . Then one of the following is true:

1. H is a closed trajectory, *e.g.* a limit cycle;
2. H asymptotically tends to a closed trajectory, *e.g.* a limit cycle;
3. H terminates on a stationary point.

Therefore, if D does not have a stationary point then there must be a limit cycle.

For a proof see P. Glendinning, *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations* [4].

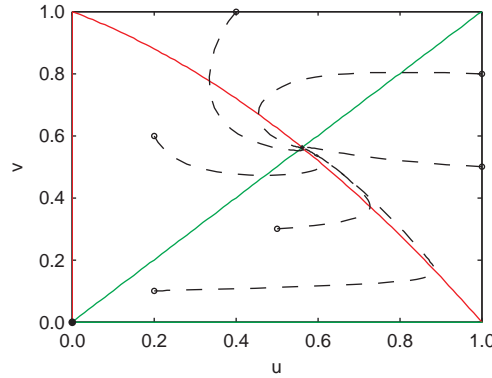


Figure 3.2: The (u, v) phase-plane for the finite predation model when the steady state is stable. The u nullclines are plotted in red and the v nullclines in green. Trajectories for a number of different initial conditions are shown as dashed lines. Parameters are as follows: $a = 2.0$, $b = 0.1$, $d = 2.0$.

Exercise. Explain why $\alpha > 0$ in the previous example (see equation (3.22)) implies we have limit cycle dynamics. What does this mean in terms of the population levels of predator and prey?

Solution. For $\alpha > 0$ the steady state is an unstable node or spiral. Further, we can find a simple, closed boundary curve, \mathcal{C} , in the positive quadrant of the (u, v) plane, such that on \mathcal{C} phase trajectories always point into the domain, \mathcal{D} , enclosed by \mathcal{C} . Applying the Poincaré-Bendixon Theorem to the domain gives the existence of a limit cycle. See J. D. Murray, *Mathematical Biology Volume I* [8] (Chapter 3.4) for more details.

3.3 Competitive exclusion

We consider an ordinary differential equation model of two competitors. An example might be populations of red squirrels and grey squirrels [8]. Here, both populations compete for the same resources and a typical model for their dynamics is

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1}{K_1} - b_{12} \frac{N_2}{K_1} \right), \quad (3.29)$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2}{K_2} - b_{21} \frac{N_1}{K_2} \right), \quad (3.30)$$

where $K_1, K_2, r_1, r_2, b_{12}, b_{21}$ are positive constants. Let us associate N_1 with red squirrels and N_2 with grey squirrels in our example.

In particular, given a range of parameter values and some initial values for N_1 and N_2 at the time $t = 0$, we would typically like to know if the final outcome is one of the following possibilities:

- the reds become extinct, leaving the greys;

- the greys become extinct, leaving the reds;
- both reds and greys become extinct;
- the reds and greys co-exist. If this system is perturbed in any way will the reds and greys continue to coexist?

After a non-dimensionalisation (exercise) we have

$$u_1' = u_1(1 - u_1 - \alpha_{12}u_2) \stackrel{\text{def}}{=} f_1(u_1, u_2), \quad (3.31)$$

$$u_2' = \rho u_2(1 - u_2 - \alpha_{21}u_1) \stackrel{\text{def}}{=} f_2(u_1, u_2), \quad (3.32)$$

where $\rho = r_2/r_1$.

The stationary states are

$$(u_1^*, u_2^*) = (0, 0), \quad (u_1^*, u_2^*) = (1, 0), \quad (u_1^*, u_2^*) = (0, 1), \quad (3.33)$$

and

$$(u_1^*, u_2^*) = \frac{1}{1 - \alpha_{12}\alpha_{21}}(1 - \alpha_{12}, 1 - \alpha_{21}), \quad (3.34)$$

if $\alpha_{12} < 1$ and $\alpha_{21} < 1$ or $\alpha_{12} > 1$ and $\alpha_{21} > 1$.

The Jacobian is

$$\mathbf{J} = \begin{pmatrix} 1 - 2u_1 - \alpha_{12}u_2 & -\alpha_{12}u_1 \\ -\rho\alpha_{21}u_2 & \rho(1 - 2u_2 - \alpha_{21}u_1) \end{pmatrix}. \quad (3.35)$$

It is a straightforward application of phase plane techniques to investigate the nature of these equilibrium points:

Steady state $(u_1^*, u_2^*) = (0, 0)$.

$$\mathbf{J} - \lambda\mathbf{I} = \begin{pmatrix} 1 - \lambda & 0 \\ 0 & \rho - \lambda \end{pmatrix} \Rightarrow \lambda = 1, \rho. \quad (3.36)$$

Therefore $(0, 0)$ is an unstable node.

Steady state $(u_1^*, u_2^*) = (1, 0)$.

$$\mathbf{J} - \lambda\mathbf{I} = \begin{pmatrix} -1 - \lambda & -\alpha_{12} \\ 0 & \rho(1 - \alpha_{21}) - \lambda \end{pmatrix} \Rightarrow \lambda = -1, \rho(1 - \alpha_{21}). \quad (3.37)$$

Therefore $(1, 0)$ is a stable node if $\alpha_{21} > 1$ and a saddle point if $\alpha_{21} < 1$.

Steady state $(u_1^*, u_2^*) = (0, 1)$.

$$\mathbf{J} - \lambda\mathbf{I} = \begin{pmatrix} 1 - \alpha_{12} - \lambda & 0 \\ -\rho\alpha_{21} & -\rho - \lambda \end{pmatrix} \Rightarrow \lambda = -\rho, 1 - \alpha_{12}. \quad (3.38)$$

Therefore $(0, 1)$ is a stable node if $\alpha_{12} > 1$ and a saddle point if $\alpha_{12} < 1$.

Steady state $(u_1^*, u_2^*) = \frac{1}{1 - \alpha_{12}\alpha_{21}}(1 - \alpha_{12}, 1 - \alpha_{21})$.

$$\mathbf{J} - \lambda \mathbf{I} = \frac{1}{1 - \alpha_{12}\alpha_{21}} \begin{pmatrix} \alpha_{21} - 1 - \lambda & \alpha_{12}(\alpha_{12} - 1) \\ \rho\alpha_{21}(\alpha_{21} - 1) & \rho(\alpha_{21} - 1) - \lambda \end{pmatrix}. \quad (3.39)$$

Stability depends on α_{12} and α_{21} .

There are several different possible behaviours. The totality of all behaviours of the above model is reflected in how one can arrange the nullclines within the positive quadrant. However, for competing populations these straight line nullclines have negative gradients. Figure 3.3 shows the model behaviour for different sets of parameter values.

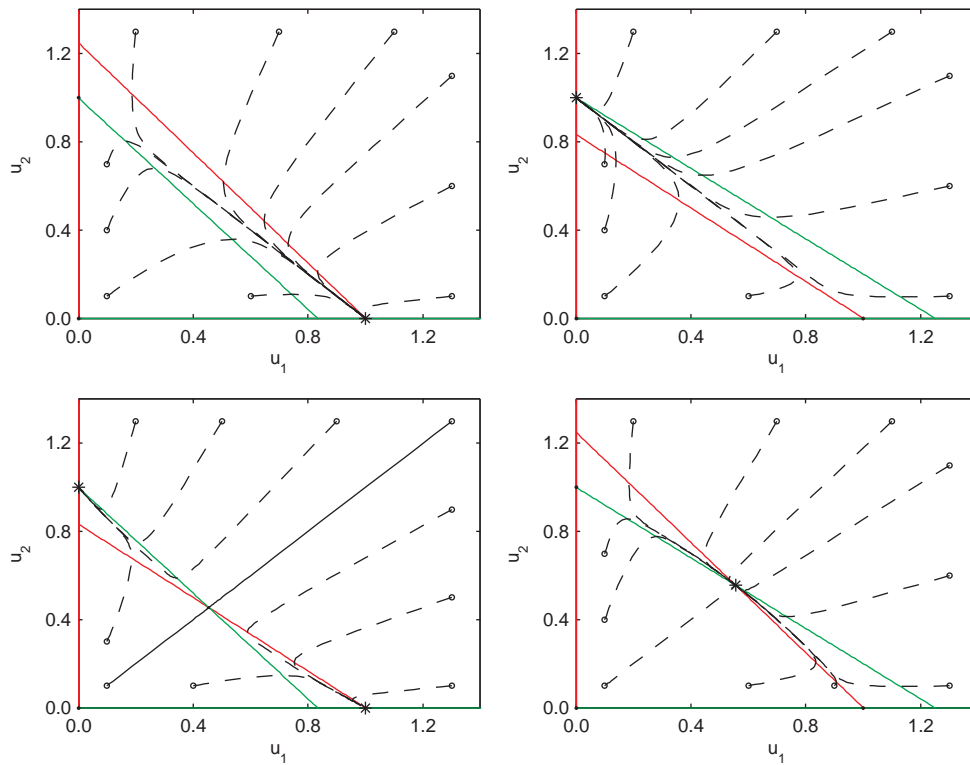


Figure 3.3: Dynamics of the non-dimensional competitive exclusion system. Top left: $\alpha_{12} = 0.8 < 1$, $\alpha_{21} = 1.2 > 1$ and u_2 is excluded. Top right: $\alpha_{12} = 1.2 > 1$, $\alpha_{21} = 0.8 < 1$ and u_1 is excluded. Bottom left: $\alpha_{12} = 1.2 > 1$, $\alpha_{21} = 1.2 > 1$ and exclusion is dependent on the initial conditions. Bottom right: $\alpha_{12} = 0.8 < 1$, $\alpha_{21} = 0.8 < 1$ and we have coexistence. The stable steady states are marked with *'s and $\rho = 1.0$ in all cases. The red lines indicate $f_1 \equiv 0$ whilst the green lines indicate $f_2 \equiv 0$.

Note. In ecology the concept of *competitive exclusion* is that two species competing for exactly the same resources cannot stably coexist. One of the two competitors will always have an ever so slight advantage over the other that leads to extinction of the second competitor in the long run (or evolution into distinct ecological niches).

3.4 Mutualism (symbiosis)

We consider the same ordinary differential equation model for two competitors, *i.e.*

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1}{K_1} + b_{12} \frac{N_2}{K_1} \right), \quad (3.40)$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2}{K_2} + b_{21} \frac{N_1}{K_2} \right), \quad (3.41)$$

where $K_1, K_2, r_1, r_2, b_{12}, b_{21}$ are positive constants or, after non-dimensionalisation,

$$u'_1 = u_1(1 - u_1 + \alpha_{12}u_2) \stackrel{def}{=} f_1(u_1, u_2), \quad (3.42)$$

$$u'_2 = \rho u_2(1 - u_2 + \alpha_{21}u_1) \stackrel{def}{=} f_2(u_1, u_2). \quad (3.43)$$

In symbiosis, the straight line nullclines will have positive gradients leading to the following two possible behaviours shown in Figure 3.4.

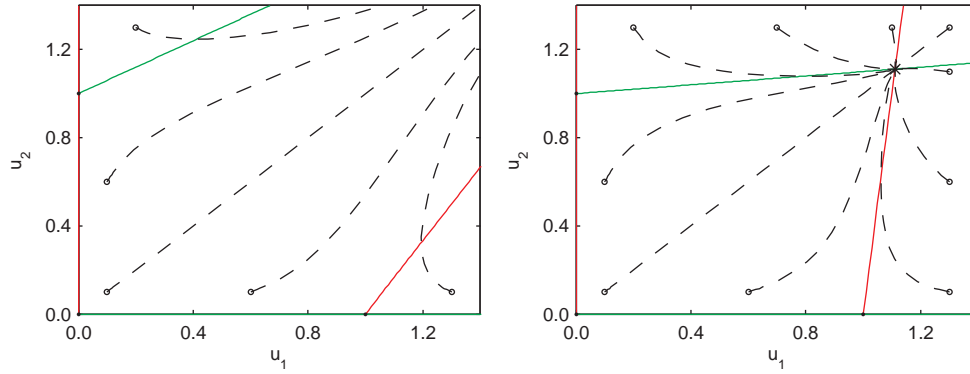


Figure 3.4: Dynamics of the non-dimensional symbiotic system. The left-hand figure shows population explosion ($\alpha_{12} = 0.6 = \alpha_{21}$) whilst the right-hand figure shows population coexistence ($\alpha_{12} = 0.1 = \alpha_{21}$). The stable steady states are marked with *'s and $\rho = 1.0$ in all cases. The red lines indicate $f_1 \equiv 0$ whilst the green lines indicate $f_2 \equiv 0$.

3.5 Interacting discrete models

It is also possible, and sometimes useful, to consider interacting discrete models which take the form

$$u_{t+1} = f(u_t, v_t), \quad (3.44)$$

$$v_{t+1} = g(u_t, v_t), \quad (3.45)$$

and possess steady states at the solutions of

$$u^* = f(u^*, v^*), \quad v^* = g(u^*, v^*). \quad (3.46)$$

It is interesting and relevant to study the linear stability of these equilibrium points, and the global dynamics, but we do not have time to pursue this here.

Chapter 4

Enzyme kinetics

In this chapter we consider enzyme kinetics, which can be thought of as a particular case of an interacting species model. In all cases here we will neglect spatial variation.

Throughout, we will consider the m chemical species C_1, \dots, C_m .

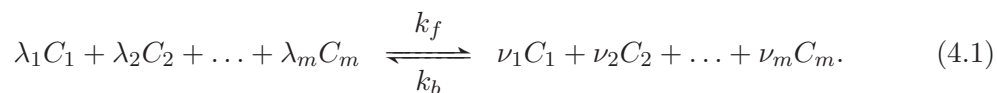
- The concentration of C_i , denoted c_i , is defined to be the number of molecules of C_i per unit volume.
- A standard unit of concentration is moles m^{-3} , often abbreviated to mol m^{-3} . Recall that 1 mole = 6.023×10^{23} molecules.

References.

- J. D. Murray, Mathematical Biology, 3rd edition, Volume I, Chapter 6 [8].
- J. P. Keener and J. Sneyd, Mathematical Physiology, Chapter 1 [7].

4.1 The Law of Mass Action

Suppose C_1, \dots, C_m undergo the reaction



The Law of Mass Action states that the forward reaction proceeds at rate

$$k_f c_1^{\lambda_1} c_2^{\lambda_2} \dots c_m^{\lambda_m}, \quad (4.2)$$

while the back reaction proceeds at the rate

$$k_b c_1^{\nu_1} c_2^{\nu_2} \dots c_m^{\nu_m}, \quad (4.3)$$

where k_f and k_b are dimensional constants that must be determined empirically.

Note 1. Strictly, to treat k_f , k_b above as constant, we have to assume that the temperature is constant. This is a very good approximation for most biochemical reactions occurring in, for example, physiological systems. However, if one wanted to model reactions that produce extensive heat for example, burning petrol, one must include the temperature dependence in k_f and k_b and subsequently keep track of how hot the system gets as the reaction proceeds. This generally makes the modelling significantly more difficult. Below we assume that we are dealing with systems where the temperature is approximately constant as the reaction proceeds.

Note 2. The Law of Mass Action for chemical reactions can be derived from statistical mechanics under quite general conditions (see for example L. E. Riechl, A Modern Course in Statistical Physics [11]).

Note 3. As we will see later, the Law of Mass Action is also used in biological scenarios to write down equations describing, for example, the interactions of people infected with, and people susceptible to, a pathogen during an epidemic. However, in such circumstances its validity must be taken as an assumption of the modelling; in such scenarios one cannot rely on thermodynamic/statistical mechanical arguments to justify the Law of Mass Action.

4.2 Michaelis-Menten kinetics

Michaelis-Menten kinetics approximately describe the dynamics of a number of enzyme systems. The reactions are



Letting c denoting the concentration of the complex SE , and s , e , p denoting the concentrations of S , E , P , respectively, we have, from the Law of Mass Action, the following ordinary differential equations:

$$\frac{ds}{dt} = -k_1se + k_{-1}c, \quad (4.6)$$

$$\frac{dc}{dt} = k_1se - k_{-1}c - k_2c, \quad (4.7)$$

$$\frac{de}{dt} = -k_1se + k_{-1}c + k_2c, \quad (4.8)$$

$$\frac{dp}{dt} = k_2c. \quad (4.9)$$

Note that the equation for p decouples and hence we can neglect it initially.

The initial conditions are:

$$s(0) = s_0, \quad e(0) = e_0 \ll s_0, \quad c(0) = 0, \quad p(0) = 0. \quad (4.10)$$

Key Point. In systems described by the Law of Mass Action, linear combinations of the variables are often conserved. In this example we have

$$\frac{d}{dt}(e + c) = 0 \quad \Rightarrow \quad e = e_0 - c, \quad (4.11)$$

and hence the equations simplify to:

$$\frac{ds}{dt} = -k_1(e_0 - c)s + k_{-1}c, \quad (4.12)$$

$$\frac{dc}{dt} = k_1(e_0 - c)s - (k_{-1} + k_2)c, \quad (4.13)$$

with the determination of p readily achievable once we have the dynamics of s and c .

4.2.1 Non-dimensionalisation

We non-dimensionalise as follows:

$$\tau = k_1 e_0 t, \quad u = \frac{s}{s_0}, \quad v = \frac{c}{e_0}, \quad \lambda = \frac{k_2}{k_1 s_0}, \quad \epsilon \stackrel{\text{def}}{=} \frac{e_0}{s_0} \ll 1, \quad K \stackrel{\text{def}}{=} \frac{k_{-1} + k_2}{k_1 s_0}, \quad (4.14)$$

which yields

$$u' = -u + (u + K - \lambda)v, \quad (4.15)$$

$$\epsilon v' = u - (u + K)v, \quad (4.16)$$

where $u(0) = 1$, $v(0) = 0$ and $\epsilon \ll 1$. Normally $\epsilon \sim 10^{-6}$. Setting $\epsilon = 0$ yields

$$v = \frac{u}{u + K}, \quad (4.17)$$

which is inconsistent with the initial conditions. Thus we have a singular perturbation problem; there must be a (boundary) region with respect to the time variable around $t = 0$ where $v' \approx \mathcal{O}(1)$. Indeed for the initial conditions given we find $v'(0) \sim \mathcal{O}(1/\epsilon)$, with $u(0)$, $v(0) \leq \mathcal{O}(1)$. This gives us the scaling we need for a singular perturbation investigation.

4.2.2 Singular perturbation investigation

We consider

$$\sigma = \frac{\tau}{\epsilon}, \quad (4.18)$$

with

$$u(\tau, \epsilon) = \tilde{u}(\sigma, \epsilon) = \tilde{u}_0(\sigma) + \epsilon \tilde{u}_1(\sigma) + \dots, \quad (4.19)$$

$$v(\tau, \epsilon) = \tilde{v}(\sigma, \epsilon) = \tilde{v}_0(\sigma) + \epsilon \tilde{v}_1(\sigma) + \dots \quad (4.20)$$

Proceeding in the usual way, we find that \tilde{u}_0 , \tilde{v}_0 satisfy

$$\frac{d\tilde{u}_0}{d\sigma} = 0 \quad \Rightarrow \quad \tilde{u}_0 = \text{constant} = 1, \quad (4.21)$$

and

$$\frac{d\tilde{v}_0}{d\sigma} = \tilde{u}_0 - (1 + K)\tilde{v}_0 = 1 - (1 + K)\tilde{v}_0 \quad \Rightarrow \quad \tilde{v}_0 = \frac{1 - e^{-(1+K)\sigma}}{1 + K}, \quad (4.22)$$

which gives us the *inner* solution.

To find the *outer* solution we expand

$$u(\tau, \epsilon) = u_0(\tau) + \epsilon u_1(\tau) + \dots, \quad (4.23)$$

$$v(\tau, \epsilon) = v_0(\tau) + \epsilon v_1(\tau) + \dots, \quad (4.24)$$

within the equations

$$u' = -u + (u + K - \lambda)v, \quad (4.25)$$

$$\epsilon v' = u - (u + K)v, \quad (4.26)$$

to find that

$$\frac{du_0}{d\tau} = -u_0 + (u_0 + K - \lambda)v_0, \quad (4.27)$$

and

$$0 = u_0 - (u_0 + K)v_0. \quad (4.28)$$

This gives

$$v_0 = \frac{u_0}{u_0 + K} \quad \text{and} \quad \frac{du_0}{d\tau} = -\frac{\lambda u_0}{u_0 + K}. \quad (4.29)$$

In order to match the solutions as $\sigma \rightarrow \infty$ and $\tau \rightarrow 0$ we require

$$\lim_{\sigma \rightarrow \infty} \tilde{u}_0 = \lim_{\tau \rightarrow 0} u_0 = 1 \quad \text{and} \quad \lim_{\sigma \rightarrow \infty} \tilde{v}_0 = \lim_{\tau \rightarrow 0} v_0 = \frac{1}{1 + K}. \quad (4.30)$$

Thus the solution looks like that shown in Figure 4.1.

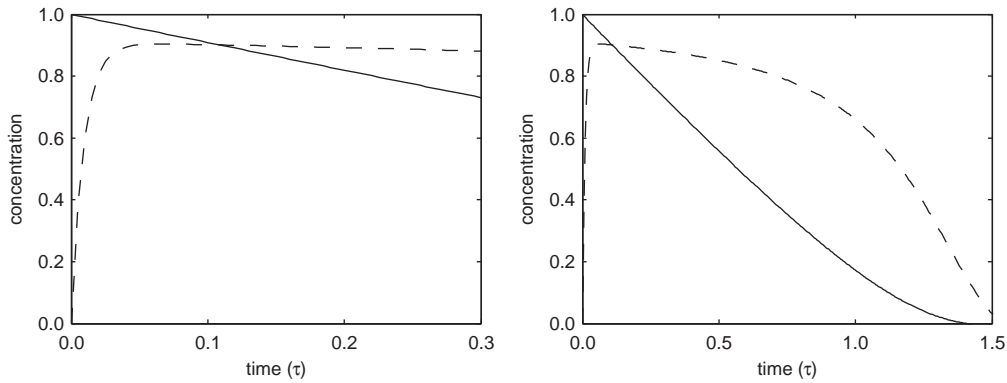


Figure 4.1: Numerical solution of the non-dimensional Michaelis-Menten equations clearly illustrating the two different time scales. The u dynamics are indicated by the solid line and the v dynamics by the dashed line. Parameters are $\epsilon = 0.01$, $K = 0.1$ and $\lambda = 1.0$.

Often the initial, fast, transient is not seen or modelled and one considers just the outer equations with a suitably adjusted initial condition (ultimately determined from consistency/matching with the inner solution). Thus one often uses *Michaelis-Menten kinetics* where the equations are simply:

$$\frac{du}{dt} = -\frac{\lambda u}{u + K} \quad \text{with} \quad u(0) = 1 \quad \text{and} \quad v = \frac{u}{u + K}. \quad (4.31)$$

Definition. We have, approximately, that $dv/d\tau \simeq 0$ using Michaelis-Menten kinetics. Taking the temporal dynamics to be trivial,

$$\frac{dv}{d\tau} \simeq 0, \quad (4.32)$$

when the time derivative is fast, *i.e.* of the form

$$\epsilon \frac{dv}{d\tau} = g(u, v), \quad (4.33)$$

where $\epsilon \ll 1$, $g(u, v) \sim \mathcal{O}(1)$, is known as the *pseudo-steady state hypothesis* and is a common assumption in the literature. We have seen its validity in the case of enzyme kinetics about at least away from the inner region.

Note. One must remember that the Michaelis-Menten kinetics derived above are a very useful approximation, but that they hinge on the validity of the Law of Mass Action. Even in simple biological systems the Law of Mass Action may breakdown. One (of many) reasons, and one that is potentially relevant at the sub-cellular level, is that the system in question has too few reactant molecules to justify the statistical mechanical assumptions underlying the Law of Mass Action. Another reason is that the reactants are not well-mixed, but vary spatially as well as temporally. We will see what happens in this case later in the course.

4.3 More complex systems

Here we consider a number of other simple systems involving enzymatic reactions. In each case the Law of Mass Action is used to write down a system of ordinary differential equations describing the dynamics of the various reactants. See J. Keener and J. Sneyd, *Mathematical Physiology* [7], for more details.

4.3.1 Several enzyme reactions and the pseudo-steady state hypothesis

We can have multiple enzymes. In general the system of equations reduces to

$$u' = f(u, v_1, \dots, v_n), \quad (4.34)$$

$$\epsilon_i v_i' = g_i(u, v_1, \dots, v_n), \quad (4.35)$$

for $i \in \{1, \dots, n\}$, while the pseudo-steady state hypothesis gives a single ordinary differential equation

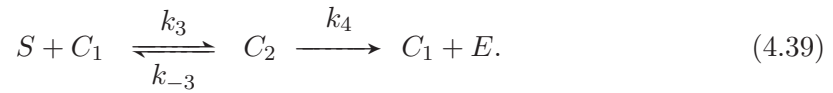
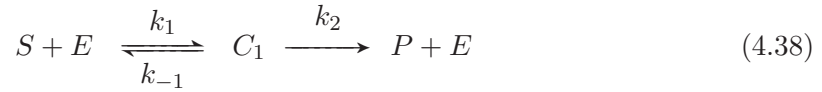
$$u' = f(u, v_1(u), \dots, v_n(u)), \quad (4.36)$$

where $v_1(u), \dots, v_n(u)$ are the appropriate roots of the equations

$$g_i(u, v_1, \dots, v_n) = 0, \quad i \in \{1, \dots, n\}. \quad (4.37)$$

4.3.2 Allosteric enzymes

Here the binding of one substrate molecule at one site affects the binding of another substrate molecules at other sites. A typical reaction scheme is:



Further details on the investigation of such systems can be found in J. D. Murray, Mathematical Biology Volume I [8], and J. P. Keener and J. Sneyd, Mathematical Physiology [7].

4.3.3 Autocatalysis and activator-inhibitor systems

Here a molecule catalyses its own production. The simplest example is the reaction scheme



though of course the positive feedback in autocatalysis is usually ameliorated by inhibition from another molecule. This leads to an example of an activator-inhibitor system which have a very rich behaviour. Other examples of these systems are given below.

Example 1

This model qualitatively incorporates activation and inhibition:

$$\frac{du}{dt} = \frac{a}{b+v} - cu, \quad (4.41)$$

$$\frac{dv}{dt} = du - ev. \quad (4.42)$$

Example 2

This model is commonly referred to as the Gierer-Meinhardt model [3]:

$$\frac{du}{dt} = a - bu + \frac{u^2}{v}, \quad (4.43)$$

$$\frac{dv}{dt} = u^2 - v. \quad (4.44)$$

Example 3

This model is commonly referred to as the Thomas model [8]. Proposed in 1975, it is an empirical model based on a specific reaction involving uric acid and oxygen:

$$\frac{du}{dt} = a - u - \rho R(u, v), \quad (4.45)$$

$$\frac{dv}{dt} = \alpha(b - v) - \rho R(u, v), \quad (4.46)$$

where

$$R(u, v) = \frac{uv}{1 + u + Ku^2}, \quad (4.47)$$

represents the interactive uptake.

Chapter 5

Introduction to spatial variation

We have initially considered biological, biochemical and ecological phenomena with negligible spatial variation. This is, however, often not the case. Consider a biochemical reaction as an example. Suppose this reaction is occurring among solutes in a relatively large, *unstirred* solution. Then the dynamics of the system is not only governed by the dynamics of the rate at which the biochemical react, but also by the fact there can be spatial variation in solute concentrations, which entails that diffusion of the reactants can occur. Thus modelling such a system requires taking into account both reaction and diffusion.

We have a similar problem for population and ecological models when we wish to incorporate the tendency of a species to spread into a region it has not previously populated. Key examples include modelling ecological invasions, where one species invades another's territory (as with grey and red squirrels in the UK [10]), or modelling the spread of disease. In some, though by no means all, of these ecological and disease-spread models the appropriate transport mechanism is again diffusion, once more requiring that we model both reaction and diffusion in a spatially varying system.

In addition, motile cells can move in response to external influences, such as chemical concentrations, light, mechanical stress and electric fields, among others. Of particular interest is modelling when motile cells respond to gradients in chemical concentrations, a process known as chemotaxis, and we will also consider this scenario.

Thus, in the following chapters, we will study how to model such phenomena and how (when possible) to solve the resulting equations in detail, for various models motivated from biology, biochemistry and ecology.

References.

- J. D. Murray, Mathematical Biology Volume I, Chapter 11 [8].
- N. F. Britton, Essential Mathematical Biology, Chapter 5 [1].

5.1 Derivation of the reaction-diffusion equations

Let $i \in \{1, \dots, m\}$. Suppose the chemical species C_i , of concentration c_i , is undergoing a reaction such that, in the absence of diffusion, one has

$$\frac{dc_i}{dt} = R_i(c_1, c_2, \dots, c_m). \quad (5.1)$$

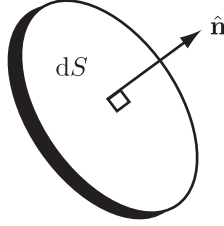
Recall that $R_i(c_1, c_2, \dots, c_m)$ is the total rate of production/destruction of C_i *per unit volume*, *i.e.* it is the rate of change of the concentration c_i .

Let t denote time, and \mathbf{x} denote the position vector of a point in space. We define

- $c(\mathbf{x}, t)$ to be the concentration of (say) a chemical (typically measured in mol m^{-3}).
- $\mathbf{q}(\mathbf{x}, t)$ to be the flux of the same chemical (typically measured in $\text{mol m}^{-2} \text{s}^{-1}$).

Recall that the flux of a chemical is defined to be such that, for a given infinitesimal surface element of area dS and unit normal $\hat{\mathbf{n}}$, the amount of chemical flowing through the surface element in an infinitesimal time interval, of duration dt , is given by

$$\hat{\mathbf{n}} \cdot \mathbf{q} \, dS dt. \quad (5.2)$$



Definition. *Fick's Law of Diffusion* relates the flux \mathbf{q} to the gradient of c via

$$\mathbf{q} = -D \nabla c, \quad (5.3)$$

where D , the diffusion coefficient, is independent of c and ∇c .

Bringing this together, we have, for any closed volume V (fixed in time and space), with bounding surface ∂V ,

$$\frac{d}{dt} \int_V c_i \, dV = - \int_{\partial V} \mathbf{q} \cdot \mathbf{n} \, dS + \int_V R_i(c_1, c_2, \dots, c_m) \, dV, \quad i \in \{1, \dots, m\}. \quad (5.4)$$

Hence

$$\frac{d}{dt} \int_V c_i \, dV = \int_V \nabla \cdot \mathbf{q} \, dV + \int_V R_i(c_1, c_2, \dots, c_m) \, dV \quad (5.5)$$

$$= \int_V \{ \nabla \cdot (D \nabla c_i) + R_i(c_1, c_2, \dots, c_m) \} \, dV, \quad (5.6)$$

and thus for any closed volume, V , with surface ∂V , one has

$$\int_V \left\{ \frac{\partial c_i}{\partial t} - \nabla \cdot (D \nabla c_i) - R_i \right\} dV = 0, \quad i \in \{1, \dots, m\}. \quad (5.7)$$

Hence

$$\frac{\partial c_i}{\partial t} = \nabla \cdot (D \nabla c_i) + R_i, \quad x \in \mathcal{D}, \quad (5.8)$$

which constitutes a system of reaction-diffusion equations for the m chemical species in the finite domain \mathcal{D} . Such equations must be supplemented with initial and boundary conditions for each of the m chemicals.

Warning. Given, for example, that

$$\int_0^{2\pi} \cos \theta d\theta = 0 \quad \not\Rightarrow \quad \cos \theta = 0, \quad \theta \in [0, 2\pi], \quad (5.9)$$

are you sure one can deduce equation (5.8)?

Suppose

$$\frac{\partial c_i}{\partial t} - \nabla \cdot (D \nabla c_i) - R_i \neq 0, \quad (5.10)$$

at some $\mathbf{x} = \mathbf{x}^*$. Without loss of generality, we can assume the above expression is positive *i.e.* the left-hand side of equation (5.10) is positive.

Then $\exists \epsilon > 0$ such that

$$\frac{\partial c_i}{\partial t} - \nabla \cdot (D \nabla c_i) - R_i > 0, \quad (5.11)$$

for all $\mathbf{x} \in \mathcal{B}_\epsilon(\mathbf{x}^*)$.

In this case

$$\int_{\mathcal{B}_\epsilon(\mathbf{x}^*)} \left[\frac{\partial c_i}{\partial t} - \nabla \cdot (D \nabla c_i) - R_i \right] dV > 0, \quad (5.12)$$

contradicting our original assumption, equation (5.7).

Hence our initial supposition is false and equation (5.8) holds for $\mathbf{x} \in \mathcal{D}$.

Remark. With one species, with a constant diffusion coefficient, in the absence of reactions, we have the diffusion equation which in one dimension reduces to

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2}. \quad (5.13)$$

For a given length scale, L , and diffusion coefficient, D , the timescale of the system is $T = L^2/D$. For a cell, $L \sim 10^{-5}\text{m} = 10^{-3}\text{cm}$, and for a typical protein $D \sim 10^{-7}\text{cm}^2\text{s}^{-1}$ would not be unreasonable. Thus the timescale for diffusion to homogenise spatial gradients of this protein within a cell is

$$T \sim \frac{L^2}{D} \sim \frac{10^{-6} \text{ cm}^2}{10^{-7} \text{ cm}^2 \text{ s}^{-1}} \sim 10 \text{ s}, \quad (5.14)$$

therefore we can often neglect diffusion in a cell. However, as the scale doubles the time scale squares *e.g.* $L \times 10 \Rightarrow T \times 100$ and $L \times 100 \Rightarrow T \times 10^4$.

Note. The above derivation generalises to situations more general than modelling chemical or biochemical diffusion. For example, let $I(x, y, t)$ denote the number of infected people per unit area. Assume the infectives, on average, spread out via a random walk mechanism and interact with susceptibles, as described in Section (6.2.1). One has that the flux of infectives, \mathbf{q}_I , is given by

$$\mathbf{q}_I = -D_I \nabla I, \quad (5.15)$$

where D_I is a constant, with dimensions of $(\text{length})^2 (\text{time})^{-1}$. Thus, one has, via precisely the same ideas and arguments as above, that

$$\frac{\partial I}{\partial t} = \nabla \cdot (D_I \nabla I) + rIS - aI, \quad (5.16)$$

where $S(x, y, t)$ is the number of susceptibles per unit area, and r and a have the same interpretation as in Section 6.2.1.

Fisher's Equation. A very common example is the combination of logistic growth and diffusion which, in one spatial dimension, gives rise to Fisher's Equation:

$$\frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} + ru \left(1 - \frac{u}{K}\right), \quad (5.17)$$

which was first proposed to model the spread of an advantageous gene through a population. See Section 6.1 for more details.

5.2 Chemotaxis

As briefly mentioned earlier, motile cells can move in response to gradients in chemical concentrations, a process known as chemotaxis. This leads to slightly more complicated transport equations, as we shall see.

The diffusive flux for the population density of the cells, n , is as previously: $\mathbf{J}_D = -D \nabla n$. The flux due to chemotaxis (assuming it is an attractant rather than a repellent) is taken to be of the form:

$$\mathbf{J}_C = n\chi(a)\nabla a = n\nabla\Phi(a), \quad (5.18)$$

where a is the chemical concentration and $\Phi(a)$ increases monotonically with a . Clearly $\chi(a) = \Phi'(a)$; the cells move in response to a gradient of the chemical in the direction in which the function $\Phi(a)$ is increasing at the fastest rate.

Thus the total flux is

$$\mathbf{J}_D + \mathbf{J}_C = -D \nabla n + n\chi(a)\nabla a. \quad (5.19)$$

Combining the transport of the motile cells, together with a term describing their reproduction and/or death, plus an equation for the chemical which also diffuses and, typically,

is secreted and degrades leads to the following equations

$$\frac{\partial n}{\partial t} = \nabla \cdot (D \nabla n) - \nabla \cdot (n \chi(a) \nabla a) + f(n, a), \quad (5.20)$$

$$\frac{\partial a}{\partial t} = \nabla \cdot (D_a \nabla a) + \lambda n - \mu a. \quad (5.21)$$

In the above the above $f(n, a)$ is often taken to be a logistic growth term while the function $\chi(a)$ describing the chemotaxis has many forms, including

$$\chi(a) = \frac{\chi_0}{a}, \quad (5.22)$$

$$\chi(a) = \frac{\chi_0}{(k + a)^2}, \quad (5.23)$$

where the latter represents a receptor law, with $\Phi(a)$ taking a Michaelis-Menten form [5].

Chapter 6

Travelling waves

Certain types of models can be seen to display wave-type behaviour. Here we will be interested in travelling waves, those that travel without change in shape and at constant speed.

References.

- J. D. Murray, Mathematical Biology Volume I, Chapter 13 [8].
- J. D. Murray, Mathematical Biology Volume II, Chapter 1 [9].
- N. F. Britton, Essential Mathematical Biology, Chapter 5 [1].

6.1 Fisher's equation: an investigation

Fisher's equation, after suitable non-dimensionalisation, is

$$\frac{\partial \beta}{\partial t} = \frac{\partial^2 \beta}{\partial z^2} + \beta(1 - \beta), \quad (6.1)$$

where β , z , t are all non-dimensionalised variables.

Clearly the solution of these equations will depend on the initial and boundary conditions we impose. We state these conditions for the time being as

$$\beta(z, t) \rightarrow \beta_{\pm\infty} \quad \text{as} \quad z \rightarrow \pm\infty \quad \text{and} \quad \beta(z, \tau = 0) = \beta_0(z), \quad (6.2)$$

where $\beta_{\pm\infty}$, β_0 , are constants.

6.1.1 Key points

- We will investigate whether such a wave solution exists for the above equations which propagates *without a change in shape* and at a constant (but as yet unknown) speed v . Such wave solutions are defined to be *travelling wave solutions*.

- The investigation of the potential existence of a travelling wave solution will be substantially easier to investigate on performing the transformation to the moving coordinate frame $y = z - v\tau$ as, by the definition of a travelling wave, the wave profile will be independent of time in a frame moving at speed v .
- Using the chain rule and noting that we seek a solution that is time independent with respect to the y variable, we have

$$\frac{\partial \beta}{\partial t} = \frac{\partial \beta}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial \beta}{\partial \tau} \frac{\partial \tau}{\partial t} \quad \text{and} \quad \frac{\partial \beta}{\partial z} = \frac{\partial \beta}{\partial y} \frac{\partial y}{\partial z} + \frac{\partial \beta}{\partial \tau} \frac{\partial \tau}{\partial z}, \quad (6.3)$$

i.e.

$$\frac{\partial}{\partial t} \mapsto -v \frac{\partial}{\partial y} + \frac{\partial}{\partial \tau} \quad \text{and} \quad \frac{\partial}{\partial z} \mapsto \frac{\partial}{\partial y}. \quad (6.4)$$

Assuming $\beta = \beta(y)$ so that $\partial \beta / \partial \tau = 0$ the partial differential equation, (6.1), reduces to

$$\beta'' + v\beta' + \beta(1 - \beta) = 0 \quad \text{where} \quad ' = \frac{d}{dy}. \quad (6.5)$$

- One must choose appropriate boundary conditions at $\pm\infty$ for the travelling wave equations. These are the same as the boundary conditions for the full partial differential equation (but rewritten in terms of y), *i.e.*

$$\beta(y) \rightarrow \beta_{\pm\infty} \quad \text{as} \quad y \rightarrow \pm\infty, \quad (6.6)$$

where $\beta_{\pm\infty}$, are the same constants as specified in equation (6.2).

- One must have that $\beta_{+\infty}, \beta_{-\infty}$ only take the values zero or unity:

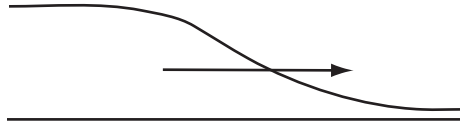
$$\int_{-\infty}^{\infty} [\beta'' + v\beta' + \beta(1 - \beta)] dy = 0, \quad (6.7)$$

gives

$$[\beta' + v\beta]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \beta(1 - \beta) dy = 0. \quad (6.8)$$

If we want $\beta \rightarrow \text{constant}$ as $y \rightarrow \pm\infty$ and β, β' finite for $\forall y$ we must have either $\beta \rightarrow 0$ or $\beta \rightarrow 1$ as $y \rightarrow \infty$ and similarly for $y \rightarrow -\infty$.

- With the boundary conditions $(\beta(-\infty), \beta(\infty)) = (1, 0)$, we physically anticipate $v > 0$.



Indeed there are no solutions of the Fisher travelling wave equations for these boundary conditions and $v \leq 0$.

- Solutions to equations (6.1) and (6.2) are unique. The proof would be an exercise in the theory of partial differential equations.
- The solutions of the travelling wave equations are not unique. One may have solutions for different values of the unknown v . Also, if $\beta(y)$ solves (6.5) for any fixed value of v then, for the same value of v , so does $\beta(y + A)$, where A is any constant. For both v and A fixed the solution of the travelling wave equations are normally unique.
- Note that the solutions of the travelling wave equations, (6.5), can only possibly be solutions of the full partial differential equation, when considered on an infinite domain. [Realistically one requires that the length scale of variation of the system in question is much less than the length scale of the physical domain for a travelling wave to (have the potential to) be an excellent approximation to the reaction-diffusion wave solutions on the physical, *i.e.* finite, domain].
- One “loses” the partial differential equation initial conditions associated with (6.1) and (6.2). The solution of the travelling wave equations given above for β are only a solution of the full partial differential equation, (6.1), for all time if the travelling wave solution is consistent with the initial conditions specified in (6.2).
- However, often (or rather usually!), one finds that for a particular choice of v the solutions of the full partial differential equation system, (6.1) and (6.2), tend, as $t \rightarrow \infty$, to a solution of the travelling wave equations (6.5), with fixed v and A , for a very large class of initial conditions.
- The Russian mathematician Kolmogorov proved that solutions of the full partial differential equation system, (6.1) and (6.2), do indeed tend, as $t \rightarrow \infty$, to a solution of the travelling wave equations for $v = 2$ for a large class of initial conditions.

6.1.2 Existence and the phase plane

We will investigate the existence of solutions of Fisher’s equation, equation (6.5), with the boundary conditions $(\beta(-\infty), \beta(\infty)) = (1, 0)$ and $v > 0$, by means of an extended exercise involving the phase plane (β', β) .

Consider the travelling wave equation

$$\frac{d^2\beta}{dy^2} + v\frac{d\beta}{dy} + \beta(1 - \beta) = 0, \quad (6.9)$$

with $v > 0$ and the boundary conditions $(\beta(-\infty), \beta(\infty)) = (1, 0)$.

Exercise 1. Show that the stationary point at $(\beta, \beta') = (1, 0)$ is always a saddle point and the stationary point at $(\beta, \beta') = (0, 0)$ is a stable node for $v \geq 2$ and a stable focus for $v < 2$.

Solution. Writing $\beta' = \gamma$ gives

$$\frac{d}{dy} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \frac{d}{dy} \begin{pmatrix} \beta \\ \beta' \end{pmatrix} = \begin{pmatrix} \gamma \\ -v\gamma - \beta(1 - \beta) \end{pmatrix}. \quad (6.10)$$

The Jacobian, \mathbf{J} , is given by

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f}{\partial \beta} & \frac{\partial f}{\partial \gamma} \\ \frac{\partial g}{\partial \beta} & \frac{\partial g}{\partial \gamma} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 + 2\beta & -v \end{pmatrix}. \quad (6.11)$$

At $(0,0)$ we have

$$\det(\mathbf{J} - \lambda \mathbf{I}) = \det \begin{pmatrix} -\lambda & 1 \\ -1 & -v - \lambda \end{pmatrix} \Rightarrow \lambda^2 + v\lambda + 1 = 0, \quad (6.12)$$

and hence

$$\lambda = \frac{-v \pm \sqrt{v^2 - 4}}{2}. \quad (6.13)$$

Therefore:

- if $v < 2$ we have $\lambda = -v/2 \pm i\mu$ and hence a stable spiral;
- if $v > 2$ we have $\lambda = -v/2 \pm \mu$ and hence a stable node;
- if $v = 2$ we have $\lambda = -1$ and hence a stable node.

At $(1,0)$ we have

$$\det(\mathbf{J} - \lambda \mathbf{I}) = \det \begin{pmatrix} -\lambda & 1 \\ 1 & -v - \lambda \end{pmatrix} \Rightarrow \lambda^2 + v\lambda - 1 = 0, \quad (6.14)$$

and hence

$$\lambda = \frac{-v \pm \sqrt{v^2 + 4}}{2}. \quad (6.15)$$

Therefore $(1,0)$ is a saddle point.

Exercise 2. Explain why solutions of Fisher's travelling wave equations must tend to phase plane stationary points as $y \rightarrow \pm\infty$. Hence explain why solutions of (6.9) with $v < 2$ are unphysical.

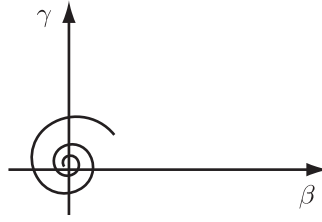
Solution. $(\beta, \gamma) \equiv (\beta, \beta')$ will change as y increases, unless at a stationary point. Therefore they will keep moving along a phase space trajectory as $y \rightarrow \infty$ unless the $y \rightarrow \infty$ limit evolves to a stationary point.

To satisfy $\lim_{y \rightarrow 0} \beta(y) = 0$, we need to be on a phase space trajectory which “stops” at $\beta = 0$. Therefore we must be on a phase space trajectory which tends to a stationary point with $\beta = 0$ as $y \rightarrow \infty$.

Hence we must tend to $(0,0)$ as $y \rightarrow \infty$ to satisfy $\lim_{y \rightarrow \infty} \beta(y) = 0$ as $y \rightarrow \infty$.

An analogous argument holds as $y \rightarrow -\infty$.

If $v < 2$ then $\beta < 0$ at some point on the trajectory which is unphysical:



Exercise 3. Show that the gradient of the unstable manifold at $(\beta, \beta') = (1, 0)$ is given by

$$\frac{1}{2} \left(-v + \sqrt{v^2 + 4} \right). \quad (6.16)$$

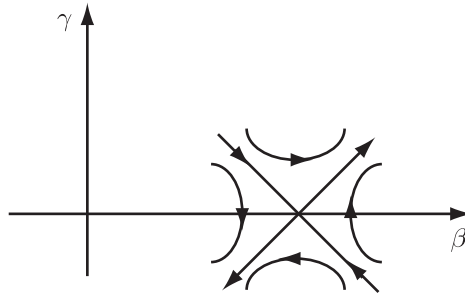
Sketch the qualitative form of the phase plane trajectories near to the stationary points for $v \geq 2$.

Solution. We require the eigenvectors of the Jacobian at $(1, 0)$:

$$\begin{pmatrix} 0 & 1 \\ 1 & -v \end{pmatrix} \begin{pmatrix} 1 \\ q_{\pm} \end{pmatrix} = \lambda_{\pm} \begin{pmatrix} 1 \\ q_{\pm} \end{pmatrix} \Rightarrow q_{\pm} = \lambda_{\pm} \quad \text{and} \quad 1 - vq_{\pm} = \lambda_{\pm}q_{\pm}. \quad (6.17)$$

Hence

$$\mathbf{v}_{\pm} = \begin{pmatrix} 1 \\ \frac{1}{2} \left[-v \pm \sqrt{v^2 + 4} \right] \end{pmatrix}. \quad (6.18)$$



Exercise 4. Explain why any physically relevant phase plane trajectory must leave $(\beta, \beta') = (1, 0)$ on the unstable manifold pointing in the direction of decreasing β .

Solution. Recall that, close to the stationary point,

$$\begin{pmatrix} \beta \\ \gamma \end{pmatrix} - \begin{pmatrix} \beta^* \\ \gamma^* \end{pmatrix} = a_- e^{\lambda_- y} \mathbf{v}_- + a_+ e^{\lambda_+ y} \mathbf{v}_+. \quad (6.19)$$

The solution moves away from the saddle along the unstable manifold, which corresponds to a_- .

Exercise 5. Consider $v \geq 2$. With $\gamma \stackrel{\text{def}}{=} \beta'$ show that for $\beta \in (0, 1]$ the phase plane trajectories, with gradient

$$\frac{d\gamma}{d\beta} = -v - \frac{\beta(1-\beta)}{\gamma}, \quad (6.20)$$

satisfy the constraint $d\gamma/d\beta < -1$ whenever $\gamma = -\beta$.

Solution.

$$\left. \frac{d\gamma}{d\beta} \right|_{\beta=-\gamma} = -v + (1-\beta) \leq (-v+2) - (1+\beta) < -1. \quad (6.21)$$

Exercise 6. Hence show that with $v \geq 2$ the unstable manifold leaving $(\beta, \beta') = (1, 0)$ and entering the region $\beta' < 0$, $\beta < 1$ enters, and can never leave, the region

$$\mathcal{R} \stackrel{\text{def}}{=} \{(\beta, \gamma) \mid \gamma \leq 0, \beta \in [0, 1], \gamma \geq \beta\}. \quad (6.22)$$

Solution. Along $\mathcal{L}_1 = \{(\beta, \gamma) \mid \gamma = 0, \beta \in (0, 1)\}$ the trajectories point vertically into \mathcal{R} as

$$\left| \frac{d\gamma}{d\beta} \right| \rightarrow \infty \quad \text{as we approach } \mathcal{L}_1 \text{ and } \gamma' = -\beta(1-\beta) < 0. \quad (6.23)$$

Along $\mathcal{L}_2 = \{(\beta, \gamma) \mid \beta = 1, \gamma \in (-1, 0)\}$ we have

$$\left. \frac{d\gamma}{d\beta} \right|_{\mathcal{L}_2} = -v - \frac{\beta(1-\beta)}{\gamma} = -v < 0. \quad (6.24)$$

Hence trajectories that enter \mathcal{R} cannot leave. There any trajectory must end at a stationary point and trajectories are forced to the point $(\beta, \gamma) = (0, 0)$.

Exercise 7. Thus prove that that there exists a monotonic solution, with $\beta \geq 0$, to equation (6.9) for every value of $v \geq 2$ and, with $v \geq 2$ fixed, the phase space trajectory is unique.

Solution. The above analysis is valid for $v \geq 2$. For v fixed a trajectory enters the region \mathcal{R} along the unstable manifold (only one unstable manifold enters \mathcal{R}). The solution is monotonic as $\gamma < 0$ throughout \mathcal{R} .

Figure 6.1 shows the results of numerical simulation of the Fisher equation (6.1) with initial and boundary conditions given by (6.2) at a series of time points.

6.1.3 Relation between the travelling wave speed and initial conditions

We have seen that, for v fixed, the phase space trajectory of Fisher's travelling wave equation is unique. The non-uniqueness associated with the fact that if $\beta(y)$ solves Fisher's travelling wave equation then so does $\beta(y + A)$ for A constant simply corresponds to a shift along the phase space trajectory. This, in turn, corresponds simply to translation of the travelling wave.

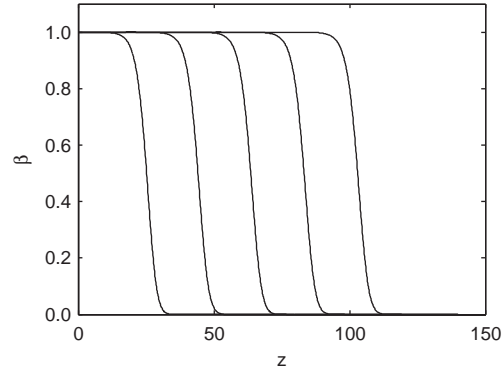


Figure 6.1: Solution of the Fisher equation (6.1) with initial and boundary conditions given by (6.2) at times $t = 10, 20, 30, 40, 50$.

Key question. Do solutions of the full system, equations (6.1) and (6.2), actually evolve to a travelling wave solution, and if so, what is its speed?

Non-Examinable: initial conditions of compact support

Kolmogorov considered the equation

$$\frac{\partial \psi}{\partial \tau} = \frac{\partial^2 \psi}{\partial z^2} + \psi(1 - \psi), \quad (6.25)$$

with the boundary conditions

$$\psi(z, \tau) \rightarrow 1 \quad \text{as } z \rightarrow -\infty \quad \text{and} \quad \psi(z, \tau) \rightarrow 0 \quad \text{as } z \rightarrow \infty, \quad (6.26)$$

and non-negative initial conditions satisfying the following: there is a K , with $0 < K < \infty$, such that

$$\psi(z, \tau = 0) = 0 \quad \text{for } z > K \quad \text{and} \quad \psi(z, \tau = 0) = 1 \quad \text{for } z < -K. \quad (6.27)$$

He proved that $\psi(z, \tau)$ tends to a Fisher travelling wave solution with $v = 2$ as $t \rightarrow \infty$.

This can be applied to equations (6.1) and (6.2) providing the initial conditions are non-negative and the initial condition for β satisfies the above constraint, *i.e.* there is a K , with $0 < K < \infty$, such that

$$\beta(z, \tau = 0) = 0 \quad \text{for } z > K \quad \text{and} \quad \beta(z, \tau = 0) = 1 \quad \text{for } z < -K. \quad (6.28)$$

Under such constraints β also tends to a Fisher travelling wave solution with $v = 2$.

6.2 Models of epidemics

The study of infectious diseases has a long history and there are numerous detailed models of a variety of epidemics and epizootics (*i.e.* animal epidemics). We can only possibly scratch the surface. In the following, we consider a simple, framework model but even this is capable of highlighting general comments about epidemics and, in fact, approximately describes some specific epidemics.

6.2.1 The SIR model

Consider a disease for which the population can be placed into three compartments:

- the susceptible compartment, S , who can catch the disease;
- the infective compartment, I , who have and transmit the disease;
- the removed compartment, R , who have been isolated, or who have recovered and are immune to the disease, or have died due to the disease during the course of the epidemic.

Assumptions

- The epidemic is of short duration course so that the population is constant (counting those who have died due to the disease during the course of the epidemic).
- The disease has a negligible incubation period.
- If a person contracts the disease and recovers, they are immune (and hence remain in the removed compartment).
- The numbers involved are sufficiently large to justify a continuum approximation.
- The ‘dynamics’ of the disease can be described by applying the Law of Mass Action to:



The model

Then the equations describing the time evolution of numbers in the susceptible, infective and removed compartments are given by

$$\frac{dS}{dt} = -rIS, \quad (6.30)$$

$$\frac{dI}{dt} = rIS - aI, \quad (6.31)$$

$$\frac{dR}{dt} = aI, \quad (6.32)$$

subject to

$$S(t=0) = S_0, \quad I(t=0) = I_0, \quad R(t=0) = 0. \quad (6.33)$$

Note that

$$\frac{d}{dt}(S + I + R) = 0 \implies S + I + R = S_0 + I_0. \quad (6.34)$$

Key questions in an epidemic situation are, given r , a , S_0 and I_0 ,

1. Will the disease spread, *i.e.* will the number of infectives increase, at least in the short-term?

Solution.

$$\frac{dS}{dt} = -rIS \Rightarrow S \text{ is decreasing and therefore } S \leq S_0. \quad (6.35)$$

$$\frac{dI}{dt} = I(rS - a) < I(rS_0 - a). \quad (6.36)$$

Therefore, if $S_0 < a/r$ the infectives never increase, at least initially.

2. If the disease spreads, what will be the maximum number of infectives at any given time?

Solution.

$$\frac{dI}{dS} = -\frac{(rS - a)}{rS} = -1 + \frac{\rho}{S} \quad \text{where } \rho \stackrel{\text{def}}{=} \frac{a}{r}. \quad (6.37)$$

Integrating gives

$$I + S - \rho \ln S = I_0 + S_0 - \rho \ln S_0, \quad (6.38)$$

and so, noting that $dI/dS = 0$ for $S = \rho$, the maximum number of infectives is given by

$$I_{max} = \begin{cases} I_0 & S_0 \leq \rho \\ I_0 + S_0 - \rho \ln S_0 - \rho \ln \rho - \rho & S_0 > \rho \end{cases}. \quad (6.39)$$

3. How many people in total catch the disease?

Solution. From 2, $I \rightarrow 0$ as $t \rightarrow \infty$. Therefore the total number who catch the disease is

$$R(\infty) = N_0 - S(\infty) - I(\infty) = N_0 - S(\infty), \quad (6.40)$$

where $S(\infty) < S_0$ is the root of

$$S_\infty - \rho \ln S_\infty = N_0 - \rho \ln S_0, \quad (6.41)$$

obtained by setting $S = S_\infty$ and $N_0 = I_0 + S_0$ in equation (6.38).

6.2.2 An SIR model with spatial heterogeneity

We consider an application to fox rabies. We will make the same assumptions as for the standard SIR model, plus:

- healthy, *i.e.* susceptible, foxes are territorial and, on average, do not move from their territories;
- rabid, *i.e.* infective, foxes undergo behavioural changes and migrate randomly, with an effective, constant, diffusion coefficient D ;
- rabies is fatal, so that infected foxes do not return to the susceptible compartment but die, and hence the removed compartment does not migrate.

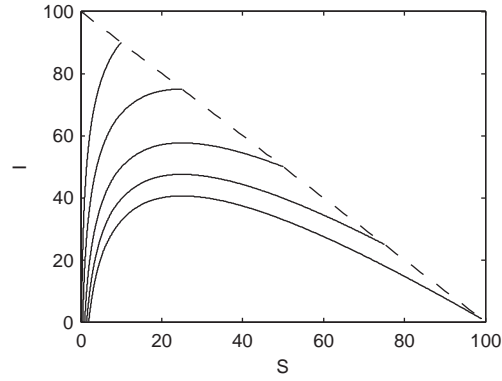


Figure 6.2: Numerical solution of the SIR model, equations (6.30)-(6.32), where the solid lines indicate the phase trajectories and the dashed line $S + I = S_0 + I_0$. Parameters are as follows: $r = 0.01$ and $a = 0.25$.

Taking into account rabid foxes' random motion, the SIR equations become

$$\frac{\partial S}{\partial t} = -rIS, \quad (6.42)$$

$$\frac{\partial I}{\partial t} = D\nabla^2 I + rIS - aI, \quad (6.43)$$

$$\frac{\partial R}{\partial t} = aI. \quad (6.44)$$

The I and S equations decouple, and we consider these in more detail. We assume a one-dimensional spatial domain $x \in (-\infty, \infty)$ and apply the following scalings/non-dimensionalisations,

$$I_* = \frac{I}{S_0}, \quad S_* = \frac{S}{S_0}, \quad x_* = \sqrt{\frac{D}{rS_0}}x, \quad t_* = rS_0t, \quad \lambda = \frac{a}{rS_0}, \quad (6.45)$$

where S_0 is the population density in the absence of rabies, to obtain

$$\frac{\partial S}{\partial t} = -IS, \quad (6.46)$$

$$\frac{\partial I}{\partial t} = \nabla^2 I + I(S - \lambda), \quad (6.47)$$

where asterisks have been dropped for convenience in the final expression.

Travelling waves

We seek travelling wave solutions with

$$S(x, t) = S(y), \quad I(x, t) = I(y), \quad y = x - ct, \quad c > 0, \quad (6.48)$$

which results in the system

$$0 = cS' - IS, \quad (6.49)$$

$$0 = I'' + cI' + I(S - \lambda), \quad (6.50)$$

where $' = d/dy$.

We assume $\lambda = a/(rS_0) < 1$ below. This is equivalent to the condition for disease spread in the earlier SIR model.

Boundary conditions

We assume a healthy population as $y \rightarrow \infty$:

$$S \rightarrow 1 \quad \text{and} \quad I \rightarrow 0, \quad (6.51)$$

and as $y \rightarrow -\infty$ we require

$$I \rightarrow 0. \quad (6.52)$$

Bound on travelling wave speed

We write $S = 1 - P$ and linearise about the wavefront:

$$-cP' - I = 0 \quad \text{and} \quad I'' + cI' + I(1 - \lambda). \quad (6.53)$$

The I equation decouples and analysis of this equation gives a stable focus at $(I, I') = (0, 0)$ if the eigenvalues

$$\mu = \frac{-c \pm \sqrt{c^2 - 4(1 - \lambda)}}{2}, \quad (6.54)$$

are complex. This requires

$$c \geq 2\sqrt{1 - \lambda}. \quad (6.55)$$

Severity of epidemic

$S(\infty)$ is a measure of the severity of the epidemic. We have $I = cS'/S$ and therefore

$$\frac{d}{dy}(I' + cI) + cS' \left(\frac{S - \lambda}{S} \right) = 0. \quad (6.56)$$

Therefore

$$(I' + cI) + c(S - \lambda \ln S) = \text{constant} = c, \quad (6.57)$$

by evaluating the equation as $y \rightarrow \infty$.

In this case

$$S(-\infty) - \lambda \ln S(-\infty) = 1, \quad \text{where} \quad S(-\infty) < 1, \quad (6.58)$$

gives the severity of the epidemic.

Further comments on travelling wave speed

Typically, the wave evolves to have minimum wave speed:

$$c \simeq c_{min}. \quad (6.59)$$

Chapter 7

Pattern formation

Examples of the importance of spatial pattern and structure can be seen just about everywhere in the natural world. Here we will be concerned with building and analysing models which can generate patterns; understanding how self-organising principles may lead to the generation of shape and form.

References.

- J. D. Murray, Mathematical Biology Volume II, Chapter 2 and Chapter 3 [9].
- N. F. Britton, Essential Mathematical Biology, Chapter 7 [1].

7.1 Minimum domains for spatial structure

Consider the non-dimensionalised, one dimensional, budworm model, but with a diffusive spatial structure:

$$u_t = Du_{xx} + f(u), \quad \text{where} \quad f(u) = ru \left(1 - \frac{u}{q}\right) - \frac{u^2}{1 + u^2}. \quad (7.1)$$

We also suppose that exterior to our domain conditions are very hostile to budworm so that we have the boundary conditions

$$u(0, t) = 0, \quad u(L, t) = 0, \quad (7.2)$$

where $L > 0$ is the size of the domain. Note that $f'(0) = r > 0$.

Question. Clearly $u = 0$ is a solution. However, if we start with a small initial distribution of budworm, will we end up with no budworm, or an outbreak of budworm? In particular, how does this depend on the domain size?

Solution. For initial conditions with $0 \leq u(x, t = 0) \ll 1$, sufficiently small, we can approximate $f(u)$ by $f'(0)u$ at least while $u(x, t)$ remains small. Thus our equations are, approximately,

$$u_t = Du_{xx} + f'(0)u, \quad u(0, t) = 0, \quad u(L, t) = 0. \quad (7.3)$$

We look for a solution of the form (invoking completeness of Fourier series):

$$u(x, t) = \sum_{n=1}^{\infty} a_n(t) \sin\left(\frac{n\pi x}{L}\right). \quad (7.4)$$

This gives that the time-dependent coefficients satisfy

$$\frac{da_n}{dt} = \frac{Dn^2\pi^2}{L^2}a_n + f'(0)a_n = \sigma_n a_n, \quad (7.5)$$

and hence

$$u(x, t) = \sum_{n=1}^{\infty} a_n^{(0)} \exp\left[\left(f'(0) - \frac{Dn^2\pi^2}{L^2}\right)t\right] \sin\left(\frac{n\pi x}{L}\right). \quad (7.6)$$

For the solution to decay to zero, we require that all Fourier modes decay to zero as $t \rightarrow \infty$, and hence we require that

$$\sigma_n < 0 \quad \forall n \quad \Rightarrow \quad f'(0) - \frac{Dn^2\pi^2}{L^2} < 0 \quad \forall n, \quad (7.7)$$

and thus that

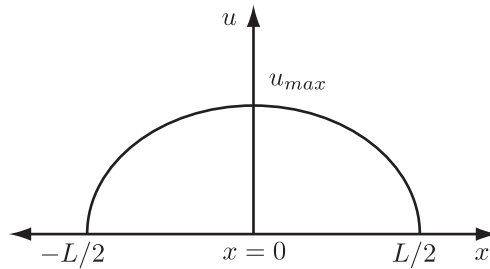
$$f'(0) < \frac{Dn^2\pi^2}{L^2} \quad \Rightarrow \quad L \leq \left[\frac{D\pi^2}{f'(0)}\right] \stackrel{\text{def}}{=} L_{crit}. \quad (7.8)$$

Hence there is a critical lengthscale, L_{crit} , beyond which an outburst of budworm is possible in a spatially distributed system.

7.1.1 Domain size

On first inspection one probably should be surprised to see that L_{crit} increases linearly with the diffusion coefficient, *i.e.* diffusion is destabilising the zero steady state.

We can further investigate how the nature of a steady state pattern depends on the diffusion coefficient. Suppose $L > L_{crit}$ and that the steady state pattern is of the form:



We therefore have

$$0 = Du_{xx} + f(u). \quad (7.9)$$

Multiplying by u_x and integrating with respect to x , we have

$$0 = \int Du_x u_{xx} dx + \int u_x f(u) dx. \quad (7.10)$$

Thus we have

$$\frac{1}{2}Du_x^2 + F(u) = \text{constant} = F(u_{max}) \quad \text{where} \quad F'(u) = f(u). \quad (7.11)$$

We can therefore find a relation between L , D , integrals of

$$F(u) \stackrel{\text{def}}{=} \int_0^u f(y) dy, \quad (7.12)$$

and $\max(u)$, the size of the outbreak, as follows:

$$u_x = -\left(\frac{2}{D}\right)^{\frac{1}{2}} \sqrt{F(u_{max}) - F(u)} \quad \text{since } x > 0 \text{ and therefore } u_x < 0. \quad (7.13)$$

Integrating, gives

$$2 \int_0^{L/2} dx = -(2D)^{\frac{1}{2}} \int_{u_{max}}^0 \frac{1}{\sqrt{F(u_{max}) - F(\bar{u})}} d\bar{u}, \quad (7.14)$$

and hence

$$L = (2D)^{\frac{1}{2}} \int_0^{u_{max}} \frac{1}{\sqrt{F(u_{max}) - F(\bar{u})}} d\bar{u}. \quad (7.15)$$

Therefore u_{max} is a function of $L/\sqrt{2D}$ and the root of equation (7.15).

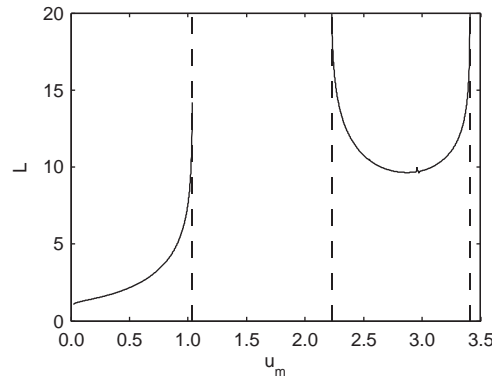


Figure 7.1: Numerical simulation of the u_m - L space, equation (7.15) with $r = 0.6$, $q = 6.2$ and $D = 0.1$.

7.2 Diffusion-driven instability

Consider a two component system

$$u_t = D_u \nabla^2 u + f(u, v), \quad (7.16)$$

$$v_t = D_v \nabla^2 v + g(u, v), \quad (7.17)$$

for $\mathbf{x} \in \Omega$, $t \in [0, \infty)$ and Ω bounded.

The initial conditions are

$$u(x, 0) = u_0(x), \quad v(x, 0) = v_0(x), \quad (7.18)$$

and the boundary conditions are either Dirichlet, *i.e.*

$$u = u_B, \quad v = v_B, \quad \mathbf{x} \in \partial\Omega, \quad (7.19)$$

or homogeneous Neumann, *i.e.*

$$\mathbf{n} \cdot \nabla u = 0, \quad \mathbf{n} \cdot \nabla v = 0, \quad \text{for } \mathbf{x} \in \partial\Omega, \quad (7.20)$$

where \mathbf{n} is the outward pointing normal on $\partial\Omega$.

Definition. *Patterns* are stable, time-independent, spatially heterogeneous solutions of equations (7.16)-(7.17).

Definition. A *diffusion-driven instability*, also referred to as a *Turing instability*, occurs when a steady state, stable in the absence of diffusion, goes unstable when diffusion is present.

Remark. Diffusion-driven instabilities, in particular, can drive pattern formation in chemical systems and there is significant, but not necessarily conclusive, evidence that it can drive pattern formation in a variety of biological systems. A key point is that this mechanism can drive the system from close to a homogeneous steady state to a state with spatial pattern and structure. The fact that diffusion is responsible for this is initially quite surprisingly. Diffusion, in isolation, disperses a pattern; yet diffusion, when in combination with the kinetic terms, often can drive a system towards a state with spatial structure.

7.2.1 Linear analysis

We wish to understand when a diffusion-driven instability occurs. Using vector and matrix notation we define

$$\mathbf{u} = \begin{pmatrix} u \\ v \end{pmatrix}, \quad \mathbf{F}(\mathbf{u}) = \begin{pmatrix} f(u, v) \\ g(u, v) \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} D_u & 0 \\ 0 & D_v \end{pmatrix}, \quad (7.21)$$

and write the problem with homogeneous Neumann boundary conditions as follows:

$$\mathbf{u}_t = \mathbf{D} \nabla^2 \mathbf{u} + \mathbf{F}(\mathbf{u}), \quad (7.22)$$

i.e.

$$\frac{\partial}{\partial t} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} D_u & 0 \\ 0 & D_v \end{pmatrix} \nabla^2 \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} f(u, v) \\ g(u, v) \end{pmatrix}, \quad (7.23)$$

with

$$\mathbf{n} \cdot \nabla \mathbf{u} = 0, \quad \mathbf{x} \in \partial\Omega, \quad (7.24)$$

i.e.

$$\mathbf{n} \cdot \nabla u = 0 = \mathbf{n} \cdot \nabla v \quad \mathbf{x} \in \partial\Omega. \quad (7.25)$$

Let \mathbf{u}^* be such that $\mathbf{F}(\mathbf{u}^*) = \mathbf{0}$. Implicit in this definition is the assumption that \mathbf{u}^* is a constant vector.

Let $\mathbf{w} = \mathbf{u} - \mathbf{u}^*$ with $|\mathbf{w}| \ll 1$. Then we have

$$\frac{\partial \mathbf{w}}{\partial t} = D \nabla^2 \mathbf{w} + \mathbf{F}(\mathbf{u}^*) + \mathbf{J} \mathbf{w} + \text{higher order terms}, \quad (7.26)$$

where

$$\mathbf{J} = \left(\begin{array}{cc} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{array} \right) \bigg|_{\mathbf{u}=\mathbf{u}^*}, \quad (7.27)$$

is the Jacobian of \mathbf{F} evaluated at $\mathbf{u} = \mathbf{u}^*$. Note that \mathbf{J} is a *constant* matrix.

Neglecting higher order terms in $|\mathbf{w}|$, we have the equation

$$\mathbf{w}_t = D \nabla^2 \mathbf{w} + \mathbf{J} \mathbf{w}, \quad \mathbf{n} \cdot \nabla \mathbf{w} = 0, \quad \mathbf{x} \in \partial\Omega. \quad (7.28)$$

This is a linear equation and so we look for a solution in the form of a linear sum of separable solutions. To do this, we first need to consider a general separable solution given by

$$\mathbf{w}(\mathbf{x}, t) = A(t) \mathbf{p}(\mathbf{x}), \quad (7.29)$$

where $A(t)$ is a scalar function of time. Substituting this into equation (7.28) yields

$$\frac{1}{A} \frac{dA}{dt} \mathbf{p} = D \nabla^2 \mathbf{p} + \mathbf{J} \mathbf{p}. \quad (7.30)$$

Clearly to proceed, with \mathbf{p} dependent on \mathbf{x} only, we require \dot{A}/A to be time independent. It must also be independent of \mathbf{x} as A is a function of time only. Thus \dot{A}/A is constant.

We take $\dot{A} = \lambda A$, where λ is as yet an undetermined constant. Thus

$$A = A_0 \exp(\lambda t), \quad (7.31)$$

for $A_0 \neq 0$ constant. Hence we require that our separable solution is such that

$$[\lambda \mathbf{p} - \mathbf{J} \mathbf{p} - D \nabla^2 \mathbf{p}] = 0. \quad (7.32)$$

Suppose \mathbf{p} satisfies the equation

$$\nabla^2 \mathbf{p} + k^2 \mathbf{p} = 0, \quad \mathbf{n} \cdot \nabla \mathbf{p} = 0, \quad \mathbf{x} \in \partial\Omega, \quad (7.33)$$

where $k \in \mathbb{R}$. This is motivated by the fact in one-dimensional on a bounded domain, we have $p'' + k^2 p = 0$; the solutions are trigonometric functions which means one immediately has a Fourier series when writing the sum of separable solutions.

Then we have

$$[\lambda \mathbf{p} - \mathbf{J}\mathbf{p} + \mathbf{D}k^2\mathbf{p}] = 0, \quad (7.34)$$

and thus

$$[\lambda \mathbf{I} - \mathbf{J} + \mathbf{D}k^2] \mathbf{p} = 0, \quad (7.35)$$

with $|\mathbf{p}|$ not identically zero. Hence

$$\det [\lambda \mathbf{I} - \mathbf{J} + k^2 \mathbf{D}] = 0. \quad (7.36)$$

This can be rewritten as

$$\det \begin{pmatrix} \lambda - f_u + D_u k^2 & -f_v \\ -g_u & \lambda - g_v + D_v k^2 \end{pmatrix} = 0, \quad (7.37)$$

which gives the following quadratic in λ :

$$\lambda^2 + [(D_u + D_v)k^2 - (f_u + g_v)] \lambda + h(k^2) = 0, \quad (7.38)$$

where

$$h(k^2) = D_u D_v k^4 - (D_v f_u + D_u g_v) k^2 + (f_u g_v - g_u f_v). \quad (7.39)$$

Note 1. Fixing model parameters and functions (*i.e.* fixing D_u , D_v , f , g), we have an equation which gives λ as a function of k^2 .

Note 2. Thus, for any k^2 such that equation (7.33) possesses a solution, denoted $\mathbf{p}_k(\mathbf{x})$ below, we can find a $\lambda = \lambda(k^2)$ and hence a general separable solution of the form

$$A_0 e^{\lambda(k^2)t} \mathbf{p}_k(\mathbf{x}). \quad (7.40)$$

The most general solution formed by the sum of separable solutions is therefore

$$\sum_{k^2} A_0(k^2) e^{\lambda(k^2)t} \mathbf{p}_k(\mathbf{x}), \quad (7.41)$$

if there are countable k^2 for which equation (7.33) possesses a solution. Otherwise the general solution formed by the sum of separable solutions is of the form

$$\int A_0(k^2) e^{\lambda(k^2)t} \mathbf{p}_{k^2}(\mathbf{x}) dk^2, \quad (7.42)$$

where k^2 is the integration variable.

Unstable points

If, for any k^2 such that equation (7.33) possesses a solution, we find $\text{Re}(\lambda(k^2)) > 0$ then:

- \mathbf{u}^* is (linearly) unstable and perturbations from the stationary state will grow;
- while the perturbations are small, the linear analysis remains valid; thus the perturbations keep growing until the linear analysis is invalid and the full non-linear dynamics comes into play;

- a small perturbation from the steady state develops into a growing spatially heterogeneous solution which subsequently seeds spatially heterogeneous behaviour of the full non-linear model;
- a spatially heterogeneous pattern can emerge from the system from a starting point which is homogeneous to a very good approximation.

Stable points

If, for all k^2 such that equation (7.33) possesses a solution, we find $Re(\lambda(k^2)) < 0$ then:

- \mathbf{u}^* is (linearly) stable and perturbations from the stationary state do not grow;
- patterning will not emerge from perturbing the homogeneous steady state solution \mathbf{u}^* ;
- the solution will decay back to the homogeneous solution¹.

7.3 Detailed study of the conditions for a Turing instability

For a Turing instability we require the homogeneous steady state to be stable without diffusion and unstable with diffusion present. Here we analyse the requirements for each of these conditions to be satisfied.

7.3.1 Stability without diffusion

We firstly require that in the absence of diffusion the system is stable. This is equivalent to

$$Re(\lambda(0)) < 0, \quad (7.43)$$

for *all* solutions of $\lambda(0)$, as setting $k^2 = 0$ removes the diffusion-driven term in equation (7.36) and the preceding equations.

We have that $\lambda(0)$ satisfies

$$\lambda(0)^2 - [f_u + g_v] \lambda(0) + [f_u g_v - f_v g_u] = 0. \quad (7.44)$$

Insisting that $Re(\lambda(0) < 0)$ gives us the conditions

$$f_u + g_v < 0 \quad (7.45)$$

$$f_u g_v - f_v g_u > 0. \quad (7.46)$$

¹Technical point: Strictly, this conclusion requires completeness of the separable solutions. This can be readily shown in 1D on bounded domains. (Solutions of $p'' + k^2 p = 0$ on bounded domains with Neumann conditions are trigonometric functions and completeness is inherited from the completeness of Fourier series). Even if completeness of the separable solutions is not clear, numerical simulations of the full equations are highly indicative and do not, for the models typically encountered, contradict the linear analysis results. With enough effort and neglecting any biological constraints on model parameters and functions, one may well be able to find D_u , D_v , f , g where there was such a discrepancy but that is not the point of biological modelling.

The simplest way of deducing (7.45) and (7.46) is by brute force.

The roots of the quadratic are given by

$$\lambda(0)_{\pm} = \frac{(f_u + g_v) \pm \sqrt{(f_u + g_v)^2 - 4(f_u g_v - f_v g_u)}}{2}. \quad (7.47)$$

7.3.2 Instability with diffusion

Now consider the effects of diffusion. In addition to $Re(\lambda(0)) < 0$, we are required to show, for diffusion-driven instability, that there exists k^2 such that

$$Re(\lambda(k^2)) > 0, \quad (7.48)$$

so that diffusion does indeed drive an instability.

We have that $\lambda(k^2)$ satisfies

$$\lambda^2 + [(D_u + D_v)k^2 - (f_u + g_v)]\lambda + h(k^2) = 0, \quad (7.49)$$

where

$$h(k^2) = D_u D_v k^4 - (D_v f_u + D_u g_v)k^2 + (f_u g_v - g_u f_v), \quad (7.50)$$

and

$$\alpha = (f_u + g_v) - (D_u + D_v)k^2 < 0. \quad (7.51)$$

Thus $Re(\lambda(k^2)) > 0$ requires that

$$Re\left(\alpha \pm \sqrt{\alpha^2 - 4h(k^2)}\right) > 0 \quad \Rightarrow \quad h(k^2) < 0. \quad (7.52)$$

Hence we must find k^2 such that

$$h(k^2) = D_u D_v k^4 - (D_v f_u + D_u g_v)k^2 + (f_u g_v - g_u f_v) < 0, \quad (7.53)$$

so that we have $k^2 \in [k_-^2, k_+^2]$ where $h(k_{\pm}^2) = 0$. Figure 7.2 shows a plot of a caricature $h(k^2)$.

This gives us that we have an instability whenever

$$k^2 \in \left[\frac{A - \sqrt{A^2 - B}}{2D_u D_v}, \frac{A + \sqrt{A^2 - B}}{2D_u D_v} \right] = [k_-^2, k_+^2], \quad (7.54)$$

where

$$A = D_v f_u + D_u g_v \quad \text{and} \quad B = 4D_u D_v (f_u g_v - g_u f_v) > 0, \quad (7.55)$$

and there exists a solution of the following

$$\nabla^2 \mathbf{p} + k^2 \mathbf{p} = 0, \quad \mathbf{n} \cdot \nabla \mathbf{p} = 0, \quad \mathbf{x} \in \partial\Omega, \quad (7.56)$$

for k^2 in the above range.

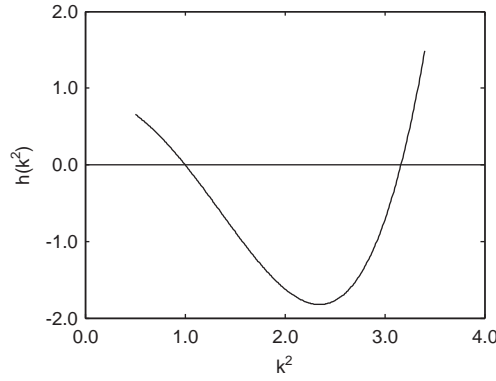


Figure 7.2: A plot of a caricature $h(k^2)$.

Insisting that k is real and non-zero (we have considered the $k = 0$ case above) we have

$$A > 0 \quad \text{and} \quad A^2 - B > 0, \quad (7.57)$$

which gives us that when $\text{Re}(\lambda(k^2)) > 0$, the following conditions hold:

$$A > 0 : \quad D_v f_u + D_u g_v > 0, \quad (7.58)$$

$$A^2 - B > 0 : \quad (D_v f_u + D_u g_v) > 2\sqrt{D_u D_v (f_u g_v - f_v g_u)}. \quad (7.59)$$

7.3.3 Summary

We have found that a diffusion-driven instability can occur when conditions (7.45), (7.46), (7.58), (7.59) hold whereupon the separable solutions, with k^2 within the range (7.54) and for which there is a solution to equation (7.33), will drive the instability.

Key point 1. Note that constraints (7.45) and (7.58) immediately gives us that $D_u \neq D_v$. Thus one cannot have a diffusion-driven instability with *identical* diffusion coefficients.

Key point 2. From constraints (7.45), (7.46), (7.58) the signs of f_u , g_v must be such that \mathbf{J} takes the form

$$\mathbf{J} = \begin{pmatrix} + & + \\ - & - \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} + & - \\ + & - \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} - & - \\ + & + \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} - & + \\ - & + \end{pmatrix}. \quad (7.60)$$

Key point 3. A Turing instability typically occurs via *long-range inhibition, short-range activation*. In more detail, suppose

$$\mathbf{J} = \begin{pmatrix} + & - \\ + & - \end{pmatrix}. \quad (7.61)$$

Then we have $f_u > 0$ and $g_v < 0$ by the signs of \mathbf{J} . In this case $D_v f_u + D_u g_v > 0 \Rightarrow D_v > D_u$. Hence the activator has a lower diffusion coefficient and spreads less quickly than the inhibitor.

7.3.4 The threshold of a Turing instability.

The threshold is defined such that equation (7.39), *i.e.*

$$D_u D_v k_c^4 - (D_v f_u + D_u g_v) k_c^2 + (f_u g_v - g_u f_v) = 0, \quad (7.62)$$

has a single root, k_c^2 .

Thus we additionally require

$$A^2 = B \quad \text{i.e.} \quad (D_v f_u + D_u g_v)^2 = 4 D_u D_v (f_u g_v - g_u f_v) > 0, \quad (7.63)$$

whereupon

$$k_c^2 = \frac{A}{2 D_u D_v} = \frac{D_v f_u + D_u g_v}{2 D_u D_v}. \quad (7.64)$$

Strictly one also requires that a solution exists for

$$\nabla^2 \mathbf{p} + k^2 \mathbf{p} = 0, \quad \mathbf{n} \cdot \nabla \mathbf{p} = 0, \quad \mathbf{x} \in \partial\Omega, \quad (7.65)$$

when $k^2 = k_c^2$. However, the above value of k_c^2 is typically an excellent approximation.

7.4 Extended example 1

Consider the one-dimensional case:

$$u_t = D_u u_{xx} + f(u, v), \quad (7.66)$$

$$v_t = D_v v_{xx} + g(u, v), \quad (7.67)$$

for $x \in [0, L]$, $t \in [0, \infty)$ and zero flux boundary conditions at $x = 0$ and $x = L$.

The analogue of

$$\nabla^2 \mathbf{p} + k^2 \mathbf{p} = 0, \quad \mathbf{n} \cdot \nabla \mathbf{p} = 0, \quad \mathbf{x} \in \partial\Omega, \quad (7.68)$$

is

$$p_{xx} + k^2 p = 0, \quad p'(0) = p'(L) = 0, \quad (7.69)$$

which gives us that

$$p_k(x) = A_k \cos(kx), \quad k = \frac{n\pi}{L}, \quad n \in \{1, 2, \dots\}, \quad (7.70)$$

where A_k is k -dependent in general but independent of t and x .

Thus the separable solution is of the form

$$\sum_k A_k e^{\lambda(k^2)t} \cos(kx), \quad (7.71)$$

where the sum is over the allowed values of k *i.e.*

$$k = \frac{n\pi}{L}, \quad n \in \{1, 2, \dots\}. \quad (7.72)$$

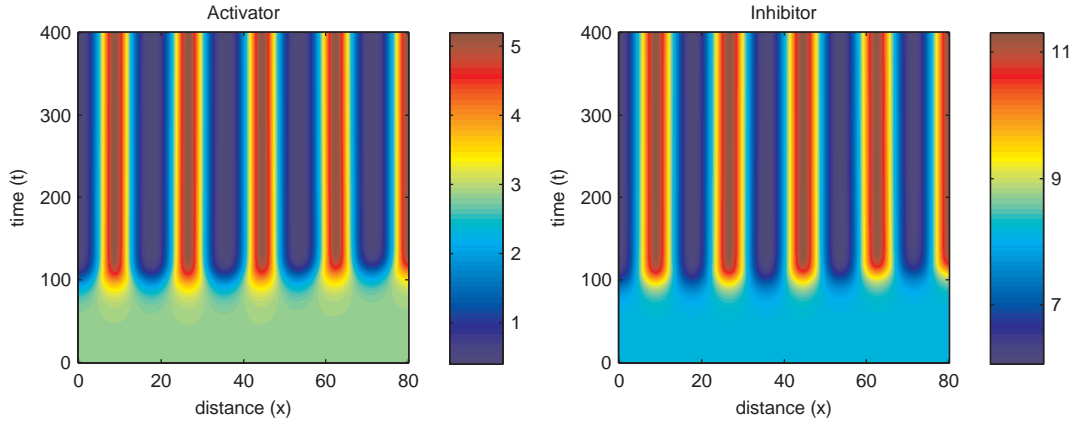


Figure 7.3: Numerical simulation of the Gierer-Meinhardt model for pattern formation.

7.4.1 The influence of domain size

If the smallest allowed value of $k^2 = \pi^2/L^2$ is such that

$$k^2 = \frac{\pi^2}{L^2} > \frac{A + \sqrt{A^2 - B}}{2D_u D_v} = k_+^2, \quad (7.73)$$

then we cannot have a Turing instability.

Thus for very small domains there is *no* pattern formation via a Turing mechanism. However, if one slowly increases the size of the domain, then L increases and the above constraint eventually breaks down and the homogeneous steady state destabilises leading to spatial heterogeneity.

This pattern formation mechanism has been observed in chemical systems. It is regularly hypothesised to be present in biological systems (*e.g.* animal coat markings, fish markings, the interaction of gene products at a cellular level, the formation of ecological patchiness) though the evidence is not conclusive at the moment.

7.5 Extended example 2

Consider the two-dimensional case with spatial coordinates $\mathbf{x} = (x, y)^T$, $x \in [0, a]$, $y \in [0, b]$, and zero flux boundary conditions. We find that the allowed values of k^2 are

$$k_{m,n}^2 = \left[\frac{m^2 \pi^2}{a^2} + \frac{n^2 \pi^2}{b^2} \right], \quad (7.74)$$

with

$$p_{m,n}(\mathbf{x}) = A_{m,n} \cos\left(\frac{m\pi x}{a}\right) \cos\left(\frac{n\pi y}{b}\right), \quad n, m \in \{0, 1, 2, \dots\}, \quad (7.75)$$

excluding the case where n, m are both zero.

Suppose the domain is long and thin, $b \ll a$. We may have a Turing instability if

$$k_{m,n}^2 = \left[\frac{m^2 \pi^2}{a^2} + \frac{n^2 \pi^2}{b^2} \right] \in [k_-^2, k_+^2] \quad \text{where} \quad h(k_{\pm}^2) = 0. \quad (7.76)$$

For b sufficiently small, this requires $n = 0$ and therefore no spatial variation in the y direction.

This means we have that the seed for pattern formation predicted by the linear analysis is a separable solution which is “stripes”; this typically invokes a striped pattern once the non-linear dynamics sets in.

For a large rectangular domain, $b \sim a$ sufficiently large, it is clear that a Turing instability can be initiated with $n, m > 0$. This means we have that the seed for pattern formation predicted by the linear analysis is a separable solution which is “spots”. This typically invokes a spotted pattern once the non-linear dynamics sets in.

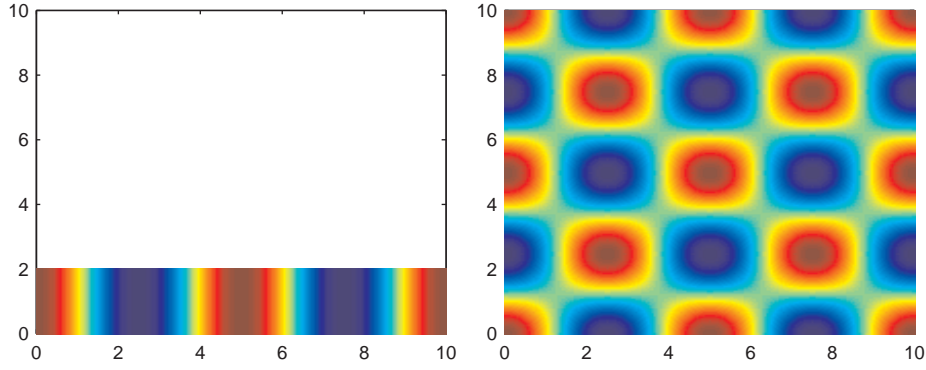
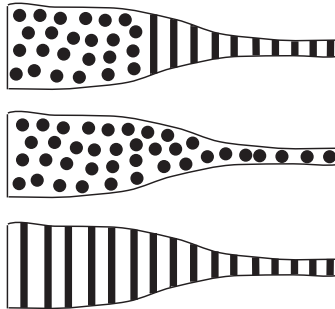


Figure 7.4: Changes in patterning as the domain shape changes.

Figure 7.4 shows how domain size may affect the patterns formed. On the left-hand side the domain is long and thin and only a striped pattern results, whilst the on the right-hand side the domain is large enough to admit patterning in both directions.

Suppose we have a domain which changes its aspect ratio from rectangular to long and thin. Then we have the following possibilities:



This leads to an interesting prediction, in the context of animal coat markings, that if it is indeed driven by a Turing instability, then one should not expect to see an animal with a striped body and a spotted tail.

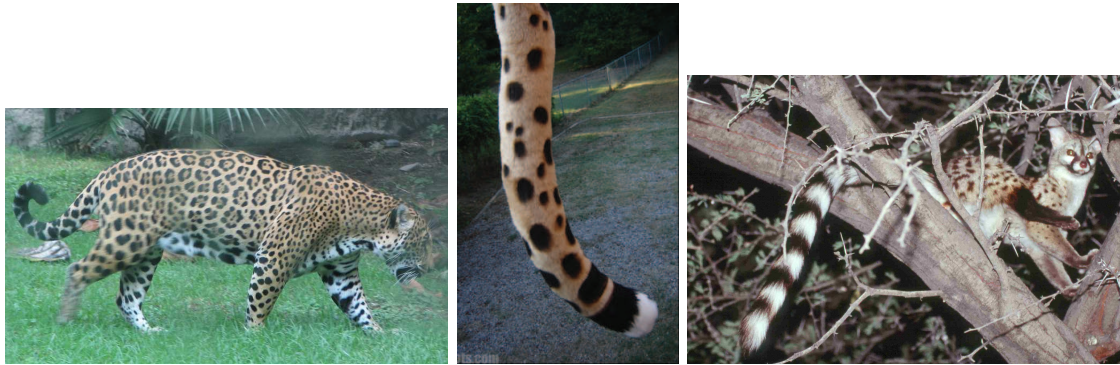


Figure 7.5: Animal coat markings which are consistent with the predictions of pattern formation by a Turing instability.

Common observation is consistent with such a prediction (see Figure 7.5) but one should not expect universal laws in the realms of biology as one does in physics (see Figure 7.6). More generally, this analysis has applications in modelling numerous chemical and biochemical reactions, in vibrating plate theory, and studies of patchiness in ecology and modelling gene interactions.



Figure 7.6: Animal coat markings which are inconsistent with the predictions of pattern formation by a Turing instability.

Chapter 8

Excitable systems: nerve pulses

Many cells communicate with one another via nerve impulses, also known as action potentials. Action potentials are brief changes in the membrane potential of a cell produced by the flow of ionic current across the cell membrane. Such type of communication is not limited to neurons but can also occur in other cells, for example cardiac and muscle cells.

See http://en.wikipedia.org/wiki/Action_potential and related links for more details.

References.

- J. P. Keener and J. Sneyd, Mathematical Physiology, Chapter 4 and Chapter 8 [7].

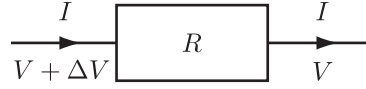
8.1 Background

Here we outline the background physics required to write down a model to describe a nerve impulse. Firstly, we note that:

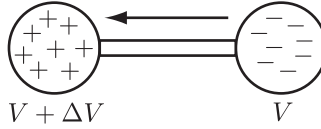
- numerous fundamental particles, ions and molecules have an electric charge, *e.g.* the electron, e^- , and the sodium ion, Na^+ ;
- it is an empirical fact that total charge is conserved;
- electric charges exert electrical forces on one another such that like charges repel and unlike charges attract. The electric potential, denoted V , is the potential energy of a unit of charge due to such forces and is measured in volts;
- a concentration of positive particles has a large *positive* potential, while a concentration of negative particles has a large, *but negative* potential;
- electric current is defined to be the rate of flow of electric charge, measured in Amps.

8.1.1 Resistance

Ohms Law, $\Delta V = IR$, holds in most situations, where ΔV is the change in potential, I is the current flowing and R , which may depend on material properties and geometries *but not on I nor V* is the resistance.

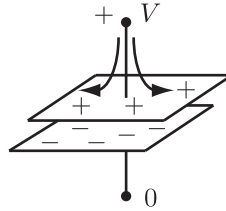


Key point. Suppose one uses a wire of low resistance to connect a region with a concentration of positive charges to a region with a concentration of negative charges. The charges will, very quickly, flow onto/off the wire until the potential is constant and there is no further flow of charge.



8.1.2 Capacitance

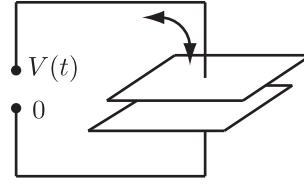
A simple example of capacitor is two conducting plates, separated by an insulator, for example, an air gap.



Connecting a battery to the plates, as illustrated, using wires of low resistance leads to charge flowing onto/off the plates. It will equilibrate (very quickly!); let Q_{eqm} denote the difference in the total value of the charge stored on the two plates. The capacitance of the plates, C , is defined to be

$$C = \frac{Q_{eqm}}{V} > 0, \quad (8.1)$$

where C is a constant, independent of V . Thus the higher the capacitance, the better the plates are at storing charge, for a given potential.



Suppose now the potential is a function of time, $V = V(t)$. Charge will flow on and off the plates in response to time dependent changes in $V(t)$. If the time for a significant change in $V(t)$ is far longer than the time it takes for the difference in the total value of the charge stored on the two plates, Q , to reach its equilibrium value, Q_{eqm} , as is essentially always the case, one has

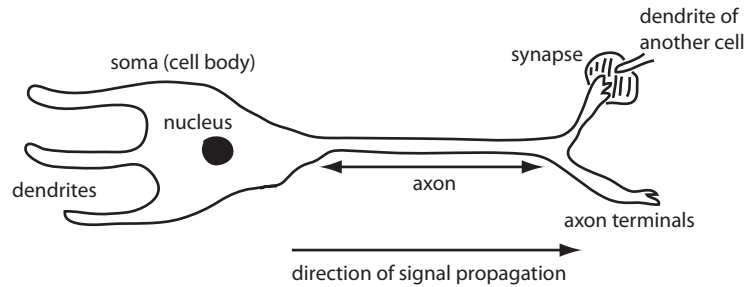
$$Q = Q_{eqm} = CV(t). \quad (8.2)$$

Hence, the current, J , *i.e.* the rate of flow of charge on/off the plates is given by

$$J = \dot{Q} = C\dot{V}. \quad (8.3)$$

8.2 Deducing the Fitzhugh Nagumo equations

An axon is a part of nerve cell:

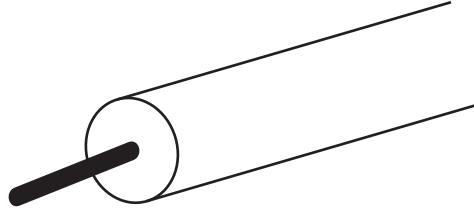
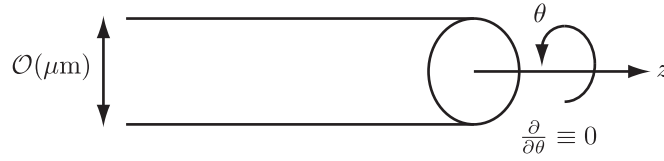


The nerve signal along the axon is, in essence, a propagating pulse in the potential difference across the plasma (*i.e.* outer) membrane of the axon. This potential difference, V , arises due to the preferential permeability of the axon plasma membrane which allows potassium and sodium ions, K^+ and Na^+ , to pass through the membrane at rates which differ between the two ions and vary with V . In the rest state, $V = V_{rest} \simeq -70\text{mV}$ (millivolts); in a nerve signal pulse in V rises to a peak of $\sim 15\text{mV}$. It is this pulse we are interested in modelling.

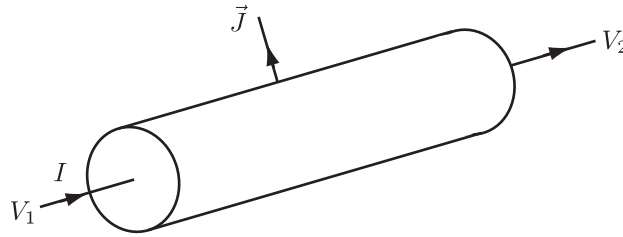
The geometry of the axon can be treated as a cylindrical tube. An axon is axisymmetric, so we have no θ dependence in any of our models of the axon.

8.2.1 Space-clamped axon

A common, simplifying, experimental scenario is to *space-clamp* the axon, *i.e.* to place a conducting wire along the axon's axis of symmetry.



- The interior of the axon will quickly equilibrate, and there will be no spatial variation in the potential difference, nor any current, along the inside of the axon.



- Thus, by conservation of charge, the *total current flowing across the axon membrane must be zero*.
- Note that any changes in the interior due to, for example the axon allowing K^+ and Na^+ to pass through its membrane, will occur on a much slower timescale and hence one has that the interior of the space-clamped axon has no spatial variation in its potential difference, no current flowing along the inside of the axon, and, most importantly, the *total current flowing through the axon membrane is zero*.

The basic model for the *space-clamped* axon plasma membrane potential is given by

$$\begin{aligned} 0 &= \text{total transmembrane current per unit area,} \\ &= c \frac{dV}{dt} + I_{Na} + I_K + I_0 + I_{applied}(t), \end{aligned} \quad (8.4)$$

where

- $I_{applied}(t)$ is the applied current, *i.e.* the current injected through the axon plasma membrane in the experiment, which is only function of time in most experimental set-ups. We will take $I_{applied}(t) = 0$ below.

- $c dV/dt$ is the capacitance current through a unit area of the membrane. *Recall:*

Rate of flow on/off capacitor = $C dV/dt$.

Therefore rate of flow of charge per unit area of membrane = $c dV/dt$ where c is the membrane capacitance per unit area.

- I_{Na} , I_K are the voltage dependent Na^+ and K^+ currents. I_0 is a voltage dependent background current.

These currents actually take complicated forms involving numerous other variables which satisfy complex equations, that can be simplified, if somewhat crudely. An excellent account is given in T. F. Weiss, Cellular Biophysics [12].

The resulting equations written in terms of the non-dimensional variables $v = (V - V_{rest})/|V_{rest}|$ and $\tau = t/T$ where $T = 6$ ms, the time scale of a typical nerve pulse, are

$$\epsilon \frac{dv}{d\tau} = Av(\delta - v)(v - 1) - n, \quad (8.5)$$

$$\frac{dn}{d\tau} = -\gamma n + v, \quad (8.6)$$

where A , γ , ϵ , δ are positive parameters such that $A, \gamma \sim \mathcal{O}(1)$, $0 < \epsilon \ll \delta \ll 1$.

Key point. The spatially independent behaviour of a *space-clamped* axon is approximated by the above Fitzhugh Nagumo equations, (8.5)-(8.6).

8.3 A brief look at the Fitzhugh Nagumo equations

We have

$$\epsilon \frac{dv}{d\tau} = Av(\delta - v)(v - 1) - n, \quad (8.7)$$

$$\frac{dn}{d\tau} = -\gamma n + v, \quad (8.8)$$

where A , γ , ϵ , δ are positive parameters such that $A, \gamma \sim \mathcal{O}(1)$, $0 < \epsilon \ll \delta \ll 1$.

8.3.1 The (n, v) phase plane

The *nullclines* of equations (8.5)-(8.6) are the lines where $\dot{v} = 0$ and $\dot{n} = 0$. A plot of the nullclines separates the (v, n) phase plane into four regions, as shown in Figure 8.1.

There are several things to note about the dynamics.

- There is one stationary point which is a stable focus.
- Thus, with initial conditions sufficiently close to the stationary point, the system evolves to the stationary point in a simple manner.

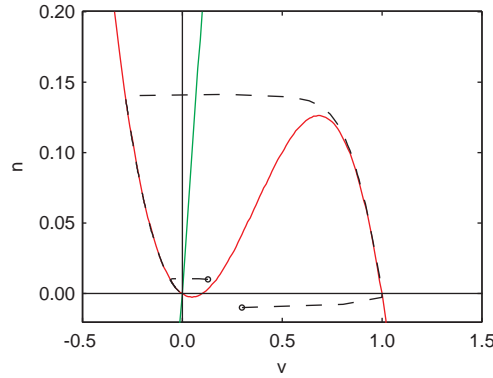


Figure 8.1: The phase plane for the Fitzhugh-Nagumo equations with the v nullcline shown in red and the n nullcline in green. The trajectories for two different initial perturbations from the steady state are shown as dashed lines. Parameters are as follows: $A = 1$, $\gamma = 0.5$, $\delta = 0.1$ and $\epsilon = 0.001$.

- Consider initial conditions with $n \sim 0$, but v increased sufficiently. The system does not simply relax back to the equilibrium. However, one can understand how the qualitative behaviour of the system by considering the phase plane.
- We anticipate that $v = (V - V_{rest})/|V_{rest}|$ behaves in the manner shown in Figure 8.2 for a sufficiently large perturbation in v .

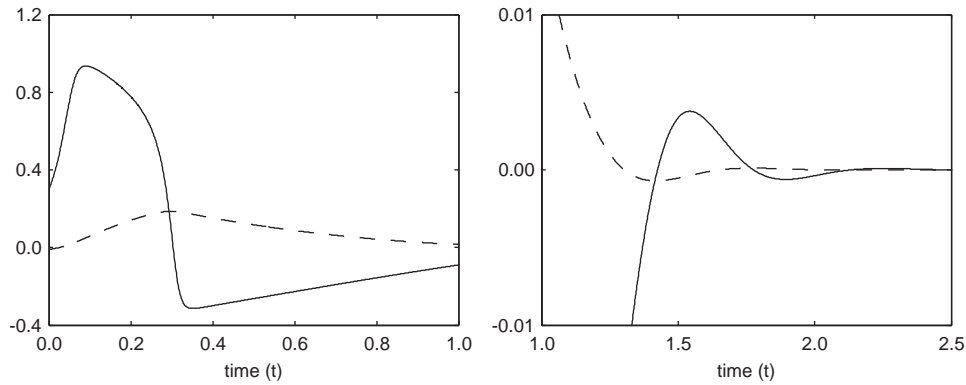


Figure 8.2: Solutions of the Fitzhugh Nagumo equations with v dynamics indicated by the solid line and n dynamics by the dashed line. The right-hand figure shows the oscillations that arise for large t . Parameters are as follows: $A = 1$, $\gamma = 0.5$, $\delta = 0.1$ and $\epsilon = 0.01$.

- This is essentially a nerve pulse (although because of the space clamping all the nerve axon is firing at once).

Definition. A system which, for a sufficiently large perturbation from a stationary point, undergoes a large change before eventually returning to the same stationary point is referred to as *excitable*.

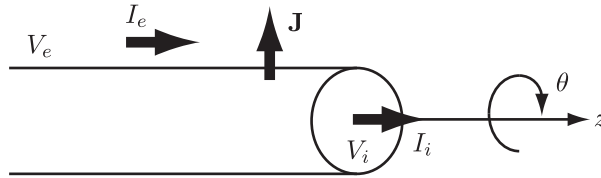
8.4 Modelling the propagation of nerve signals

In the following, we generalise the ideas we have seen for modelling the plasma membrane potential of an axon to scenarios where this potential can vary along the axon.

8.4.1 The cable model

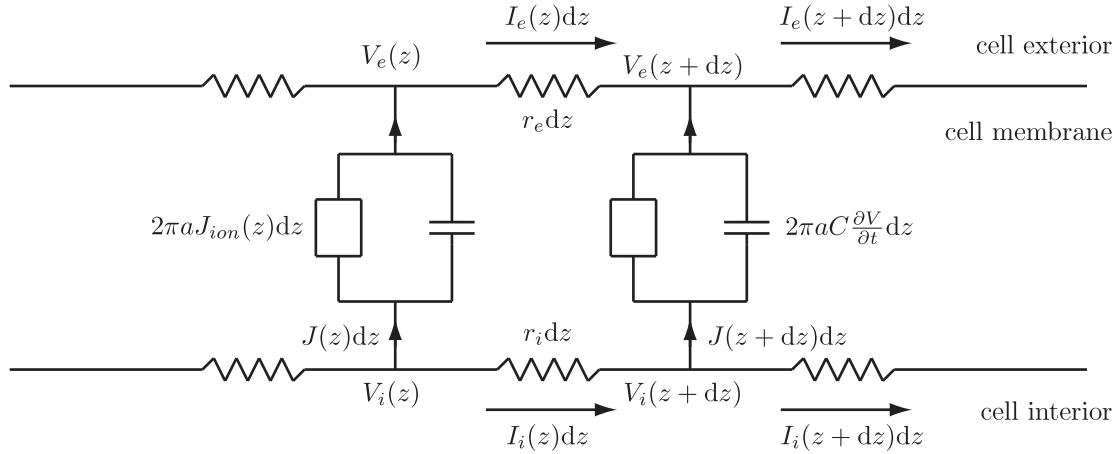
In the model we are about to develop we make following assumptions.

- The cell membrane is a cylindrical membrane separating two conductors of electric current, namely the extracellular and intracellular mediums. These are assumed to be homogeneous and to obey Ohm's law.
- The model has no θ dependence.
- A circuit theory description of current and voltages is adequate, *i.e.* quasi-static terms of Maxwell's equations are adequate; for example, electromagnetic radiation effects are totally negligible.
- Currents flow through the membrane in the radial direction only.
- Currents flow through the extracellular medium in the axial direction only and the potential in the extracellular medium is a function of z only. Similarly for the potential in the intracellular medium.



These assumptions are appropriate for unmyelinated nerve axons. Deriving the model requires considering the following variables:

- $I_e(z, t)$ – external current;
- $I_i(z, t)$ – internal current;
- $J(z, t)$ – total current through the membrane per unit length;
- $J_{ion}(z, t)$ – total ion current through the membrane per unit area;
- $V(z, t) = V_i(z, t) - V_e(z, t)$ – transmembrane potential;
- r_i – internal resistance per unit length;
- r_e – external resistance per unit length;
- C – membrane capacitance per unit area.



Consider the axial current in the extracellular medium, which has resistance r_e per unit length. We have

$$V_e(z + dz) - V_e(z) = -r_e I_e(z) dz \quad \Rightarrow \quad r_e I_e(z) = -\frac{\partial V_e}{\partial z}, \quad (8.9)$$

where the minus sign appears because of the convention that positive current is a flow of positive charges in the direction of increasing z . Hence, if $V_e(z + dz) > V_e(z)$ then positive charges flow in the direction of decreasing z giving a negative current. Similarly,

$$r_i I_i(z) = -\frac{\partial V_i}{\partial z}. \quad (8.10)$$

Using conservation of current, we have

$$I_e(z + dz, t) - I_e(z, t) = J(z, t) dz = I_i(z, t) - I_i(z + dz, t), \quad (8.11)$$

which gives

$$J(z, t) = -\frac{\partial I_i}{\partial z} = \frac{\partial I_e}{\partial z}. \quad (8.12)$$

Hence

$$J = \frac{1}{r_i} \frac{\partial^2 V_i}{\partial z^2} = -\frac{1}{r_e} \frac{\partial^2 V_e}{\partial z^2}, \quad (8.13)$$

and so

$$\frac{\partial^2 V}{\partial z^2} = (r_i + r_e) J. \quad (8.14)$$

Putting this all together gives

$$0 = -\frac{\partial(I_i + I_e)}{\partial z} = \frac{\partial}{\partial z} \left(\frac{1}{r_e} \frac{\partial V_e}{\partial z} + \frac{1}{r_i} \frac{\partial V_i}{\partial z} \right) = \left(\frac{r_e + r_i}{r_e r_i} \right) \frac{\partial^2 V_e}{\partial z^2} + \frac{1}{r_i} \frac{\partial^2 V}{\partial z^2}, \quad (8.15)$$

and so

$$0 = \frac{1}{r_i} \frac{\partial^2 V}{\partial z^2} - \left(\frac{r_e + r_i}{r_i} \right) \frac{\partial I_e}{\partial z} = \frac{1}{r_i} \left(\frac{\partial^2 V}{\partial z^2} + (r_e + r_i) J(z, t) \right). \quad (8.16)$$

We also have that

$$J(z, t) = 2\pi a \left(J_{ion}(V, z, t) + C \frac{\partial V}{\partial t} \right), \quad (8.17)$$

and finally, therefore,

$$\frac{1}{2\pi a(r_i + r_e)} \frac{\partial^2 V}{\partial z^2} = C \frac{\partial V}{\partial t} + J_{ion}(V, z, t). \quad (8.18)$$

This gives an equation relating the cell plasma membrane potential, V , to the currents across the cell plasma membrane due to the flow of ions, $J_{ion}(V, z, t)$.

Note 1. Note even though, physically, there is no diffusion, we still have a parabolic partial differential equation, so the techniques we have previously studied are readily applicable.

Note 2. From the above equation one can model cell plasma membrane potentials given suitable initial and boundary conditions, and a suitable expression for $J_{ion}(z, t)$.

We use the same expression for $J_{ion}(z, t)$, *i.e.* the expression for $I_{Na} + I_K + I_0$ as in the Fitzhugh Nagumo model of a space-clamped axon.

Thus with $v = (V - V_{rest})/|V_{rest}|$ and $x = Kz$, where K is a constant, we have

$$\epsilon \frac{\partial v}{\partial \tau} = \epsilon^2 \frac{\partial^2 v}{\partial x^2} + Av(\delta - v)(v - 1) - n, \quad (8.19)$$

$$\frac{dn}{d\tau} = -\gamma n + v, \quad (8.20)$$

where $0 < A, \gamma \sim \mathcal{O}(1)$, $0 < \epsilon \ll \delta \ll 1$.

Note that K has been chosen so that the coefficient in front of the v_{xx} term is ϵ^2 . This means, with respect to such variables, the front of the nerve pulse is extremely sharp. Hence, for such a scaling to exist, the extent of the nerve pulse must be less than ϵL , where L is the length of the axon; this constraint holds true for typical parameter estimates. The reason for the choice of this scaling is simply mathematical convenience in a travelling wave analysis.

We are interested in *nerve pulses*, so we take the boundary conditions to be $n, v \rightarrow 0$ as $x \rightarrow \pm\infty$.

We thus again have a system of parabolic partial differential equations to solve, and we are particularly interested in travelling pulse solutions. This entails that a travelling wave analysis would be most insightful. With the travelling wave coordinate $y = x - c\tau$ and $v(y) = v(x, \tau)$, $n(y) = n(x, \tau)$, we obtain

$$\epsilon^2 \frac{d^2 v}{dy^2} + \epsilon c \frac{dv}{dy} + Av(\delta - v)(v - 1) - n = 0, \quad (8.21)$$

$$c \frac{dn}{dy} - \gamma n + v = 0. \quad (8.22)$$

We have $0 < A \sim \mathcal{O}(1)$, $0 < \gamma^{-1}$, $\delta, \epsilon \ll 1$. One can readily investigate these ordinary differential equations to find that the travelling wave speed is *unique*, giving a unique prediction for the speed of a nerve pulse in terms of biophysical parameters.

Appendix A

The phase plane

Throughout this appendix we will be concerned with systems of two coupled, first-order, autonomous, non-linear ordinary differential equations.

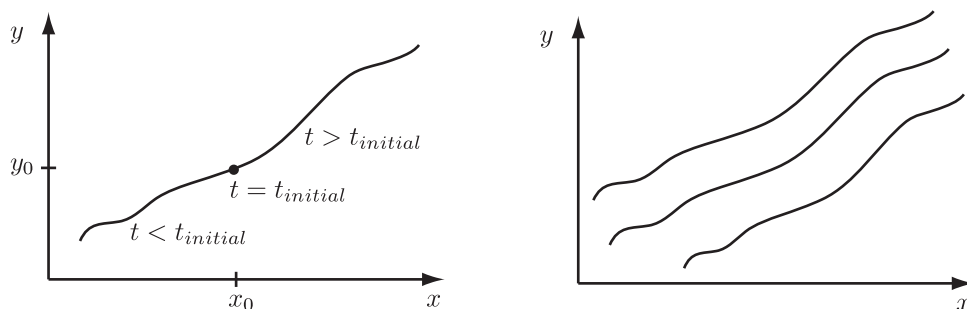
Disclaimer. This material should have been covered elsewhere (for example in your course on differential equations) and hence below is intended to review, rather than introduce and lecture this topic.

We can represent solutions to the equations

$$\frac{dx}{dt} = X(x, y), \quad (\text{A.1})$$

$$\frac{dy}{dt} = Y(x, y), \quad (\text{A.2})$$

as trajectories (or “integral paths”) in the phase plane, that is the (x, y) plane. Suppose, for the initial condition $x(t = t_{\text{initial}}) = x_0$, $y(t = t_{\text{initial}}) = y_0$ we plot, in the (x, y) plane, the solution of (A.1):



We can do exactly the same for all the values of $\{t_{\text{initial}}, x_{\text{initial}}, y_{\text{initial}}\}$, to build-up a graphical representation of the solutions to the equations (A.1) and (A.2) for many initial conditions. This plot is referred to as the “*phase plane portrait*”.

A.1 Properties of the phase plane portrait

The gradient of the integral path through the point (x_0, y_0) is given by

$$\frac{dy}{dx} = \frac{dy}{dt} \bigg/ \frac{dx}{dt} = \left(\frac{Y(x, y)}{X(x, y)} \right) \bigg|_{(x_0, y_0)} = \frac{Y(x_0, y_0)}{X(x_0, y_0)}. \quad (\text{A.3})$$

Key point 1. Note that if $Y(x_0, y_0) = 0$ and $X(x_0, y_0) \neq 0$ then

$$\left(\frac{dy}{dx} \right) \bigg|_{(x_0, y_0)} = 0, \quad (\text{A.4})$$

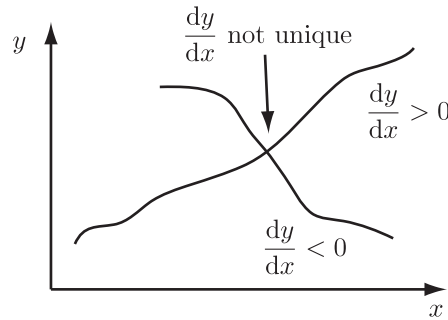
which corresponds to a horizontal line segments in the phase plane.

Key point 2. If $Y(x_0, y_0) \neq 0$ and $X(x_0, y_0) = 0$ then

$$\left| \frac{dy}{dx} \right| \rightarrow \infty \quad \text{as} \quad (x, y) \rightarrow (x_0, y_0), \quad (\text{A.5})$$

which corresponds to a vertical line segment in the phase plane.

Key point 3. Assuming that either $X(x_0, y_0) \neq 0$ or $Y(x_0, y_0) \neq 0$, then two path integral curves do not cross at the point (x_0, y_0) . This is because under these circumstances dy/dx takes a unique value, *i.e.* the following is *not* possible:



A.2 Equilibrium points

Definition. A point in the phase plane where $X(x_0, y_0) = Y(x_0, y_0) = 0$ is defined to be an *equilibrium point*, or equivalently, a *stationary point*.

The reason for the above definition is because if $(x, y) = (x_0, y_0)$ then both dx/dt and dy/dt are zero, and hence (x, y) do not change as t increases; hence $x(t)$, $y(t)$ remain at (x_0, y_0) for all time.

Key point 1. Integral curves cannot cross at points which are not equilibrium points.

Key point 2. If an integral path ends it must end on a stationary point.

Key point 3. As we shall see below, equilibrium points are only approached as $t \rightarrow \infty$ or $t \rightarrow -\infty$.

However, what about the gradient of integral paths at (x_0, y_0) ? We informally have

$$\frac{dy}{dx} = \frac{0}{0}, \quad (\text{A.6})$$

which is not uniquely defined—the value ultimately depends on the details of how quickly $X(x, y)$ and $Y(x, y)$ approach zero as $(x, y) \rightarrow (x_0, y_0)$, and this generally depends on the direction upon which (x, y) approaches (x_0, y_0) .

A.2.1 Equilibrium points: further properties

Suppose the equations (A.1) and (A.2) have an equilibrium point at (x_0, y_0) . Thus $X(x_0, y_0) = Y(x_0, y_0) = 0$. To determine the behaviour of integral paths close to the equilibrium point we write

$$x = x_0 + \bar{x}, \quad y = y_0 + \bar{y}, \quad (\text{A.7})$$

where it is assumed that \bar{x}, \bar{y} are sufficiently small to allow the approximations that we will make below.

By Taylor expansion, we have

$$\begin{aligned} X(x, y) &= X(x_0 + \bar{x}, y_0 + \bar{y}) = X(x_0, y_0) + \bar{x} \frac{\partial X}{\partial x}(x_0, y_0) + \bar{y} \frac{\partial X}{\partial y}(x_0, y_0) + \text{h.o.t.} \\ &= \bar{x} \frac{\partial X}{\partial x}(x_0, y_0) + \bar{y} \frac{\partial X}{\partial y}(x_0, y_0) + \text{h.o.t.}, \end{aligned} \quad (\text{A.8})$$

using the fact $X(x_0, y_0) = 0$. Similarly, we have

$$\begin{aligned} Y(x, y) &= Y(x_0 + \bar{x}, y_0 + \bar{y}) = Y(x_0, y_0) + \bar{x} \frac{\partial Y}{\partial x}(x_0, y_0) + \bar{y} \frac{\partial Y}{\partial y}(x_0, y_0) + \text{h.o.t.}, \\ &= \bar{x} \frac{\partial Y}{\partial x}(x_0, y_0) + \bar{y} \frac{\partial Y}{\partial y}(x_0, y_0) + \text{h.o.t.} \end{aligned} \quad (\text{A.9})$$

Note that x_0 and y_0 are constant, and hence have zero time derivative. Hence, by use of Taylor expansions and neglecting higher orders (*i.e.* taking \bar{x}, \bar{y} sufficiently small), we can neglect terms of the order $\mathcal{O}(\bar{x}\bar{y}, \bar{x}^2, \bar{y}^2)$ and hence we can write equations (A.1) and (A.2) in the form

$$\frac{d\mathbf{u}}{dt} = \begin{pmatrix} \frac{\partial X}{\partial x}(x_0, y_0) & \frac{\partial X}{\partial y}(x_0, y_0) \\ \frac{\partial Y}{\partial x}(x_0, y_0) & \frac{\partial Y}{\partial y}(x_0, y_0) \end{pmatrix} \mathbf{u} \stackrel{\text{def}}{=} \mathbf{J}\mathbf{u} \quad \text{where} \quad \mathbf{u} \stackrel{\text{def}}{=} \begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}. \quad (\text{A.10})$$

Definition. The matrix

$$\mathbf{J} = \begin{pmatrix} \frac{\partial X}{\partial x}(x_0, y_0) & \frac{\partial X}{\partial y}(x_0, y_0) \\ \frac{\partial Y}{\partial x}(x_0, y_0) & \frac{\partial Y}{\partial y}(x_0, y_0) \end{pmatrix}, \quad (\text{A.11})$$

is defined to be the *Jacobian* matrix at the equilibrium point (x_0, y_0) .

A.3 Summary

The key points thus far are as follows.

1. We have taken the full *non-linear* equation system, (A.1) and (A.2), and expanded about one of its (possibly many) equilibrium points taken to be located at (x_0, y_0) , using Taylor expansions of $X(x, y)$, $Y(x, y)$.
2. We assume that we are sufficiently close to (x_0, y_0) to enable us to only consider linear terms of the order of $(x - x_0)$, $(y - y_0)$.
3. In this way, we obtain a set of two coupled, *linear*, autonomous ordinary differential equations, *i.e.* equation (A.10) above, which in principle we can solve!
4. This procedure is sometimes referred to as “a *linearisation* of equations (A.1) and (A.2) about the point (x_0, y_0) ”.
5. In *virtually all* cases the behaviour of the linearised system is the same as the behaviour of the full non-linear equations sufficiently close to the point (x_0, y_0) . In this respect one should note that the statement immediately above can be formulated more rigorously and proved for all the types of stationary points except:
 - centre type equilibrium points, *i.e.* case [3c] below;
 - the degenerate cases where $\lambda_1 = 0$ and/or $\lambda_2 = 0$, which are briefly mentioned in item 2 on page (87). These stationary points can be considered non-examinable.

The relevant theorem is “Hartmann’s theorem”, as discussed further in P. Glendinning, *Stability, Instability and Chaos* [4].

6. However, one should also note that the solution of the linearised equations may behave substantially differently from the solutions of the full *non-linear* equations, (A.1) and (A.2), sufficiently far from (x_0, y_0) .

A.4 Investigating solutions of the linearised equations

We now have a set of two coupled, linear, autonomous ordinary differential equations, (A.10). It is useful to look for a solution of the form

$$\mathbf{u} = \mathbf{u}_0 e^{\lambda t}, \quad (\text{A.12})$$

for some constant, λ . Substituting this into equation (A.10) we obtain

$$\lambda \mathbf{u}_0 e^{\lambda t} = \mathbf{J} \mathbf{u}_0 e^{\lambda t} \quad i.e. \quad (\mathbf{J} - \lambda \mathbf{I}) \mathbf{u}_0 = 0. \quad (\text{A.13})$$

For a non-zero solution, we must have $\mathbf{u}_0 \neq (0, 0)$ and hence we require

$$\det(\mathbf{J} - \lambda \mathbf{I}) = 0, \quad (\text{A.14})$$

where \mathbf{I} is the 2×2 identity matrix.

This quadratic equation has two roots for λ , denoted λ_1, λ_2 , which are possibly equal and possibly complex; these are, of course, the eigenvalues of \mathbf{J} evaluated at the point (x_0, y_0) .

A.4.1 Case I

λ_1, λ_2 real, with $\lambda_1 \neq 0, \lambda_2 \neq 0, \lambda_1 \neq \lambda_2$. *Without loss of generality* we take $\lambda_2 > \lambda_1$ below.

We have two distinct, real eigenvalues. Let the corresponding eigenvectors be denoted by \mathbf{e}_1 and \mathbf{e}_2 . We thus have

$$\mathbf{J} \mathbf{e}_1 = \lambda_1 \mathbf{e}_1, \quad \mathbf{J} \mathbf{e}_2 = \lambda_2 \mathbf{e}_2. \quad (\text{A.15})$$

We seek a solution of the form

$$\mathbf{u} = A_1 \mathbf{e}_1 + A_2 \mathbf{e}_2. \quad (\text{A.16})$$

Substituting this into equation (A.10), we find, by comparing coefficients of \mathbf{e}_1 and \mathbf{e}_2 , that

$$\frac{dA_1}{dt} = \lambda_1 A_1, \quad \frac{dA_2}{dt} = \lambda_2 A_2, \quad (\text{A.17})$$

and hence

$$A_1 = A_1(t=0)e^{\lambda_1 t}, \quad A_2 = A_2(t=0)e^{\lambda_2 t}. \quad (\text{A.18})$$

Thus we have

$$\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix} \stackrel{def}{=} \mathbf{u} = A_1(t=0)e^{\lambda_1 t} \mathbf{e}_1 + A_2(t=0)e^{\lambda_2 t} \mathbf{e}_2, \quad (\text{A.19})$$

which gives us a representation of the solution of (A.10) for general initial conditions. This information is best displayed graphically, and we do so below according to the values of λ_2, λ_1 .

Note. The equilibrium point *i.e.* $(\bar{x}, \bar{y}) = (0, 0)$ can only be reached either as $t \rightarrow \infty$ or $t \rightarrow -\infty$.

1. $\lambda_1 < \lambda_2 < 0$. The phase plot of the linearised equations in the (\bar{x}, \bar{y}) plane looks like one of the two possibilities in Figure A.1.

Definition. An equilibrium point which results in this case is called a *stable node*, with the word “stable” referring to the fact that integral paths *enter* the node, *i.e.* the equilibrium point at $(0, 0)$.

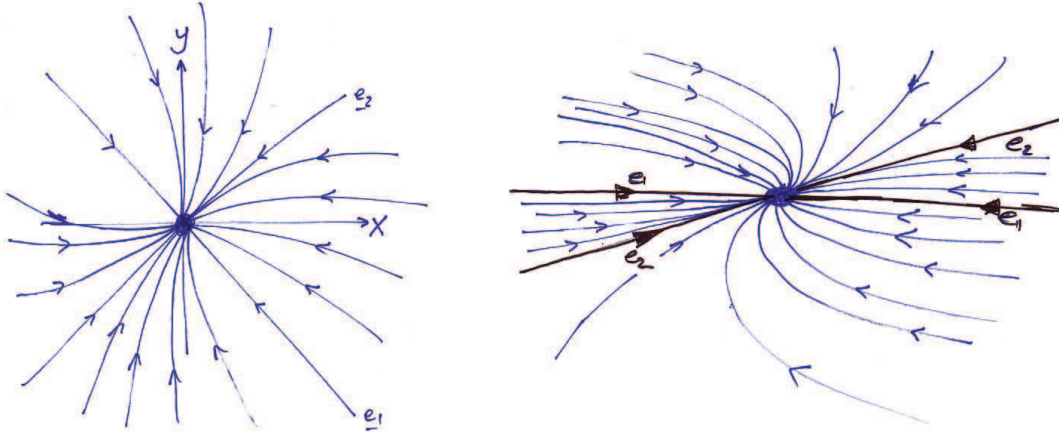


Figure A.1: Possible phase portraits of a stable node. The equilibrium point in each case is denoted by the large dot.

2. $\lambda_2 > \lambda_1 > 0$. We still have

$$\mathbf{u} = A_1(t=0)e^{\lambda_1 t}\mathbf{e}_1 + A_2(t=0)e^{\lambda_2 t}\mathbf{e}_2. \quad (\text{A.20})$$

However, the direction of the arrows is reversed as the signs of λ_1, λ_2 are changed. The phase plane portraits are the same as in Figure A.1 except the direction of the arrows is reversed.

Definition. An equilibrium point which results in this case is called an *unstable node*, with the word “unstable” referring to the fact that integral paths *leave* the node, *i.e.* the equilibrium point at $(0, 0)$.

3. $\lambda_2 > 0 > \lambda_1$. Once more, we still have

$$\mathbf{u} = A_1(t=0)e^{\lambda_1 t}\mathbf{e}_1 + A_2(t=0)e^{\lambda_2 t}\mathbf{e}_2, \quad (\text{A.21})$$

but again the phase plane portrait is slightly different—see Figure A.2.

Definition. An equilibrium point which results in this case is called a *saddle point*.

Definition. The two integral paths originating from the saddle point are sometimes referred to as the *unstable manifolds* of the saddle point. Conversely, the integral paths tending to the saddle point are sometimes referred to as the *stable manifolds* of the saddle point. This forms part of a nomenclature system commonly used in more advanced dynamical systems theory; see P. Glendinning, *Stability, Instability and Chaos* [4].

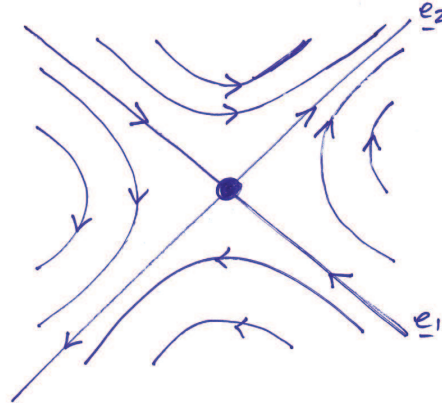


Figure A.2: The phase portrait of a saddle point. The equilibrium point is denoted by the large dot.

A.4.2 Case II

λ_2, λ_1 real. One, or more, of the following also holds:

$$\lambda_2 = \lambda_1, \quad \lambda_1 = 0, \quad \lambda_2 = 0. \quad (\text{A.22})$$

We typically will not encounter these degenerate cases in this course. We briefly note that behaviour of the full equations, (A.1), can be highly nontrivial when the linearisation reduces to these degenerate cases. Further details of such cases can be found in P. Glendinning, *Stability, Instability and Chaos* [4], which is on the reading list for this course. When $\lambda_1, \lambda_2 = 0$, Hartmann's theorem doesn't hold.

A.4.3 Case III

λ_2, λ_1 complex. The complex eigenvalues of a real matrix always occur in complex conjugate pairs. Thus we take, without loss of generality,

$$\lambda_1 = a - ib = \lambda_2^*, \quad \lambda_2 = a + ib = \lambda_1^*, \quad (\text{A.23})$$

where a, b real, $b \neq 0$, and $*$ denotes the complex conjugate.

We also have two associated complex eigenvectors $\mathbf{e}_1, \mathbf{e}_2$, satisfying

$$\mathbf{J}\mathbf{e}_1 = \lambda_1\mathbf{e}_1, \quad \mathbf{J}\mathbf{e}_2 = \lambda_2\mathbf{e}_2, \quad (\text{A.24})$$

which are complex conjugates of each other, *i.e.* $\mathbf{e}_1 = \mathbf{e}_2^*$.

Using the same idea as in Case I above, we have

$$\mathbf{u} = A_1(t=0)e^{\lambda_1 t}\mathbf{e}_1 + A_2(t=0)e^{\lambda_2 t}\mathbf{e}_2, \quad (\text{A.25})$$

though now, in general, $A_1(t=0), \lambda_1, \mathbf{e}_1, A_2(t=0), \lambda_2, \mathbf{e}_2$ are complex, and hence so is \mathbf{u} .

Restricting \mathbf{u} to be real gives

$$\mathbf{u} = A_1(t=0)e^{\lambda_1 t}\mathbf{e}_1 + A_1^*(t=0)e^{\lambda_2 t}\mathbf{e}_2 = A_1(t=0)e^{\lambda_1 t}\mathbf{e}_1 + (A_1(t=0)e^{\lambda_1 t}\mathbf{e}_1)^*, \quad (\text{A.26})$$

and this is real, as for any complex number z , we have $z + z^* \in \mathbb{R}$.

After some algebra this reduces to

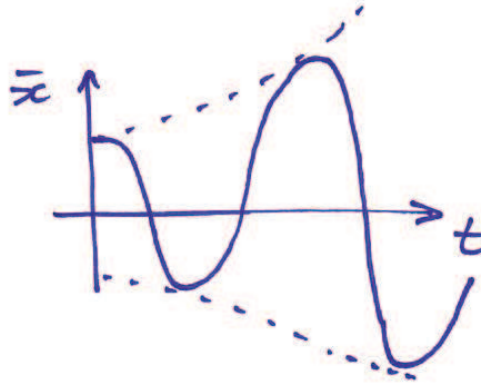
$$\mathbf{u} = e^{at} [\mathbf{M} \cos(bt) + \mathbf{K} \sin(bt)] = e^{at} \left[\begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \cos(bt) + \begin{pmatrix} K_1 \\ K_2 \end{pmatrix} \sin(bt) \right], \quad (\text{A.27})$$

where $\mathbf{M} = (M_1, M_2)^T$, $\mathbf{K} = (K_1, K_2)^T$ are real, constant vectors, which can be expressed in terms of $A_1(t=0)$, $A_2(t=0)$ and the components of the eigenvectors \mathbf{e}_1 and \mathbf{e}_2 . Equivalently, we have

$$\bar{x} = e^{at} [\cos(bt)M_1 + \sin(bt)K_1], \quad \bar{y} = e^{at} [\cos(bt)M_2 + \sin(bt)K_2], \quad (\text{A.28})$$

where M_1, M_2, K_1, K_2 are real constants.

1. $a > 0$. We have \bar{x}, \bar{y} are, overall, increasing exponentially but are oscillating too. For, example, with $K_1 = 0, M_1 = 1$ we have $\bar{x} = e^{at} \cos(bt)$, which looks like:



Note that the overall growth of \bar{x} is exponential at rate a . Thus, in general, the phase plane portrait looks like one of the examples shows in Figure A.3.

Note. The sense of the rotation, clockwise or anti-clockwise, is easily determined by calculating $d\bar{y}/dt$ when $\bar{y} = 0$ or $d\bar{x}/dt$ when $\bar{x} = 0$.

Definition. An equilibrium point which results in the above, is called an *unstable spiral* or, equivalently, an *unstable focus*. The word “unstable” refers to the fact that integral paths *leave* the equilibrium point.

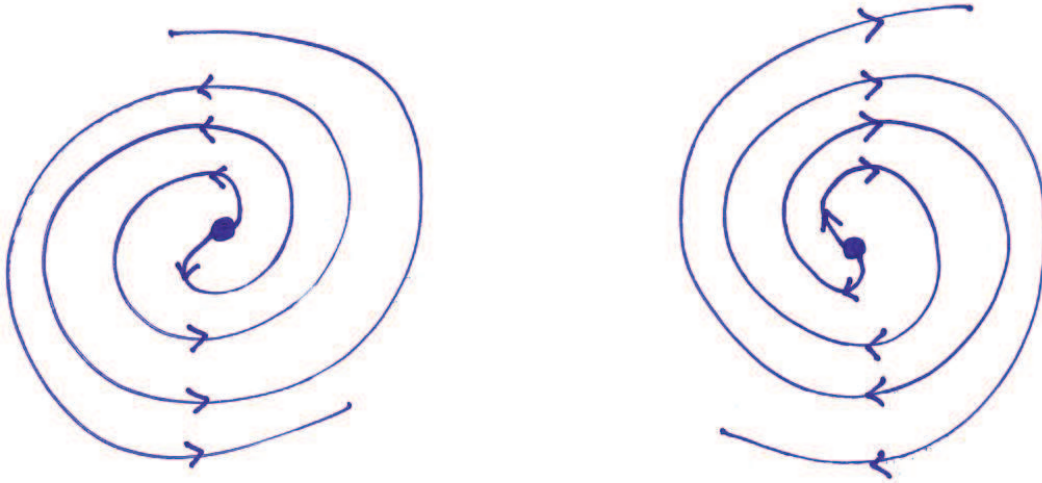


Figure A.3: Possible phase portraits of a focus. The equilibrium point in each case is denoted by the large dot.

2. $a < 0$. This is the same as 1. except now the phase plane portrait arrows point towards the equilibrium point as \bar{x} and \bar{y} are exponentially decaying as time increases rather than exponentially growing.

Definition. An equilibrium point which results in this case is called a *stable spiral* or, equivalently, a *stable focus*. The word “stable” refers to the fact that integral paths *enter* the equilibrium point.

3. $a = 0$. Thus we have $\lambda_2 = -ib = -\lambda_1$, $b \neq 0$, b real, and

$$\bar{x} = [\cos(bt)M_1 + \sin(bt)K_1], \quad \bar{y} = [\cos(bt)M_2 + \sin(bt)K_2], \quad (\text{A.29})$$

where M_1, M_2, K_1, K_2 are constants. Note that

$$K_2\bar{x} - K_1\bar{y} = L \cos(bt), \quad -M_2\bar{x} + M_1\bar{y} = L \sin(bt), \quad (\text{A.30})$$

where $L = K_2M_1 - K_1M_2$. Letting $x^* = K_2\bar{x} - K_1\bar{y}$ and $y^* = -M_2\bar{x} + M_1\bar{y}$, we have

$$(x^*)^2 + (y^*)^2 = L^2, \quad (\text{A.31})$$

i.e. a circle in the (x^*, y^*) plane, enclosing the origin, which is equivalent to, in general, a closed ellipse, in the (\bar{x}, \bar{y}) plane enclosing the origin.

Note. As with 3. above, the sense of the rotation, clockwise or anti-clockwise, is easily determined by calculating $d\bar{y}/dt$ when $\bar{y} = 0$ or $d\bar{x}/dt$ when $\bar{x} = 0$.

Definition. An equilibrium point which results in this case, is called a *centre*. A centre is an example of a limit cycle.

Definition. A *limit cycle* is an integral path which is closed (and which does not have any equilibrium points).

A.5 Linear stability

Definition. An equilibrium point is *linearly stable* if the real parts of both eigenvalues λ_1, λ_2 are negative.

From the expressions for \mathbf{u} above, for example

$$\mathbf{u} = A_1(t=0)e^{\lambda_1 t}\mathbf{e}_1 + A_2(t=0)e^{\lambda_2 t}\mathbf{e}_2, \quad (\text{A.32})$$

when λ_1, λ_2 real, we see that any perturbation away from the equilibrium decays back to the equilibrium point.

Definition. An equilibrium point is *linearly unstable* if the real parts of at least one of the eigenvalues λ_1, λ_2 is positive (and the other is non-zero).

Other situations are in general governed by the non-linear behaviour of the full equations and we do not need to consider them here.

A.5.1 Technical point

The behaviour of the linearised equations and the behaviour of the non-linear equations sufficiently close to the equilibrium point are guaranteed to be the same for any of the equilibrium points [I 1-3], [III 1,2] or [II] with $\lambda_1 = \lambda_2 \neq 0$. All these equilibrium points are such that $Re(\lambda_1) Re(\lambda_2) \neq 0$. This is the essence of Hartmann's theorem. This guarantee does not hold for centres, or the equilibrium points described in [II] with $\lambda_1 \lambda_2 = 0$.

The underlying reasons for this are as follows.

- First, note from the above that integral paths which meet the equilibrium point can either grow/decay at exponential rate $Re(\lambda_1)$, or exponential rate $Re(\lambda_2)$, or consist of the sum of two such terms. Second, note that in the above we took a Taylor expansion. Including higher order terms in this Taylor expansion can lead to a small correction for the rate of exponential decay towards or growth away from the stationary point exhibited by the integral paths. These corrections tend to zero as one approaches the equilibrium point.
- Consider the centre equilibrium point, which has $Re(\lambda_1) = Re(\lambda_2) = 0$, and an exponential growth/decay of zero. If the corrections arising from the Taylor series are always positive, the exponential growth/decay rate of all integral paths sufficiently near the stationary point is always (slightly) positive. Hence these integral paths grow exponentially away from the stationary point. However, b is non-zero, so \bar{x} and \bar{y} are still oscillating. Hence one has the non-linear equations behave like a stable focus.

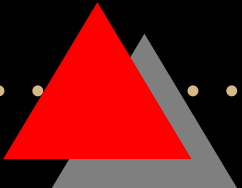

- If $Re(\lambda_1), Re(\lambda_2) \neq 0$, then for all integral paths reaching the stationary point, the above mentioned corrections, sufficiently close to the equilibrium point, are negligible, *e.g.* they cannot change exponential growth into exponential decay or *vice-versa*. This allows one to show that stationary points with $Re(\lambda_1) Re(\lambda_2) \neq 0$ are guaranteed to have the same behaviour for the linearised and the non-linear equations sufficiently close to the equilibrium point.

A.6 Summary

We will typically only encounter stationary points [I 1-3], [III 1-3]. Of these stationary points, all but centres exhibit the same behaviour for the linearised and the non-linear equations sufficiently close to the equilibrium point as plotted above and in D. W. Jordan and P. Smith, *Mathematical Techniques* [6].

Bibliography

- [1] N. F. Britton. *Essential Mathematical Biology*. Springer Undergraduate Mathematics Series. Springer, 2005.
- [2] L. Edelstein-Keshet. *Mathematical Models in Biology*. SIAM Classics in Applied Mathematics, 2005.
- [3] A. Gierer and H. Meinhardt. A theory of biological pattern formation. *Kybernetik*, 12:30–39, 1972.
- [4] P. Glendinning. *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations*. Cambridge Texts in Applied Mathematics, 1999.
- [5] T. Hillen and K. Painter. A user’s guide to PDE models for chemotaxis. *J. Math. Biol.*, 58(1):183–217, 2009.
- [6] D. W. Jordan and P. Smith. *Mathematical Techniques: An Introduction for the Engineering, Physical and Mathematical Sciences*. Oxford University Press, 3rd edition, 2002.
- [7] J. P. Keener and J. Sneyd. *Mathematical Physiology*, volume 8 of *Interdisciplinary Applied Mathematics*. Springer, New York, 1st edition, 1998.
- [8] J. D. Murray. *Mathematical Biology I: An Introduction*, volume I. Springer-Verlag, 3rd edition, 2003.
- [9] J. D. Murray. *Mathematical Biology II: Spatial Models and Biochemical Applications*, volume II. Springer-Verlag, 3rd edition, 2003.
- [10] A. Okubo, P. K. Maini, M. H. Williamson, and J. D. Murray. On the spread of the grey squirrel in Great Britain. *Proc. R. Soc. Lond. B*, 238(1291):113–125, 1989.
- [11] L. E. Reichl. *A Modern Course in Statistical Physics*. Wiley-VCH, 3rd edition, 2009.
- [12] T. F. Weiss. *Cellular Biophysics*, volume 2. MIT Press, 1996.



Ordinary Differential Equations and Introduction to Dynamical Systems

Holly D. Gaff

hgaff@tiem.utk.edu

University of Tennessee, Knoxville



Overview

- Single Species Systems
 - Solving for Equilibria
 - Evaluating Stability of Equilibria Graphically
- Two Species Systems
 - Lotka-Volterra Predator-Prey
 - Evaluating Stability of Equilibria
- Examples from Epidemiology



Single Species Systems

- Exponential Growth
- Logistic Growth
- Other Equations



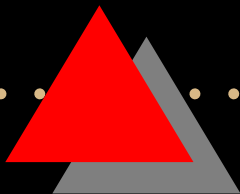
Exponential Growth

$$\frac{dN}{dt} = rN$$

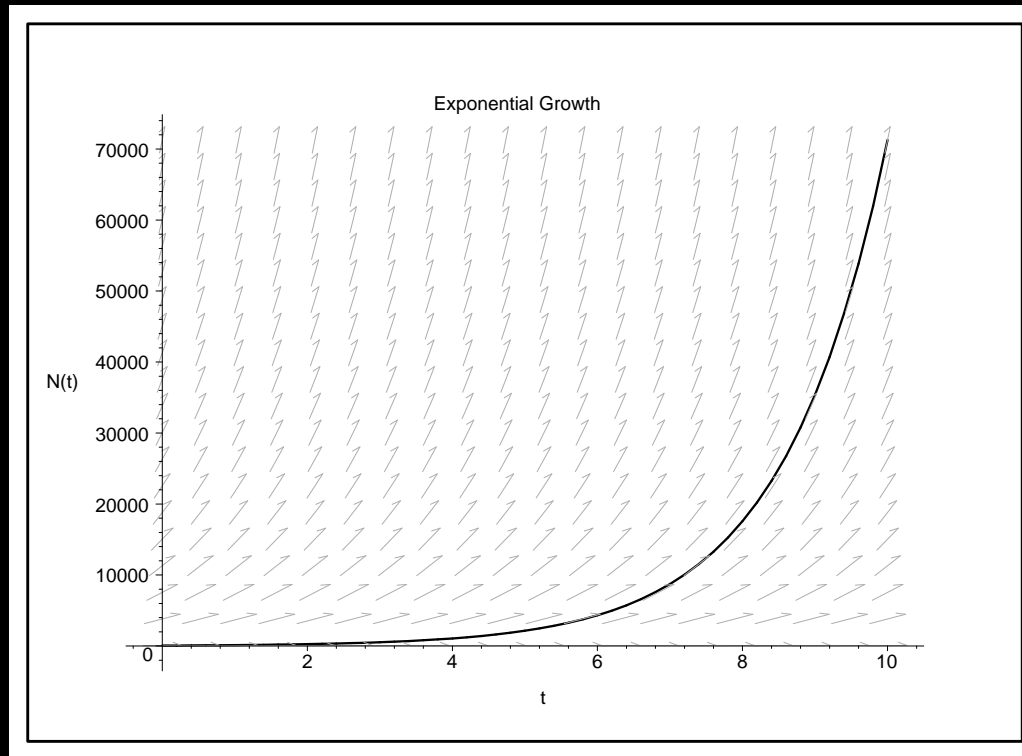
Solution:

$$N(t) = N_0 e^{rt}$$

What happens to this population?



Exponential Growth





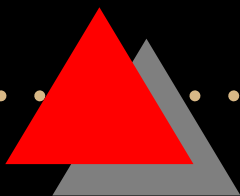
Logistic Growth

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right)$$

What do you think the solutions of this will look like?

Recall exponential growth was

$$\frac{dN}{dt} = rN$$



Equilibria

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right)$$

$$\frac{dN}{dt} = 0$$

$$rN^* \left(1 - \frac{N^*}{K} \right) = 0$$

$$N^* = 0 \quad \text{or} \quad N^* = K$$



Stability of Equilibria

First evaluate the stability of $N^* = 0$.

Near $N^* = 0$,

$$\frac{dN}{dt} \approx rN$$

So as N increase, $\frac{dN}{dt}$ grows exponentially.

Therefore, $N^* = 0$ is an unstable equilibrium.



Stability of Equilibria

What do you think will happen near $N^* = K$?

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right)$$



Stability near K

If N is just slightly above K ,

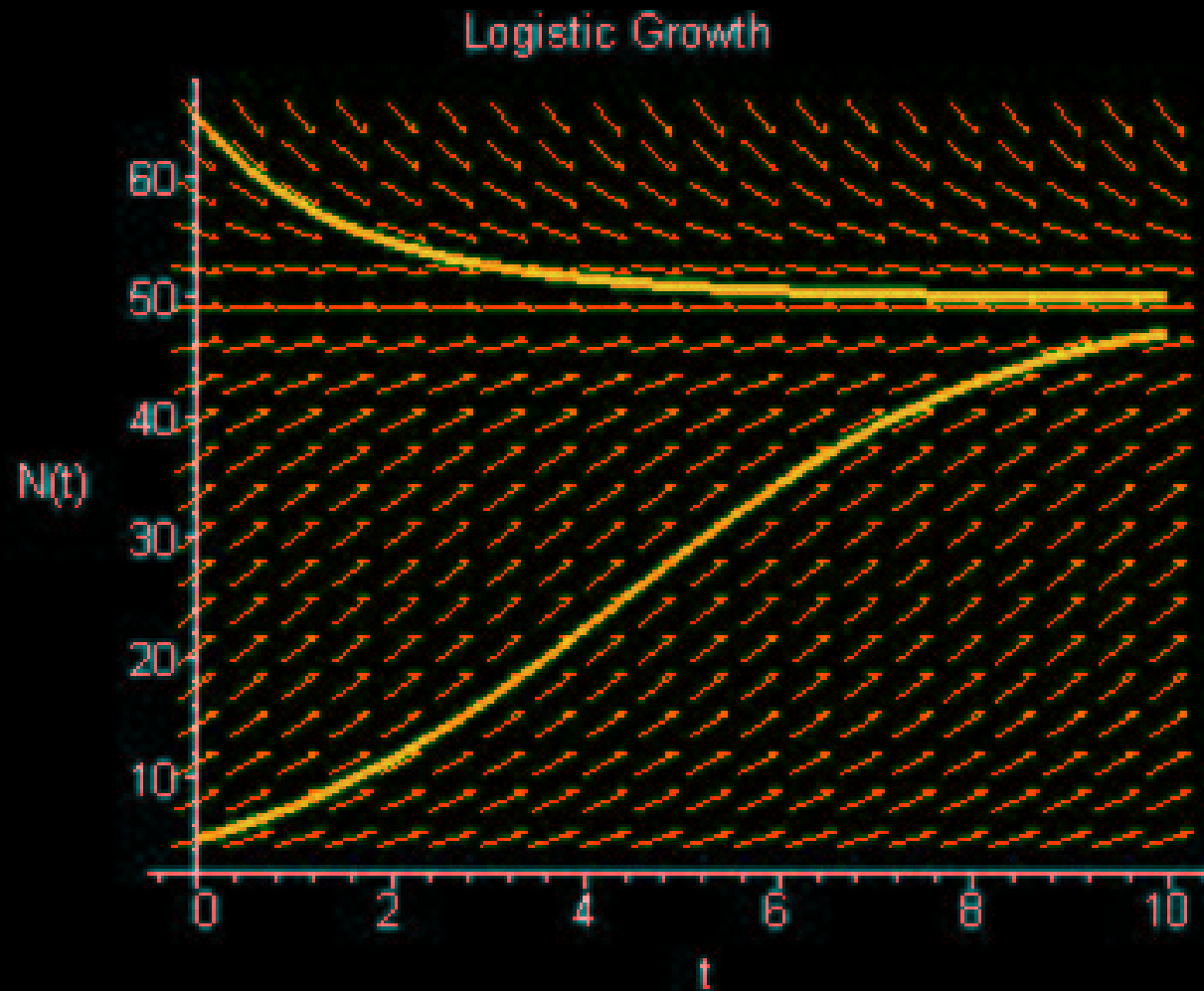
$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) < 0$$

but if N is just slightly below K ,

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) > 0$$

Therefore, $N^* = K$ is stable.

Graphical View of Stability





Other Alternatives

- Gompertz Equation

$$\frac{dN}{dt} = r_0 e^{-\alpha t} N$$

- Delay or lag time

$$\frac{dN}{dt} = F(N(t), N(t - T))$$



Other Alternatives

- Allee effect

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right) \left(\frac{N}{K_0} - 1\right)$$

- Discrete time

$$N(t+1) = F(N(t))$$

- Stochastic processes



Interacting Populations

- Predator-prey models
- Competition
- Mutualism



Classic Predator-Prey

Lotka-Volterra Predator-Prey Model

$$\begin{aligned}\frac{dN}{dt} &= rN - cNP \\ \frac{dP}{dt} &= bNP - mP\end{aligned}$$

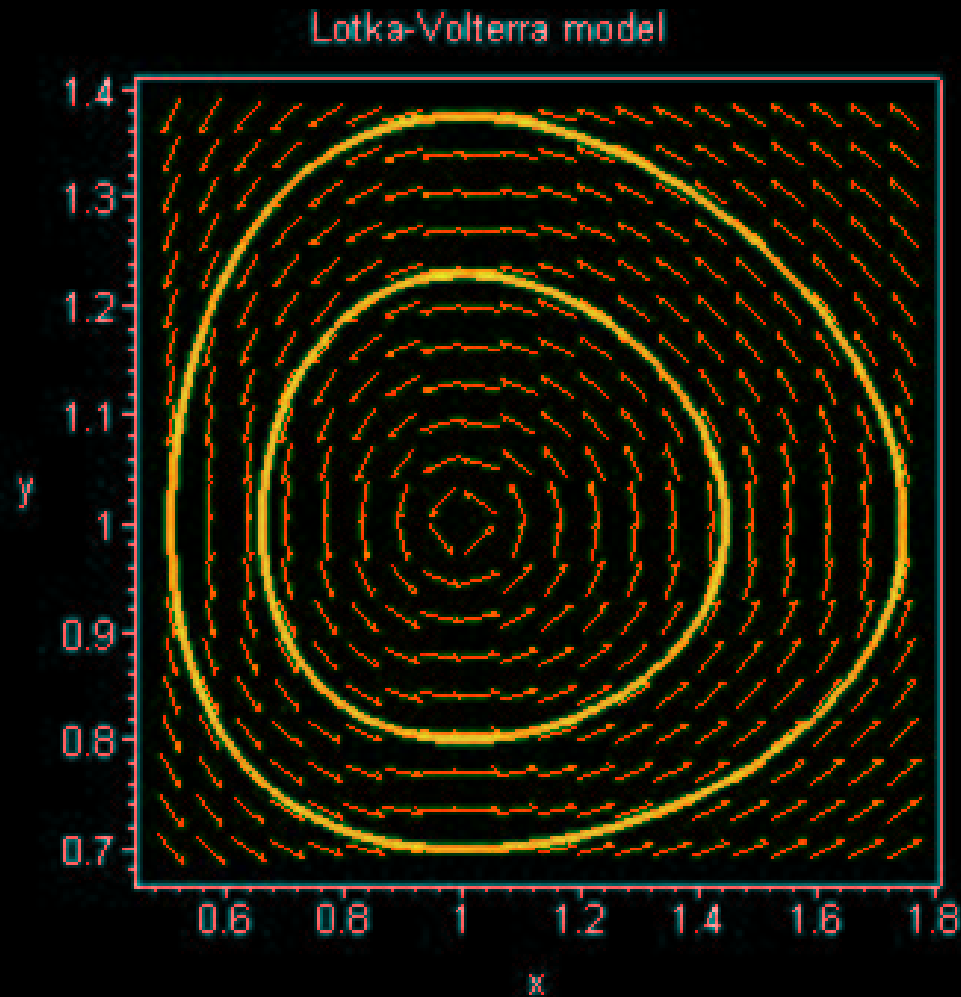


Classic Predator-Prey

Lotka-Volterra Predator-Prey Model

- Historical interest
- Mass-action term
- Bad mathematical model
- Structurally unstable

Lotka-Volterra Phase Plane





Interacting Populations

More Realistic Predator-prey models

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) - P \left(\frac{A}{N + B} \right)$$

$$\frac{dP}{dt} = eP \left(\frac{A}{N + B} \right) - dP$$



Interacting Populations

Another More Realistic Predator-prey models

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right) - P \left(\frac{AN}{N^2 + B^2} \right)$$

$$\frac{dP}{dt} = eP \left(\frac{AN}{N^2 + B^2} \right) - dP$$



Interacting Populations

Competition

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1}{K_1} - b_{12} \frac{N_2}{K_1} \right)$$
$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2}{K_2} - b_{21} \frac{N_1}{K_2} \right)$$



Interacting Populations

Mutualism

$$\begin{aligned}\frac{dN_1}{dt} &= r_1 N_1 \left(1 - \frac{N_1}{K_1} + b_{12} \frac{N_2}{K_1} \right) \\ \frac{dN_2}{dt} &= r_2 N_2 \left(1 - \frac{N_2}{K_2} + b_{21} \frac{N_1}{K_2} \right)\end{aligned}$$



Interacting Populations

To analyze these types of models

- Nondimensionalize the system
 - reduce the number of parameters
 - simplify the system
- Solve for equilibria
- Analyze stability of equilibria
- Translate back to determine biological significance



Phase-Plane Techniques

Some definitions of stability

- Stable - if start small distance from equilibrium, remain small distance as $t \rightarrow \infty$
 - Lyapunov stable
 - locally stable
- Asymptotically stable - if start small distance from equilibrium, distance from equilibrium approaches zero as $t \rightarrow \infty$
 - locally asymptotically stable



Phase-Plane Techniques

- Linearization
- Bendixson-Dulac negative criterion
- Hopf bifurcation theorem
- Poincaré-Bendixson theorem
- Routh-Hurwitz Conditions

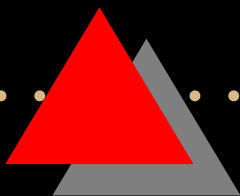


Linearization

Given:

$$\frac{dN}{dt} = F(N, P)$$

$$\frac{dP}{dt} = G(N, P)$$





Linearization

Solve:

$$F(N^*, P^*) = 0$$

$$G(N^*, P^*) = 0$$

to find the equilibria, (N^*, P^*) .

Let:

$$x(t) = N(t) - N^*$$

$$y(t) = P(t) - P^*$$




Linearization

Then linearize about the equilibrium:

$$\frac{dx}{dt} = \left. \frac{\partial F}{\partial N} \right|_{(N^*, P^*)} x + \left. \frac{\partial F}{\partial P} \right|_{(N^*, P^*)} y$$

$$\frac{dy}{dt} = \left. \frac{\partial G}{\partial N} \right|_{(N^*, P^*)} x + \left. \frac{\partial G}{\partial P} \right|_{(N^*, P^*)} y$$

Or:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$




Linearization

Let:

$$J = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

Where J is known as the Jacobian matrix or the community matrix.

We now look for solutions of the form:

$$x(t) = x_0 e^{\lambda t}$$

$$y(t) = y_0 e^{\lambda t}$$



Linearization

Substitute this back into the equations to obtain:

$$\lambda x_0 = a_{11}x_0 + a_{12}y_0$$

$$\lambda y_0 = a_{21}x_0 + a_{22}y_0$$

or

$$\begin{pmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = 0$$



Linearization

From this, we obtain the characteristic equation

$$\lambda^2 - (a_{11} + a_{22}) \lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0$$

Solving for the two roots of λ will determine the stability of the system.



Linearization

- If both roots of λ are real and negative, the equilibrium is a stable node.
- If both roots of λ are real and positive, the equilibrium is an unstable node.
- If the roots of λ are real and of opposite signs, the equilibrium is a saddle point.



Linearization

- If the roots of λ are complex with negative real parts, the equilibrium is a stable focus.
- If the roots of λ are complex with positive real parts, the equilibrium is an unstable focus.
- If the roots of λ are purely complex, the equilibrium of the linearized system is a center, but the original nonlinear system will have a center or a stable or unstable focus depending upon the exact nature of the nonlinear terms.

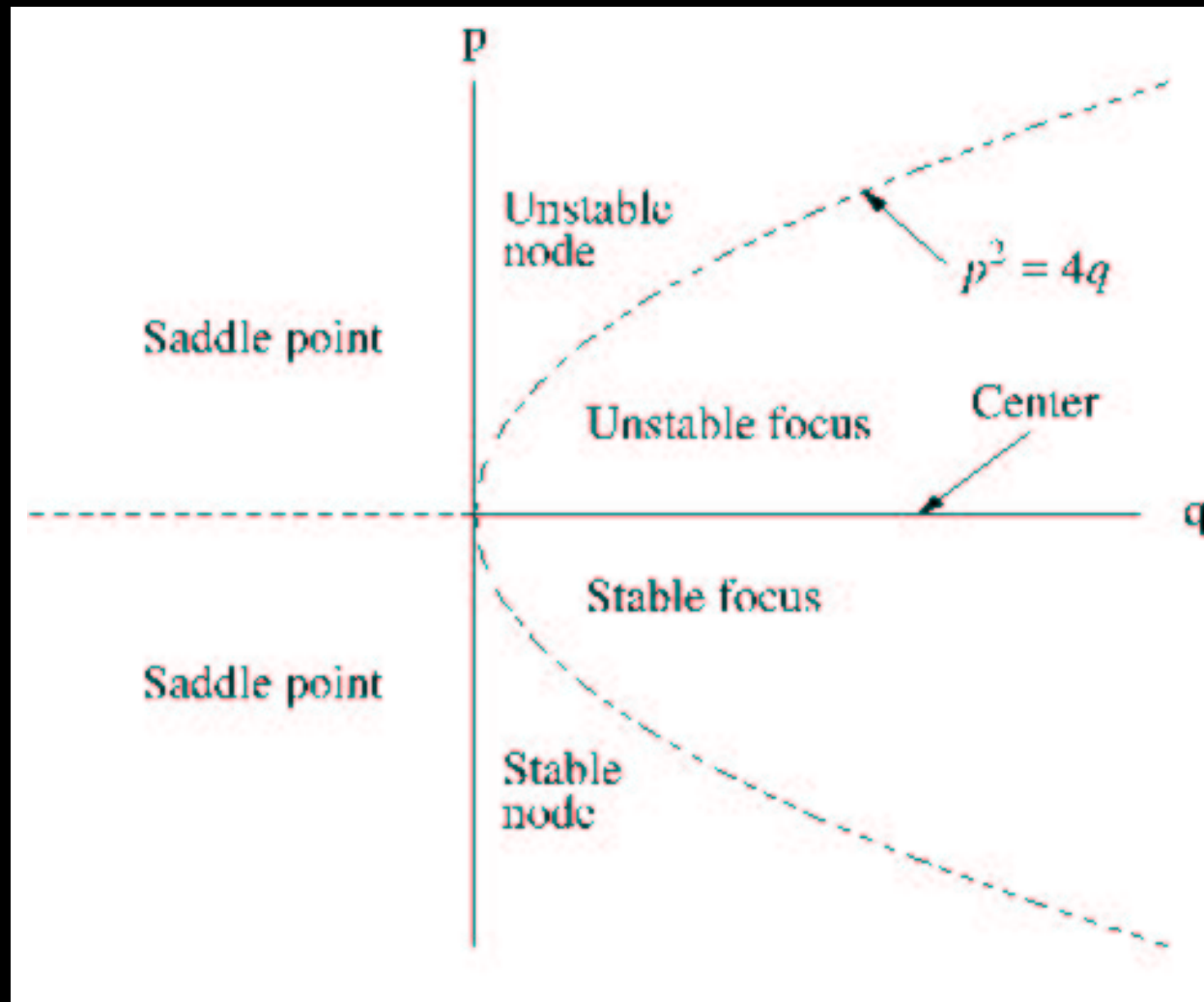


Routh-Hurwitz conditions

Routh-Hurwitz conditions give the necessary and sufficient conditions for all roots of the characteristic polynomial to have negative real roots thus implying asymptotic stability.

$$\begin{aligned} p = \text{Tr} J &= a_{11} + a_{22} < 0 \\ q = \det J &= a_{11}a_{22} - a_{12}a_{21} > 0 \end{aligned}$$

Stability





Bendixson negative criterion

Bendixson's negative criterion

Consider the dynamical system, $\frac{dx}{dt} = F(x, y), \frac{dy}{dt} = G(x, y)$, where F and G are continuously differentiable functions on some simply connected domain $D \subset \mathbb{R}^2$. If $\nabla \cdot (F, G) = \frac{\partial F}{\partial x} + \frac{\partial G}{\partial y}$ is of one sign in D , there cannot be a closed orbit contained within D .



Bendixson-Dulac negative criterion

Additionally, we have the Bendixson's-Dulac's negative criterion.

Let B be a smooth function on $D \subset \mathbb{R}^2$ (with above assumptions). If $\nabla \cdot (BF, BG) = \frac{\partial BF}{\partial x} + \frac{\partial BG}{\partial y}$ is of one sign in D , there cannot be a closed orbit contained within D .



Other theorems

- The Hopf bifurcation theorem gives conditions necessary for the existence of real periodic solutions of a real system of ordinary differential equations.
- Poincaré-Bendixson theorem can also be used to prove the existence of periodic orbits.



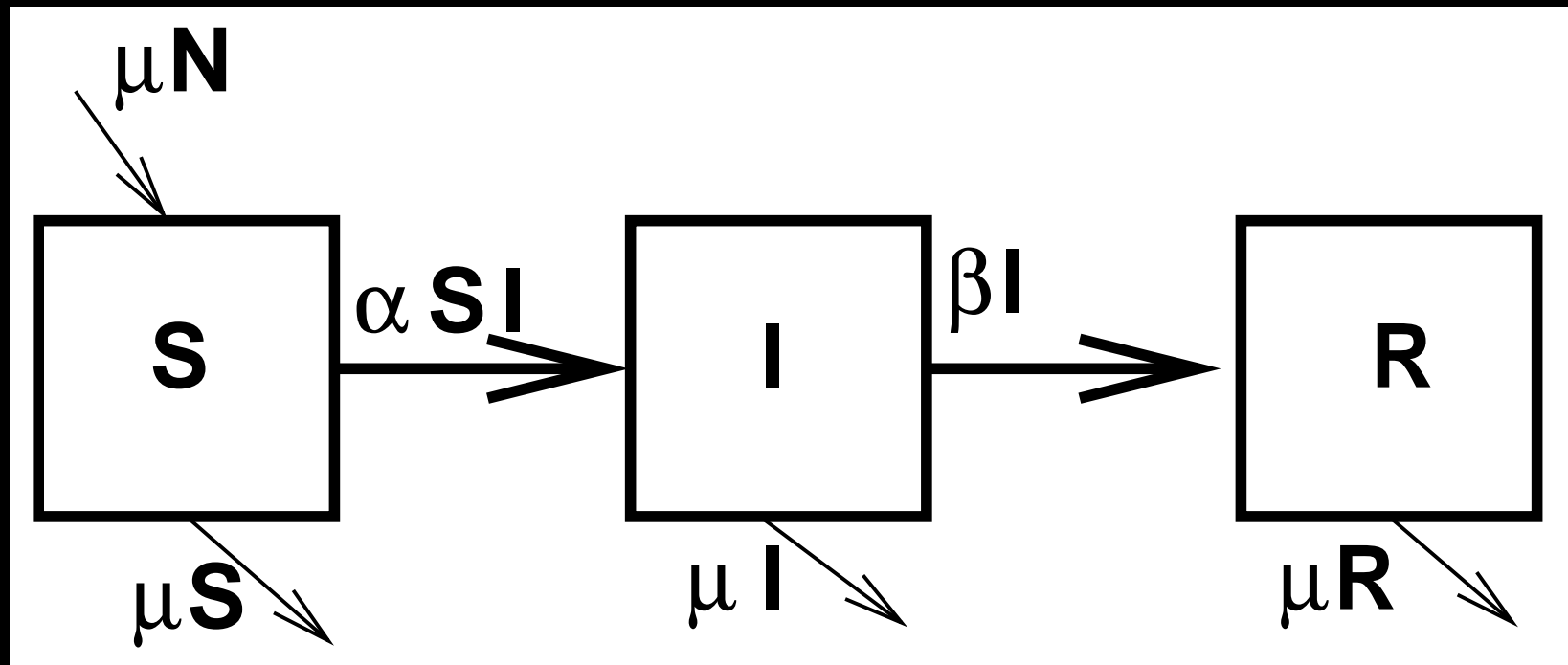
Examples from Epidemiology

Divide population up into distinct classes

- S = Susceptibles
- I = Infectives
- R = Recovered

Classes used depend on disease dynamics

SIR Model





SIR Model - constant population

$$\frac{dS}{dt} = \mu (S + I + R) - \alpha SI - \mu S$$

$$\frac{dI}{dt} = \alpha SI - \beta I - \mu I$$

$$\frac{dR}{dt} = \beta I - \mu R$$

$$N = S + I + R$$

$S(0) = S_0, I(0) = I_0, R(0) = 0$. All parameters are assumed to be positive.



Questions for Epidemic Models

Given all parameters and initial conditions

- Does the infection spread or die out?
- If it does spread, how does it develop with time?
- When will it start to decline?



Equilibria

Note: Since N is a constant, we can solve for only S and I , then if we need R , we can calculate it easily.

$$\frac{dS}{dt} = \mu N - \alpha SI - \mu S = 0$$

$$\frac{dI}{dt} = \alpha SI - \beta I - \mu I = 0$$



Equilibria

Gives two equilibria:

$$S^* = N \quad , \quad I^* = 0$$
$$S^* = \frac{\beta + \mu}{\alpha} \quad , \quad I^* = \frac{\mu (\alpha N - \beta - \mu)}{\alpha (\beta + \mu)}$$

Let

$$F(S, I) = \mu N - \alpha SI - \mu S$$
$$G(S, I) = \alpha SI - \beta I - \mu I$$

Stability

Then

$$\frac{\partial F}{\partial S} = -\alpha I - \mu$$

$$\frac{\partial F}{\partial I} = -\alpha S$$

$$\frac{\partial G}{\partial S} = \alpha I$$

$$\frac{\partial G}{\partial I} = \alpha S - \beta - \mu$$



Stability

First, let's evaluate the stability of $S^* = N, I^* = 0$
The elements of the Jacobian evaluated at this equilibrium are:

$$a_{11} = -\mu$$

$$a_{12} = -\alpha N$$

$$a_{21} = 0$$

$$a_{22} = \alpha N - \beta - \mu$$



Stability

Applying the Routh-Hurwitz conditions:

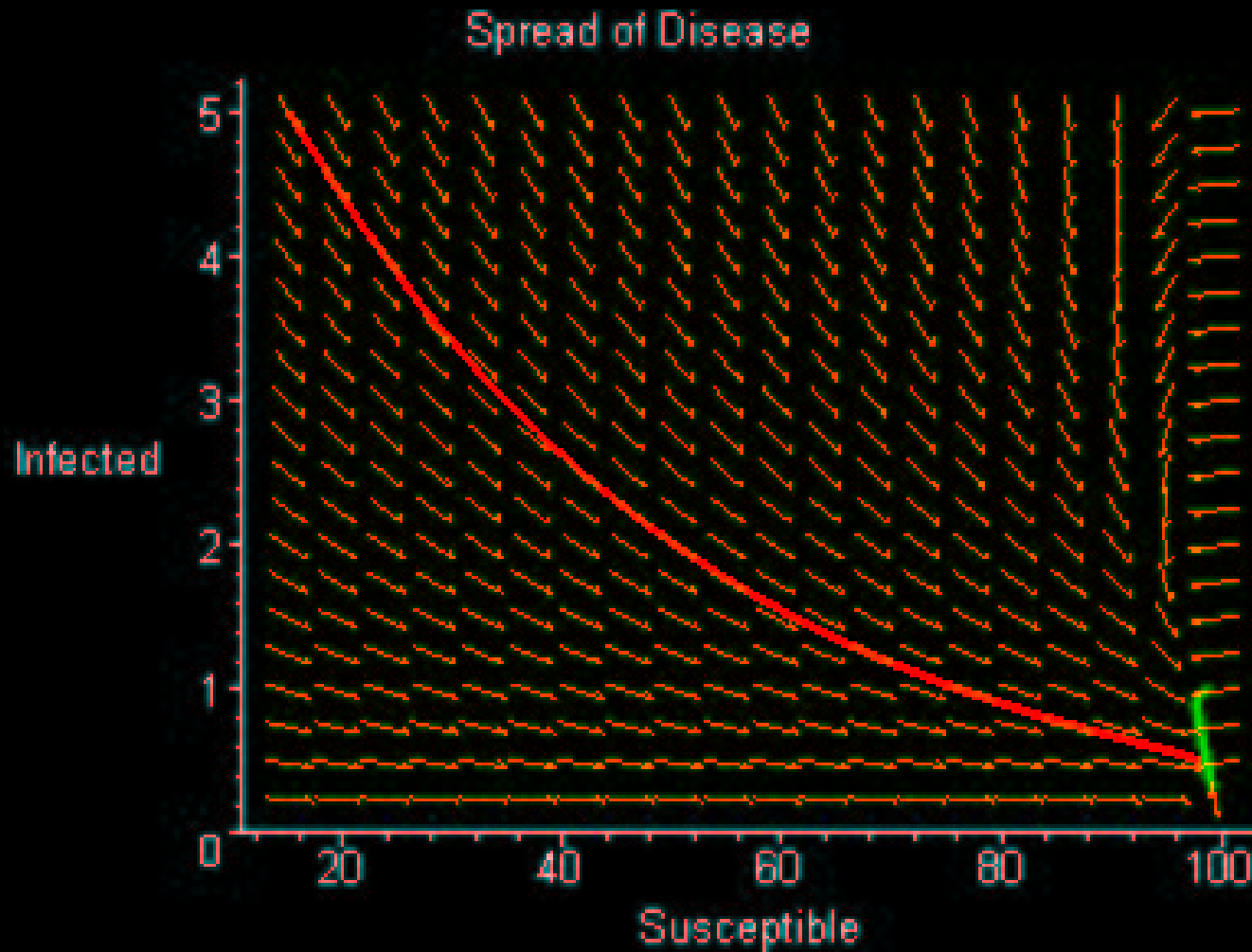
$$\begin{aligned}a_{11} + a_{22} &= -\beta - 2\mu \\ a_{11}a_{22} - a_{12}a_{21} &= \mu(\beta + \mu - \alpha N)\end{aligned}$$

Clearly, $-\beta - 2\mu < 0$

However, $\mu(\beta + \mu - \alpha N) > 0$ only if $\alpha N < \beta + \mu$

Therefore, $S^* = N, I^* = 0$ is asymptotically stable
if $\alpha N < \beta + \mu$

Disease dies out





Stability

Now, let's evaluate the stability of

$$S^* = \frac{\beta + \mu}{\alpha}, I^* = \frac{\mu(\alpha N - \beta - \mu)}{\alpha(\beta + \mu)}$$

The elements of the Jacobian evaluated at this equilibrium are:

$$a_{11} = -\mu - \frac{\mu(\alpha N - \beta - \mu)}{\beta + \mu}$$

$$a_{12} = -\beta - \mu$$

$$a_{21} = \frac{\mu(\alpha N - \beta - \mu)}{\beta + \mu}$$

$$a_{22} = 0$$




Stability

Applying the Routh-Hurwitz conditions:

$$a_{11} + a_{22} = -\mu - \frac{\mu(\alpha N - \beta - \mu)}{\beta + \mu}$$

$$a_{11}a_{22} - a_{12}a_{21} = (\beta + \mu) \left(\frac{\mu(\alpha N - \beta - \mu)}{\beta + \mu} \right)$$

Stability

Both

$$-\mu - \frac{\mu(\alpha N - \beta - \mu)}{\beta + \mu} < 0$$

$$(\beta + \mu) \left(\frac{\mu(\alpha N - \beta - \mu)}{\beta + \mu} \right) < 0$$

are true if $\alpha N > \beta + \mu$



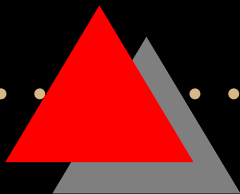
Stability

Therefore,

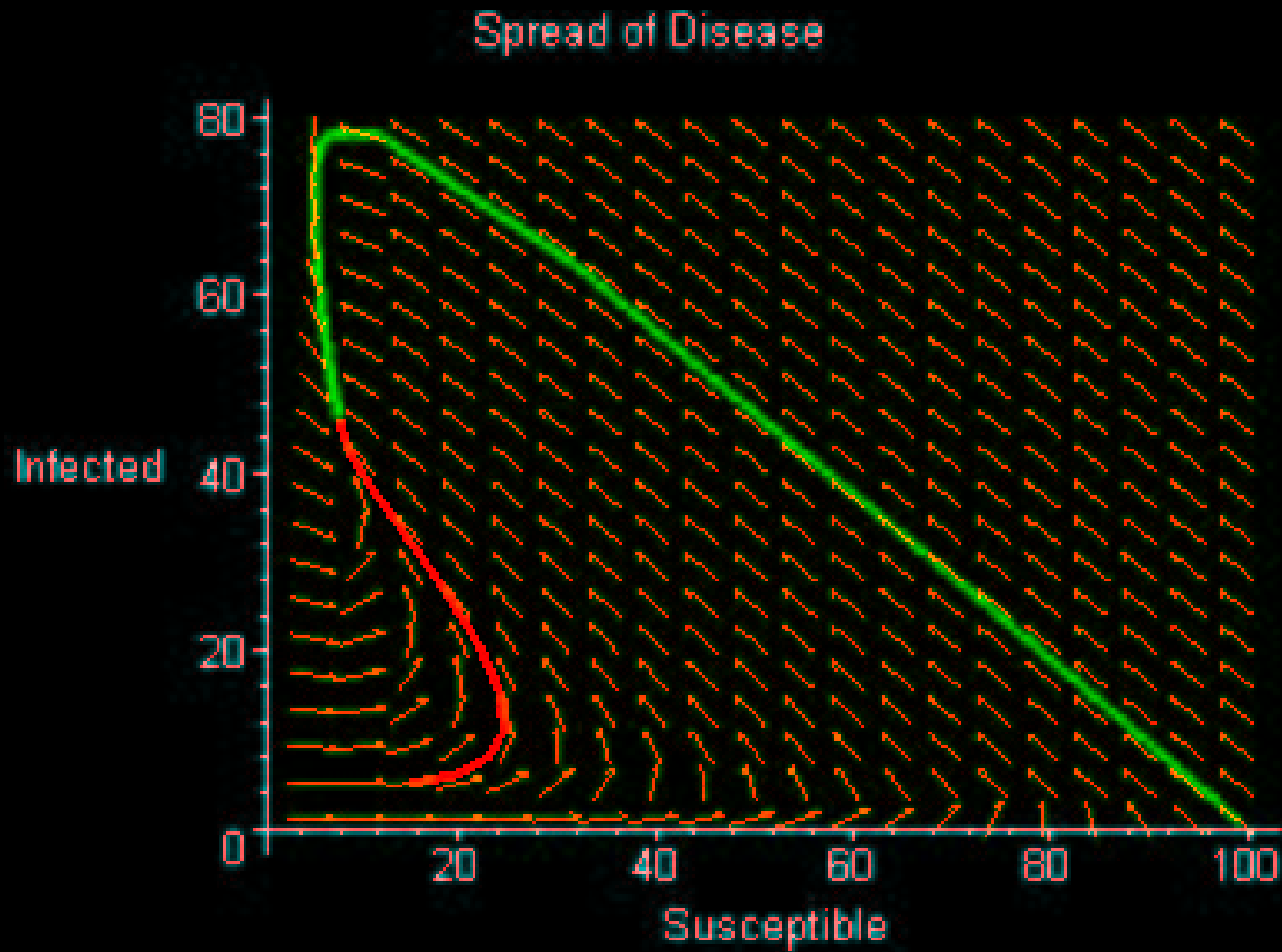
$$S^* = \frac{\beta + \mu}{\alpha} \quad , \quad I^* = \frac{\mu (\alpha N - \beta - \mu)}{\alpha (\beta + \mu)}$$

is asymptotically stable if $\alpha N < \beta + \mu$

But what about limit cycles?



Epidemic





Bendixson's Negative Criteria

Recall, we need

$$\nabla \cdot (F, G) = \frac{\partial F}{\partial S} + \frac{\partial G}{\partial I}$$

to be of one sign in our region of interest, D .

Define D to be all positive values in \mathbb{R}^2 .



Bendixson's Negative Criteria

$$\frac{\partial F}{\partial S} + \frac{\partial G}{\partial I} = -\alpha I - \mu + \alpha S - \beta - \mu$$

So this will be of one sign, negative, if
 $\alpha N < \beta + \mu$ since $S \leq N$.

Therefore there are no limit cycles in D .



R_0 *Basic reproduction rate*

- R_0 is defined to be the number of secondary infections produced by one primary infection in a wholly susceptible population.
- So if $R_0 > 1$, then the disease will spread.
- For SIR model, R_0 is calculated by linearizing the equation for $\frac{dI}{dt}$ about $I = 0$, which we have already done.
- So the criteria for determining if the epidemic will spread is, $R_0 \equiv \frac{\alpha N}{\beta + \mu}$.



Conclusions

There are many other applications of differential equation models in biology. Once a basic set of equations has been developed, there are a number of standard techniques used to analyze the stability of the equations.

We will take time in the lab to explore these and other equations.



Acknowledgements

- Murray, J. D., Mathematical Biology, Springer-Verlag, 1989.
- Kot, Mark, Elements of Mathematical Ecology, Cambridge University Press, 2001.
- Arrowsmith, D.K. and C.M. Place, Ordinary Differential Equations, Chapman and Hall, 1982.

All slides created in \LaTeX using the Prosper class.

34960 - MMB - Mathematical Models in Biology

Coordinating unit: 200 - FME - School of Mathematics and Statistics
Teaching unit: 749 - MAT - Department of Mathematics
Academic year: 2015
Degree: MASTER'S DEGREE IN ADVANCED MATHEMATICS AND MATHEMATICAL ENGINEERING (Syllabus 2010). (Teaching unit Optional)
ECTS credits: 7,5 Teaching languages: English

Teaching staff

Coordinator: ANTONI GUILLAMON GRABOLOSE

Others:

MARTA CASANELLAS RIUS - A
JESUS FERNANDEZ SANCHEZ - A
ANTONI GUILLAMON GRABOLOSE - A
GEMMA HUGUET CASADES - A

Prior skills

- * Proficiency in undergraduate mathematics: calculus, algebra, probability and statistics.
- * Ability to perform basic operations in linear algebra: eigenvalues and eigenvectors, computation of determinants, rank of matrices...
- * Ability to analyze and solve linear differential equations and discuss the stability of simple vector fields.
- * Interest towards biological applications of mathematics and/or previous working experience.

Requirements

- * Basic knowledge of undergraduate mathematics: calculus, ordinary differential equations, linear algebra, probability and statistics.
- * First course in ordinary differential equations: linear differential equations, qualitative and stability theory and numerical simulation.
- * Basic knowledge of computer programming for scientific purposes.
- * Courses and all the bibliography will be in English.

Degree competences to which the subject contributes

Specific:

1. RESEARCH. Read and understand advanced mathematical papers. Use mathematical research techniques to produce and transmit new results.
2. MODELLING. Formulate, analyse and validate mathematical models of practical problems by using the appropriate mathematical tools.
3. CALCULUS. Obtain (exact or approximate) solutions for these models with the available resources, including computational means.
4. CRITICAL ASSESSMENT. Discuss the validity, scope and relevance of these solutions; present results and defend conclusions.

Transversal:

5. SELF-DIRECTED LEARNING. Detecting gaps in one's knowledge and overcoming them through critical self-appraisal. Choosing the best path for broadening one's knowledge.
6. EFFICIENT ORAL AND WRITTEN COMMUNICATION. Communicating verbally and in writing about learning outcomes, thought-building and decision-making. Taking part in debates about issues related to the own field of

34960 - MMB - Mathematical Models in Biology

specialization.

7. THIRD LANGUAGE. Learning a third language, preferably English, to a degree of oral and written fluency that fits in with the future needs of the graduates of each course.

8. TEAMWORK. Being able to work as a team player, either as a member or as a leader. Contributing to projects pragmatically and responsibly, by reaching commitments in accordance to the resources that are available.

9. EFFECTIVE USE OF INFORMATION RESOURCES. Managing the acquisition, structure, analysis and display of information from the own field of specialization. Taking a critical stance with regard to the results obtained.

Teaching methodology

The course will be structured in five blocks each consisting of a brief introduction through theoretical lectures, the development of a short project in groups and wrap-up sessions with oral presentations, discussion and complementary lectures.

The central part intended to develop the short project will held at the computer lab. The SAGE computing environment will be used, with interfaces to Python, R and C if necessary.

Learning objectives of the subject

This course is an introduction to the most common mathematical models in biology: in populations dynamics, ecology, physiology, sequence analysis and phylogenetics. At the end of the course the student should be able to:

- * Understand and discuss basic models of dynamical systems of biological origin, in terms of the parameters.
- * Model simple phenomena, analyze them (numerically and/or analytically) and understand the effect of parameters.
- * Understand the diversity of mechanisms and the different levels of modelization of physiological activity.
- * Obtain and analyze genomic sequences of real biological species and databases containing them.
- * Use computer software for gene prediction, alignment and phylogenetic reconstruction.
- * Understand different gene prediction, alignment and phylogenetic reconstruction methods.
- * Compare the predictions given by the models with real data.
- * Communicate results in interdisciplinary teams.

Study load

Total learning time: 187h 30m	Hours large group:	60h	32.00%
	Self study:	127h 30m	68.00%

34960 - MMB - Mathematical Models in Biology

Content

Mathematical models in Genomics	Learning time: 75h Theory classes: 12h Laboratory classes: 12h Self study : 51h
Description: 1. Brief introduction to genomics (genome, gen structure, genetic code...). Genome databases online. 2. Phylogenetics: Markov models of molecular evolution (Jukes-Cantor, Kimura, Felsenstein hierarchy...), phylogenetic trees, branch lengths. Phylogenetic tree reconstruction (distance and character based methods). 3. Genomics: Markov chains and Hidden Markov models for gene prediction. Tropical arithmetics and Viterbi algorithm. Forward and Expectation-Maximization algorithms. 4. Multiple sequence alignment: dynamical programming, tropical arithmetics and Pair-HMMs	
Mathematical Models in Neurophysiology	Learning time: 56h 15m Theory classes: 9h Laboratory classes: 9h Self study : 38h 15m
Description: 1) Membrane biophysics. 2) Excitability and Action potentials: The Hodgkin-Huxley model, the Morris-Lecar model, integrate & fire models. 3) Bursting oscillations. 4) Synaptic transmission and dynamics.	
Models of Population Dynamics	Learning time: 37h 30m Theory classes: 6h Laboratory classes: 6h Self study : 25h 30m
Description: 1. Modelling interactions among populations with differential equations. Stability and bifurcations. 2. One-dimensional discrete models. Chaos in biological systems. 3. Paradigms of population dynamics in current research.	

34960 - MMB - Mathematical Models in Biology

Biological networks	Learning time: 18h 45m Theory classes: 3h Laboratory classes: 3h Self study : 12h 45m
Description: 1. Complex networks in biology. 2. Networks of neurons.	

Qualification system

50%: Each of the five blocks will give a part (10%) of the qualification, based on the performance on the short-projects.

20%: Overall evaluation of the participation, interest and proficiency evinced along the course.

30%: Final exam aiming at validating the acquisition of the most basic concepts of each block.

34960 - MMB - Mathematical Models in Biology

Bibliography

Basic:

Allman, Elizabeth S.; Rhodes, John A. Mathematical models in biology: an introduction. Cambridge: Cambridge University Press, 2004. ISBN 9780521819800.

Istas, Jacques. Mathematical modeling for the life sciences [on line]. Berlin: Springer, 2005 Available on: <<http://dx.doi.org/10.1007/3-540-27877-X>>. ISBN 354025305X.

Murray, J.D. Mathematical biology [on line]. 3rd ed. Berlin: Springer, 2002 Available on: <<http://link.springer.com/book/10.1007/b98868> (v. 1) <http://link.springer.com/book/10.1007/b98869> (v. 2)>. ISBN 978-0-387-95223-9.

Pachter, Lior; Sturmfels, Bernd. Algebraic statistics for computational biology. Cambridge: Cambridge University Press, 2005. ISBN 0521857007.

Keener, James P.; Sneyd, James. Mathematical physiology. Vol 1. 2nd ed. New York: Springer Verlag, 2009. ISBN 9780387758466.

Izhikevich, Eugene M. Dynamical systems in neuroscience : the geometry of excitability and bursting. Cambridge: MIT Press, 2007. ISBN 0262090430.

Ermentrout, Bard G.; Terman, David H. Mathematical foundations of neuroscience. New York: Springer, 2010. ISBN 978-0-387-87708-2.

Complementary:

Stein, William A. [et al.]. Sage mathematics software (Version 4.4.2) [on line]. 2010 [Consultation: 23/11/2012]. Available on: <<http://www.sagemath.org/>>.

Durbin, Richard [et al.]. Biological sequence analysis : probabilistic models of proteins and nucleic acids. Cambridge: Cambridge University Press, 1998. ISBN 0521629713.

Feng, Jianfeng. Computational neuroscience : a comprehensive approach [on line]. Boca Raton: Chapman & Hall/CRC, 2004 [Consultation: 23/11/2012]. Available on: <http://nba.uth.tmc.edu/homepage/cnjclub/2007summer/renart_chapter.pdf>.

Felsenstein, J. PHYLIP [on line]. [Consultation: 23/11/2012]. Available on: <<http://evolution.genetics.washington.edu/phylip.html>>.

European Bioinformatics Institute; Wellcome Trust Sanger Institute. Ensembl project [on line]. [Consultation: 23/11/2012]. Available on: <<http://www.ensembl.org>>.

6 Applications of Continuous Models to Population Dynamics

Each organic being is striving to increase in a geometrical ratio . . . each at some period of its life, during some season of the year, during each generation or at intervals has to struggle for life and to suffer great destruction. . . . The vigorous, the healthy, and the happy survive and multiply.

Charles R. Darwin. (1860). *On the Origin of Species by Means of Natural Selection*, D. Appleton and Company, New York, chap. 3.

The growth and decline of populations in nature and the struggle of species to predominate over one another has been a subject of interest dating back through the ages. Applications of simple mathematical concepts to such phenomena were noted centuries ago. Among the founders of mathematical population models were Malthus (1798), Verhulst (1838), Pearl and Reed (1908), and then Lotka and Volterra, whose works were published primarily in the 1920s and 1930s.

The work of Lotka and Volterra, who arrived independently at several models including those for predator-prey interactions and two-species competition, had a profound effect on the field now known as *population biology*. They were among the first to study the phenomena of interacting species by making a number of simplifying assumptions that led to nontrivial but tractable mathematical problems. Since their pioneering work, many other notable contributions were made. Among these is the work of Kermack and McKendrick (1927), who addressed the problem of outbreaks of epidemics in a population.

Today, students of ecology and population biology are commonly taught such classical models as part of their regular biology curriculum. Critics of these historical models often argue that certain biological features, such as environmental effects, chance random events, and spatial heterogeneity to mention a few, were ig-

nored. However, the importance of these models stems not from realism or the accuracy of their predictions but rather from the simple and fundamental principles that they set forth; the propensity of predator-prey systems to oscillate, the tendency of competing species to exclude one another, the threshold dependence of epidemics on population size are examples.

While appreciation of the Lotka-Volterra models in the biological community is mixed, it is nevertheless interesting to note that in subtle yet important ways they have helped to shape certain research directions in current biology. As demonstrated by the Nicholson-Bailey model of Chapter 3, a model does not have to be accurate to serve as a helpful diagnostic tool. We shall later discuss more specific ways in which the unrealistic predictions of simple models have led to new empirical as well as theoretical progress.

The classic population biology models serve several purposes in this text. Aside from being interesting in their own right, models of two interacting species or of epidemics in a fixed population are ideal illustrations of the techniques and concepts outlined in Chapters 4 and 5. The models also demonstrate how the predictions of a model change when slight alterations are made in the equations or in values of the critical quantities that appear in them. Finally, the fact that these models are fairly simple allows us to assess critically the various assumptions and their consequences.

As in previous discussions, we set the stage by a brief discussion of models for single-species populations. (Section 4.1 introduced this topic; here we somewhat broaden the context.) In Sections 6.2 and 6.3 the Lotka-Volterra predator-prey and species competition models are described and then analyzed. The story of Volterra's initiation to this biological area is well known. This Italian mathematician became interested in the area of population biology through conversations with a colleague, U. d'Ancona, who had observed a puzzling biological trend. During World War I, commercial fishing in the Adriatic Sea fell to rather low levels. It was anticipated that this would cause a rise in the availability of fish for harvest. Instead, the population of commercially valuable fish declined on average while the number of sharks, which are their predators, increased. The two populations were also perceived to fluctuate.

Volterra suggested a somewhat naive model to describe the predator-prey interactions in the fish populations and was thereby able to explain the trends d'Ancona had observed. As we shall see, the model's basic prediction is that predators tend to overrespond to increases in the population of their prey. This can give rise to oscillations in the populations of both species.

Because natural communities are composed of numerous interacting species no two of which can be entirely isolated from the rest, theoretical tools for dealing with larger systems are often required. The Routh-Hurwitz criteria and the methods of qualitative stability are thus briefly outlined in Sections 6.4 and 6.6. For rapid coverage of this chapter, these sections may be omitted without loss of continuity. In Sections 6.6 and 6.7 we study models for the spread of an epidemic in a population and then explore certain consequences of the policy of vaccinating against disease-causing agents.

Since the scope of this material is vast, a thorough documentation of sources is

impossible. There are numerous recent reviews (for example, May, 1973). An excellent companion to this chapter is Van der Vaart (1983), which contains historical, biological, and mathematical details on certain topics and which uses an instructive and guided approach. [See also Braun (1979, 1983).] All of these sources have been used repeatedly in putting together the material for this chapter.

6.1 MODELS FOR SINGLE-SPECIES POPULATIONS

Two examples of ODEs modeling continuous single-species populations have already been encountered and analyzed in Section 4.1. To summarize, these are

1. *Exponential growth (Malthus, 1798):*

$$\frac{dN}{dt} = rN, \quad (1a)$$

$$\text{Solution: } N(t) = N_0 e^{rt}. \quad (1b)$$

2. *Logistic growth (Verhulst, 1838):*

$$\frac{dN}{dt} = r \left(1 - \frac{N}{K} \right) N, \quad (2a)$$

$$\text{Solution: } N(t) = \frac{N_0 K}{N_0 + (K - N_0)e^{-rt}}. \quad (2b)$$

$N_0 = N(0)$ = the initial population. (See Figure 6.1.)

To place both of the above into a somewhat broader context we proceed from a more general assumption, namely that for an isolated population (no migration) the rate of growth depends on population density. Therefore

$$\frac{dN}{dt} = f(N). \quad (3)$$

This approach is based on an instructive summary by Lamberson and Biles (1981), which should be consulted for further details.

Observe that for equations (1a) and (2a) the function f is the polynomial

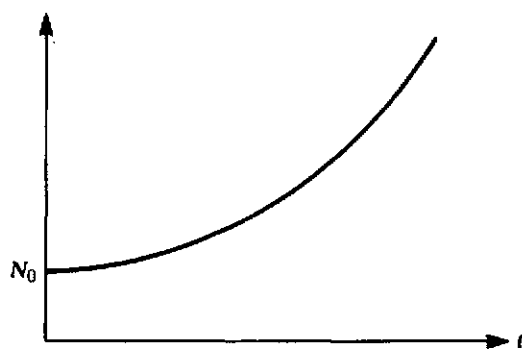
$$f(N) = a_0 + a_1 N + a_2 N^2$$

where $a_0 = 0$; for equation (1a) $a_1 = r$ and $a_2 = 0$; for equation (2a) $a_1 = r$ and $a_2 = -r/K$. More generally, it is possible to write an infinite power (Taylor) series for f if it is sufficiently smooth:

$$f(N) = \sum_{n=0}^{\infty} a_n N^n = a_0 + a_1 N + a_2 N^2 + a_3 N^3 + \cdots$$

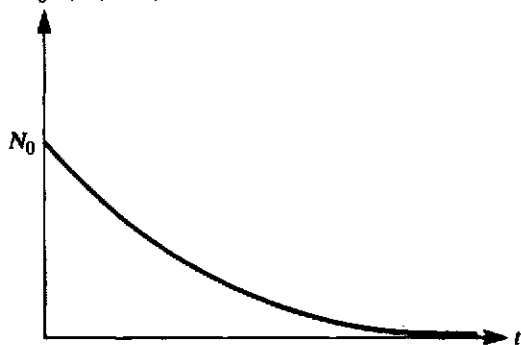
Thus any growth function may be written as a (possibly infinite) polynomial (see Lamberson and Biles, 1981).

$$N(t) = N_0 e^{rt}, \quad (r > 0)$$



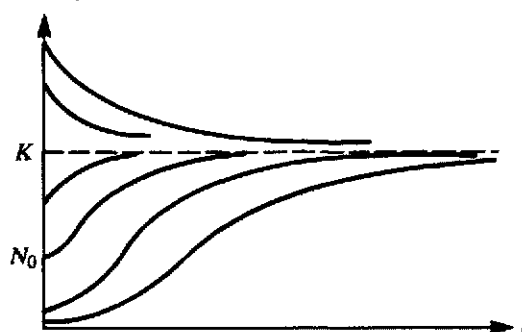
(a)

$$N(t) = N_0 e^{rt}, \quad (r < 0)$$



(b)

$$N(t) = \frac{N_0 K}{N_0 + (K - N_0)e^{-rt}}, \quad (r > 0)$$



(c)

Figure 6.1 Changes in population size $N(t)$ predicted by two models for single-species growth:

Exponential growth with (a) $r > 0$, (b) $r < 0$, and (c) logistic growth. See equations (1a) and (2a).

About (3) we require that $f(0) = 0$ to dismiss the possibility of *spontaneous generation*, the production of living organisms from inanimate matter. (See also Hutchinson, 1978 for this *Axiom of Parenthood*: every organism must have parents.) In any growth law this is equivalent to

$$\left. \frac{dN}{dt} \right|_{N=0} = f(0) = 0,$$

so that we may assume that

$$\begin{aligned} a_0 &= 0, \\ \frac{dN}{dt} &= a_1 N + a_2 N^2 + a_3 N^3 + \cdots \\ &= N(a_1 + a_2 N + a_3 N^2 + \cdots), \\ &= Ng(N). \end{aligned} \tag{4}$$

The polynomial $g(N)$ is called the *intrinsic growth rate* of the population.

Now we examine more closely several specific growth models, including those given in equations (1a) and (2a).

Malthus Model

This can be viewed as the simplest form of equation (4) in which the coefficients of $g(N)$ are $a_1 = r$ and $a_2 = a_3 = \cdots = 0$. As noted before, this model predicts exponential growth if $r > 0$ and exponential decline if $r < 0$.

Logistic Growth

To correct the prediction that a population can grow indefinitely at an exponential rate, consider a nonconstant intrinsic growth rate $g(N)$. The logistic growth model is perhaps the simplest extension of equation (1a). It can be explained by any of the following comments.

Formal mathematical justification

Equation (2a) makes use of more terms in the (possibly infinite) series for $f(N)$ and is thus more faithful to the true population growth rate.

Density-dependent growth rate

Equation (2a) takes the form of equation (4), where

$$g(N) = r \left(1 - \frac{N}{K} \right).$$

It is essentially the simplest rule in which the intrinsic growth rate g depends on the

population density (in a *linear decreasing* relationship). It thus accounts for a decreasing per capita growth rate as population size increases.

Carrying capacity

From equation (2a) we observe that

$$\frac{dN}{dt} = 0 \quad (N = K).$$

Thus $N = K$ is a steady state of the logistic equation. It is easy to establish that this steady state is stable; note in particular that

$$\frac{dN}{dt} > 0 \quad (N < K),$$

$$\frac{dN}{dt} < 0 \quad (N > K).$$

The constant K can represent the carrying capacity of the environment for the species. See also Section 4.1 for a derivation of (2a) based on nutrient consumption.

Intraspecific competition

The fact that individuals compete for food, habitat, and other limited resources means that such an increase in the net population mortality may be observed under crowded conditions. Such effects are most pronounced when there are frequent *encounters* between individuals. Equation (2a) can be written in the form

$$\frac{dN}{dt} = rN - \frac{r}{K} N^2.$$

The second term thus depicts a mortality proportional to the rate of paired encounters.

The solution of equation (2a) given by (2b) can be obtained in a relatively straightforward calculation (see problem 5 of Chapter 4). Aside from Gause's work on yeast cultures (Section 4.1), such models have been applied to a variety of populations including humans (Pearl and Reed, 1920), microorganisms (Slobodkin, 1954), and other species. See Lamberson and Biles (1981) for examples and references.

Allee Effect

A further direct extension of equations (1) and (2) is an assumption of the form

$$g(N) = a_1 + a_2 N + a_3 N^2.$$

Provided $a_2 > 0$, and $a_3 < 0$, one obtains the *Allee effect*, which represents a population that has a maximal intrinsic growth rate at intermediate density. This effect may stem from the difficulty of finding mates at very low densities.

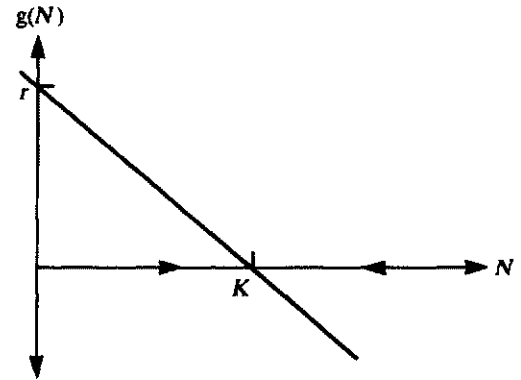
Figure 6.2 is an example of a density-dependent form of $g(N)$ that depicts the

Allee effect. Its general character can be summarized by the inequalities

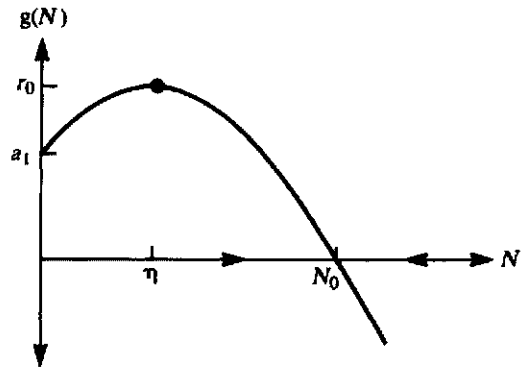
$$g'(N) > 0 \quad (N < \eta),$$

$$g'(N) < 0 \quad (N > \eta),$$

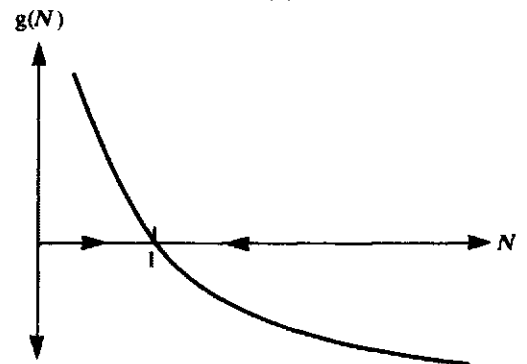
where η is the density for optimal reproduction.



(a)



(b)



(c)

Figure 6.2 A comparison between three types of density-dependent intrinsic growth rates $g(N)$. (a) Logistic growth decreases linearly with density (or population size). (b) In the Allee effect the rate of reproduction is maximal at intermediate densities. (c) The Gompertz law shows a negative logarithmic dependence of growth rate on population size. See text for details.

The simplest example of an Allee effect would be

$$g(N) = r_0 - \alpha(N - \eta)^2, \quad (\eta < \sqrt{r_0/\alpha}). \quad (5)$$

Notice that this inverted parabola, shown in Figure 6.2(b), intersects the axis at $r_0 - \alpha\eta^2$, has a maximum of r_0 when $N = \eta$, and drops below 0 when

$$N > N_0 = \eta + \sqrt{r_0/\alpha}.$$

Thus for densities above N_0 , the population begins to decline. From the curve in Figure 6.2(a) we can deduce that $N = N_0$ is a stable equilibrium for the population. (N_0 is an equilibrium point because $g(N_0) = 0$; it is stable because $g'(N_0) < 0$.)

In equation (5) we assumed that $a_1 = r_0 - \alpha\eta^2$, $a_2 = 2\alpha\eta$, and $a_3 = -\alpha$.

Other Assumptions; Gompertz Growth in Tumors

Yet a fourth growth law that frequently appears in models of single-species growth is the *Gompertz law* (introduced in Chapter 4), which is used mainly for depicting the growth of solid tumors. The problems of dealing with a complicated geometry and with the fact that cells in the interior of a tumor may not have ready access to nutrients and oxygen are simplified by assuming that the growth rate declines as the cell mass grows. Three equivalent versions of this growth rate are as follows:

$$\frac{dN}{dt} = \lambda e^{-\alpha N}, \quad (6a)$$

$$\frac{dN}{dt} = \gamma N, \quad \frac{d\gamma}{dt} = -\alpha\gamma, \quad (6b)$$

$$\frac{dN}{dt} = -\kappa N \ln N. \quad (6c)$$

See Figure 6.2(c). In (6c) we can identify the intrinsic growth rate as

$$g(N) = -\kappa \ln N.$$

Since $\ln N$ is undefined at $N = 0$, this relation is not valid for very small populations and cannot be considered a direct extension of any of the previous growth laws. It is, however, a popular model in clinical oncology. (See Braun, 1979, sec. 1.8; Newton, 1980; Aroesty et al., 1973.) Biological interpretations for these equations are discussed in problem 7. Considering their relatively simple form, the predictions of any of the Gompertz equations agree remarkably well with the data for tumor growth. (See Aroesty et al., 1973, or Newton, 1980, for examples.)

A valid remark about most of the models for population growth is that they are at best gross simplifications of true events and often are used simply as an expedient fit to the data. To be more realistic one needs a greater mathematical sophistication. For example, in Chapter 13 we will see that partial differential equations provide a

more powerful way to deal with age-dependent growth, fecundity, or mortality rates. Equations such as (3) or (6) are frequently used by modelers as a convenient first approach to complicated situations and thus are quite useful provided their limitations are not ignored.

6.2 PREDATOR-PREY SYSTEMS AND THE LOTKA-VOLTERRA EQUATIONS

The fact that predator-prey systems have a tendency to oscillate has been observed for well over a century. The Hudson Bay Company, which traded in animal furs in Canada, kept records dating back to 1840. In these records, oscillations in the populations of lynx and its prey the snowshoe hare are remarkably regular (see Figure 6.3).

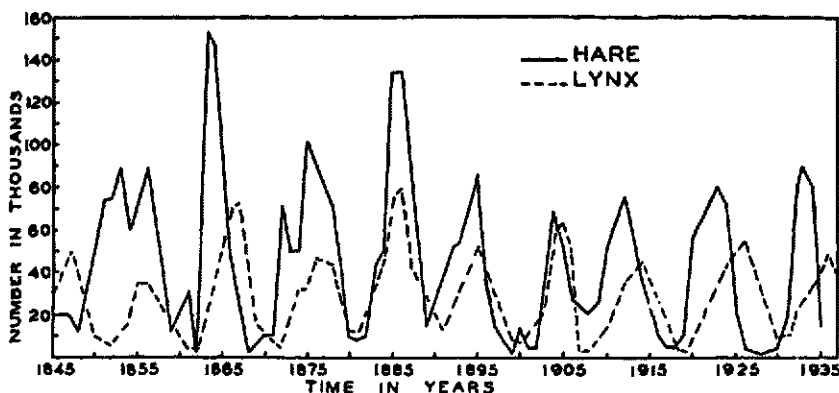


Figure 6.3 Records dating back to the 1840s kept by the Hudson Bay Company. Their trade in pelts of the snowshoe hare and its predator the lynx reveals that the relative abundance of the two

species undergoes dramatic cycles. The period of these cycles is roughly 10 years.
[From E. P. Odum (1953), fig. 39.]

In this section we explore a model for predator-prey interactions that Volterra proposed to explain oscillations in fish populations in the Mediterranean. To reconstruct his line of reasoning and arrive at the equations independently, let us list some of the simplifying assumptions he made:

1. Prey grow in an unlimited way when predators do not keep them under control.
2. Predators depend on the presence of their prey to survive.
3. The rate of predation depends on the likelihood that a victim is encountered by a predator.
4. The growth rate of the predator population is proportional to food intake (rate of predation).

Taking the simplest set of equations consistent with these assumptions, Volterra wrote down the following model:

$$\frac{dx}{dt} = ax - bxy, \quad (7a)$$

$$\frac{dy}{dt} = -cy + dxy, \quad (7b)$$

where x and y represent prey and predator populations respectively; the variables can represent, for example, biomass or population densities of the species. To acquaint ourselves with this model we proceed by answering several questions. First let us consider the meaning of parameters a , b , c , and d and of each of the four terms on the RHS of the equations.

The net growth rate a of the prey population when predators are absent is a positive quantity (with dimensions of 1/time) in accordance with assumption 1. The net death rate c of the predators in the absence of prey follows from assumption 2. The term xy approximates the likelihood that an encounter will take place between predators and prey given that both species move about randomly and are uniformly distributed over their habitat.

The form of this encounter rate is derived from the *law of mass action* that, in its original context, states that the rate of molecular collisions of two chemical species in a dilute gas or solution is proportional to the product of the two concentrations (see Chapter 7). We should bear in mind that this simple relationship may be inaccurate in describing the subtle interactions and motion of organisms. An encounter is assumed to decrease the prey population and increase the predator population by contributing to their growth. The ratio b/d is analogous to the efficiency of predation, that is, the efficiency of converting a unit of prey into a unit of predator mass.

Further practice in linear stability techniques given in Chapter 5 can be revealing:

It is clear that two possible steady states of equation (7) exist:

$$(\bar{x}_1, \bar{y}_1) = (0, 0) \quad \text{and} \quad (\bar{x}_2, \bar{y}_2) = \left(\frac{c}{d}, \frac{a}{b}\right).$$

Their stability properties are determined by the methods given in Chapter 5.

The Jacobian of this system is

$$J = \begin{pmatrix} a - by & -bx \\ dy & dx - c \end{pmatrix}_{(x, y)},$$

for steady state 1

$$J = \begin{pmatrix} a & 0 \\ 0 & -c \end{pmatrix},$$

for steady state 2

$$J = \begin{pmatrix} 0 & -bc/d \\ da/b & 0 \end{pmatrix},$$

eigenvalues are

$$\lambda_1 = a, \quad \lambda_2 = -c, \quad \lambda_{1,2} = \pm \sqrt{ca}i.$$

Thus (\bar{x}_1, \bar{y}_1) is a saddle. Thus (\bar{x}_2, \bar{y}_2) is a center¹.

From the analysis of this model we arrive at a number of somewhat counterintuitive results. First, notice that the steady-state level of prey is independent of its own growth rate or mortality; rather, it depends on parameters associated with the predator ($x_2 = c/d$). A similar result holds for steady-state levels of the predator ($y_2 = a/b$). It is the particular *coupling* of the variables that leads to this effect. To paraphrase, the presence of predator ($y \neq 0$) means that the available prey has to just suffice to make growth rate due to predation, dx , equal predator mortality c for a steady predator population to persist. Similarly, when prey are present ($x \neq 0$), predators can only keep them under control when prey growth rate a and mortality due to predation, by , are equal. This helps us to understand the steady-state equations.

A second result (see problem 10) is that the steady state (\bar{x}_2, \bar{y}_2) is neutrally stable (a center). The eigenvalues of $J(\bar{x}_2, \bar{y}_2)$ are pure imaginary and the steady state is not a spiral point. See problem 10. Note that the off-diagonal terms, $-bc/d$ and da/b , are of opposite sign (since the influence of each species on the other is opposite) and that the diagonal terms evaluated at (\bar{x}_2, \bar{y}_2) are zero. Stability analysis predicts oscillations about the steady state (\bar{x}_2, \bar{y}_2) . The factor \sqrt{ca} governs the frequency of these oscillations, so that larger prey reproduction or predator mortality (which means a greater turnover rate) result in more rapid cycles. A complete phase-plane diagram of the predator-prey system (7) can be arrived at with minimal further work. See Figure 6.4(a).

To gain deeper understanding of the neutral stability of system (7) we will examine a slight variant in which prey populations have the property of self-regulation. Assuming logistic prey growth, equations (7a, b) become

$$\frac{dx}{dt} = \frac{ax(K-x)}{K} - bxy, \quad (8a)$$

$$\frac{dy}{dt} = -cy + dxy. \quad (8b)$$

This leads to steady-state values

$$(\bar{x}_2, \bar{y}_2) = \left(\frac{c}{d}, \frac{a}{b} - \frac{ca}{dbK} \right),$$

1. To be more accurate we must include the possibility that this steady state could be a spiral point since the system is nonlinear. (See Section 5.10 for comments.) In problem 10 we will demonstrate that this option can be dismissed for the predator-prey equations.

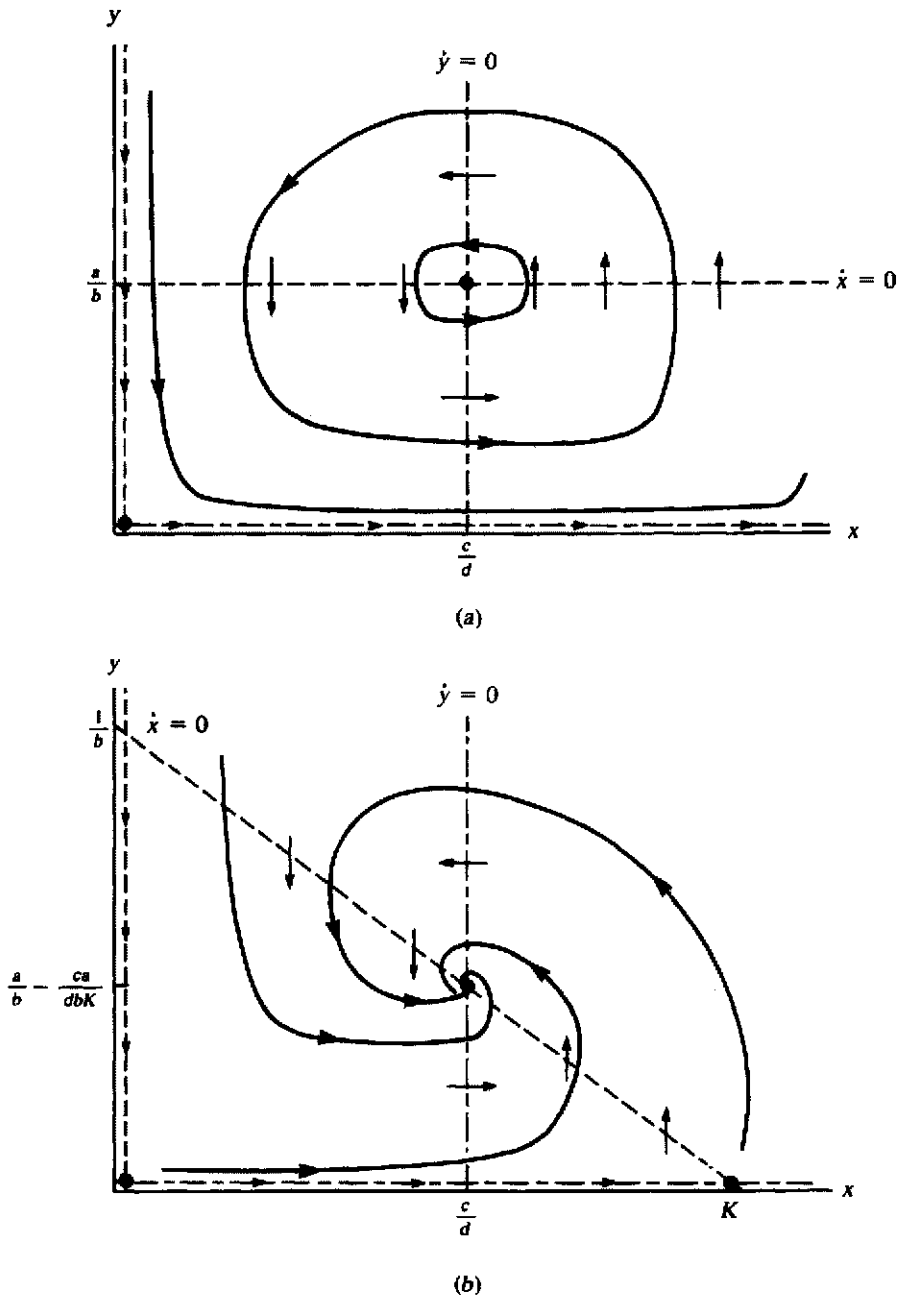


Figure 6.4 (a) The Lotka-Volterra equations (7a,b) predict neutral stability at the steady state $(c/d, a/b)$. (b) When the prey grow logistically [as in

equation (8a)], the steady state becomes a stable spiral with somewhat depressed predator population levels.

and the Jacobian is then

$$\mathbf{J} = \begin{pmatrix} \frac{-ac}{dK} & \frac{-bc}{d} \\ d\left(\frac{a}{b} - \frac{ca}{dbK}\right) & 0 \end{pmatrix}.$$

(The condition $1 > c/dK$ must be satisfied so that the steady-state predator level y is positive.) Now $\text{Tr } \mathbf{J} = -ac/dK$ is always negative, and $\det \mathbf{J} = bc\bar{y}_2$ is positive, so that the steady state is always stable. In other words, its neutral stability has been lost. In problem 14 you are asked to investigate whether oscillations accompany the return to steady state after a perturbation.

The lesson to be learned from this example is that a relatively minor change in equations (7a, b) has a major influence on the predictions. In particular, this means that neutral stability, and thus also the oscillations that accompany a neutrally stable steady state, tend to be somewhat ephemeral. This is a serious criticism of the realism of the Lotka-Volterra model.

Taking a somewhat more philosophical approach, we could argue that the Lotka-Volterra model serves a useful purpose precisely because it is so delicately balanced between stability and instability. We could use this model together with minor variants to test out a set of assumptions and so identify stabilizing and destabilizing influences. Following are some of the frequently suggested alterations. It is a relatively easy task to understand what effects such changes have on the stability of the equilibrium. More theoretical results on stable cycles due to Kolmogorov (1936) and others (briefly mentioned below) are recommended for further independent exploration and will be discussed in detail in Chapter 8.

For Further Study

Stable cycles in predator-prey systems

The main objection to the Lotka-Volterra model is that its cycles are only neutrally stable. What additional features are necessary to yield stable oscillations? As we shall see in Chapter 8, stable oscillations (usually called limit cycles) are closed trajectories that attract nearby flow in the phase plane. Kolmogorov (1936) investigated conditions on the general predator-prey system

$$\frac{dx}{dt} = xf(x, y),$$

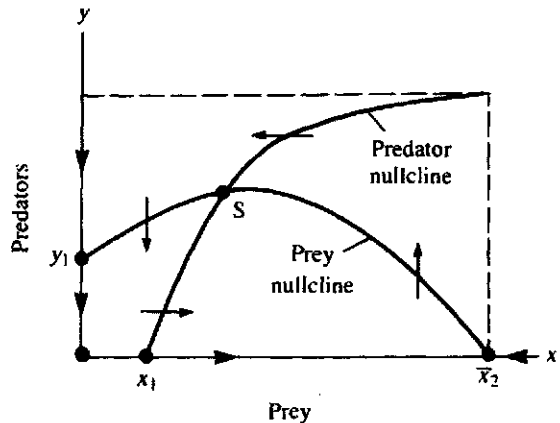
$$\frac{dy}{dt} = yg(x, y),$$

that would lead to such solutions. The functions f and g are assumed to satisfy several relations consistent with the nature of predator-prey systems:

$$\begin{aligned} \partial f / \partial x &< 0 \text{ (for large } x), & \partial g / \partial x &> 0, \\ \partial f / \partial y &< 0, & \partial g / \partial y &< 0. \end{aligned}$$

An interpretation of these is left as an exercise. Additional conditions (for example, Coleman, 1978; May 1973) are equivalent to the nullcline geometry shown in Figure 6.5 (Rosenzweig, 1969). It can be proved that when the steady state S is unstable, any trajectory winding out of its vicinity approaches a stable limit cycle that is trapped somewhere inside the rectangular region. See Chapter 8 for further details.

Figure 6.5 With a set of conditions given by Kolmogorov (1936), the phase plane for a predator-prey system has a nullcline geometry that gives rise to stable (limit cycle) oscillations. See Chapter 8.



Other modifications of Volterra's equations

Other assumptions which have been made over the years to modify Volterra's equations are listed below. Details about their effect can be found in May (1973) and in the references as follows.

1. **Density dependence:** More realistic prey growth-rate assumptions in which a is replaced by a density-dependent function f :

$$f(x) = r \left(1 - \frac{x}{\kappa} \right) \quad \text{Pielou (1969, pp. 19-21),}$$

$$f(x) = r \left[\left(\frac{\kappa}{x} \right)^{-g} - 1 \right] \quad (1 \geq g > 0) \quad \text{Rosenzweig (1971),}$$

$$f(x) = r \left(\frac{\kappa}{x} - 1 \right) \quad \text{Schoener (1973).}$$

2. **Attack rate:** More realistic rates of predation where the term bxy is replaced by a term in which the attack capacity of predators is a limited one. Terms replacing bxy in equation (7a) are:

$$ky(1 - e^{-\alpha x}) \quad \text{Ivlev (1961),}$$

$$\frac{kxy}{x + D} \quad \text{Holling (1965),}$$

$$kx^g \quad (1 \geq g > 0) \quad \text{Rosenzweig (1971),}$$

$$\frac{kxy^2}{x^2 + D^2} \quad \text{Takahashi (1964).}$$

6.3 POPULATIONS IN COMPETITION

When two or more species live in proximity and share the same basic requirements, they usually compete for resources, habitat, or territory. Sometimes only the strongest prevails, driving the weaker competitor to extinction. (This is the *principle of competitive exclusion*, a longstanding concept in population biology.) One species wins because its members are more efficient at finding or exploiting resources, which leads to an increase in population. Indirectly this means that a population of competitors finds less of the same resources and cannot grow at its maximal capacity.

In the following model, proposed by Lotka and Volterra and later studied empirically by Gause (1934), the competition between two species is depicted without direct reference to the resources they share. Rather, it is assumed that the presence of each population leads to a depression of its competitor's growth rate. We first give the equations and then examine their meanings and predictions systematically. See also Braun (1979, sec. 4.10) and Pielou (1969, sec. 5.2) for further discussion of this model.

The Lotka-Volterra model for species competition is given by the equations

$$\frac{dN_1}{dt} = r_1 N_1 \frac{\kappa_1 - N_1 - \beta_{12} N_2}{\kappa_1}, \quad (9a)$$

$$\frac{dN_2}{dt} = r_2 N_2 \frac{\kappa_2 - N_2 - \beta_{21} N_1}{\kappa_2}, \quad (9b)$$

where N_1 and N_2 are the population densities of species 1 and 2. Again we proceed to understand the equations by addressing several questions:

1. Suppose only species 1 is present. What has been assumed about its growth? What are the meanings of the parameters r_1 , κ_1 , r_2 , and κ_2 ?
2. What kind of assumption has been made about the effect of competition on the growth rate of each species? What are the parameters β_{12} and β_{21} ?

To answer these questions observe the following:

1. In the absence of a competitor ($N_2 = 0$) the first equation reduces to the logistic equation (2a). This means that the population of species 1 will stabilize at the value $N_1 = \kappa_1$ (its carrying capacity), as we have already seen in Section 6.1.
2. The term $\beta_{21} N_1$ in equation (9a) can be thought of as the contribution made by species 2 to a decline in the growth rate of species 1. β_{12} is the per capita decline (caused by individuals of species 2 on the population of species 1).

The next step will be to study the behavior of the system of equations. The task will again be divided into a number of steps, including (1) identifying steady states, (2) drawing nullclines, and (3) determining stability properties as necessary in putting together a complete phase-plane representation of equation (9) using the

Nullclines are just all point sets that satisfy one of the following equations:

$$\frac{dN_1}{dt} = 0 \quad \text{or} \quad \frac{dN_2}{dt} = 0.$$

1. From equation (9a) we arrive at the N_1 nullclines:

$$N_1 = 0 \quad \text{and} \quad \kappa_1 - N_1 - \beta_{12}N_2 = 0,$$

2. Whereas equation (9b) leads to the N_2 nullclines:

$$N_2 = 0 \quad \text{and} \quad \kappa_2 - N_2 - \beta_{21}N_1 = 0.$$

To simplify the notation slightly, we shall refer to these lines as L_{1a} , L_{1b} , L_{2a} , and L_{2b} respectively. Notice that L_{1a} and L_{2a} are just the N_2 and N_1 axes respectively, whereas L_{1b} and L_{2b} intersect the axes as follows:

L_{1b} goes through $(0, \kappa_1/\beta_{12})$ and $(\kappa_1, 0)$.

L_{2b} goes through $(0, \kappa_2)$ and $(\kappa_2/\beta_{21}, 0)$.

methods given in Chapter 5. (For practice, it is advisable to attempt this independently before continuing to the procedure in the box.)

It follows that the points $(0, 0)$, $(\kappa_1, 0)$, and $(0, \kappa_2)$ are always steady states. These correspond to three distinct situations:

$(0, 0)$ = both species absent,

$(\kappa_1, 0)$ = species 2 absent and species 1 at its carrying capacity κ_1 ,

$(0, \kappa_2)$ = species 1 absent and species 2 at its carrying capacity.

There is a fourth possible steady-state value that corresponds to coexistence of the two species. (We leave the computation of this steady state as an exercise.)

Proceeding to the second stage, we sketch the nullcline curves on a phase plane. If you have already attempted this independently, you may have hesitated slightly because numerous situations are possible. Figure 6.6 illustrates four distinct possibilities, all of them correct. In order to choose any one of the four cases we must make some assumptions about the relative magnitudes of κ_2 and κ_1/β_{12} , and of κ_1 and κ_2/β_{21} . The cases shown in Figure 6.6 correspond to the following situations:

$$\text{case 1: } \frac{\kappa_2}{\beta_{21}} > \kappa_1 \quad \text{and} \quad \kappa_2 > \frac{\kappa_1}{\beta_{12}},$$

$$\text{case 2: } \kappa_1 > \frac{\kappa_2}{\beta_{21}} \quad \text{and} \quad \frac{\kappa_1}{\beta_{12}} > \kappa_2,$$

$$\text{case 3: } \kappa_1 > \frac{\kappa_2}{\beta_{21}} \quad \text{and} \quad \kappa_2 > \frac{\kappa_1}{\beta_{12}},$$

$$\text{case 4: } \frac{\kappa_2}{\beta_{21}} > \kappa_1 \quad \text{and} \quad \frac{\kappa_1}{\beta_{12}} > \kappa_2.$$

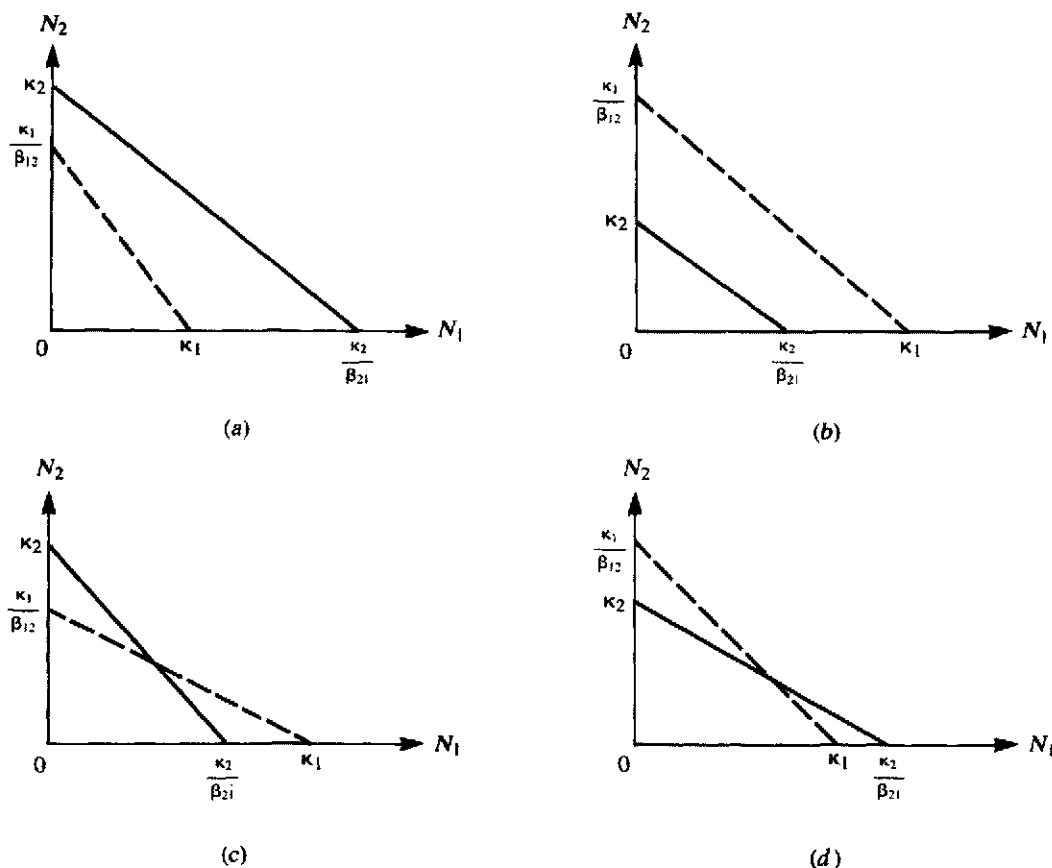


Figure 6.6 Four possible cases corresponding to four choices in the relative positions of nullclines of equations (9): (a) $\kappa_2/\beta_{21} > \kappa_1$ and $\kappa_2 > \kappa_1/\beta_{12}$; (b) $\kappa_1 > \kappa_2/\beta_{21}$ and $\kappa_1/\beta_{12} > \kappa_2$; (c) $\kappa_1 > \kappa_2/\beta_{21}$ and $\kappa_2 > \kappa_1/\beta_{12}$; (d) $\kappa_1/\beta_{12} > \kappa_2$ and $\kappa_2/\beta_{21} > \kappa_1$.

In problem 15 the reader is asked to interpret these inequalities within the biological context of the problem.

Our next step will be to identify the steady states of equations (9a,b) in Figure 6.6(a–d). By drawing arrows on the nullclines we will also indicate the directions of flow in the N_1N_2 plane for each of the four cases shown. To do this, one can combine geometric reasoning with results of analysis. We recall that steady states are located at the intersections of two nullclines (which must be of opposite types). It helps to remember that L_{1b} (the line $N_1 = 0$) is an N_1 nullcline; it is simply the N_2 axis. Thus the point at which any N_2 nullcline meets the N_2 axis will be a steady state. It is evident that this happens at $(0, 0)$ as well as at $(0, \kappa_2)$. By similar reasoning we find that $(\kappa_1, 0)$ is at the intersection of two (opposite type) nullclines. A fourth steady state occurs only when L_{1b} and L_{2b} intersect, as is true in (c) and (d) of Figure 6.6.

To sketch arrows on the N_1 and N_2 nullclines, recall that the directions of flow

on these are parallel to the N_2 and N_1 axes respectively. Arrows have been put in for cases 1 and 4 in Figure 6.7, with case 2 and 3 left as an exercise. Notice that once the flow along the N_1 and N_2 axes is drawn the rest of the picture can be completed by preserving the continuity of flow. (See remarks in Section 5.5.) For a more pedestrian approach, we can use equations (9a,b) to tabulate the directions associated with several points in the plane.

At this stage the problem is practically solved; with the directions of flow determined on the nullclines, we can draw sensible phase-plane pictures in only one distinct way for each case. For example, it should be evident in case 1 that for any starting value of (N_1, N_2) provided $N_2 \neq 0$, the populations eventually converge to the steady state on the N_2 axis. (To see this, notice that there is no other exit from the region bounded by the two slanted lines L_{1b} and L_{2b} in case 1; moreover, all flows pass through this region.) In case 4, any point within the two triangular regions must eventually converge to the steady state at the intersection of L_{1b} and L_{2b} . (What can be said about other regions of the plane in case 4?)

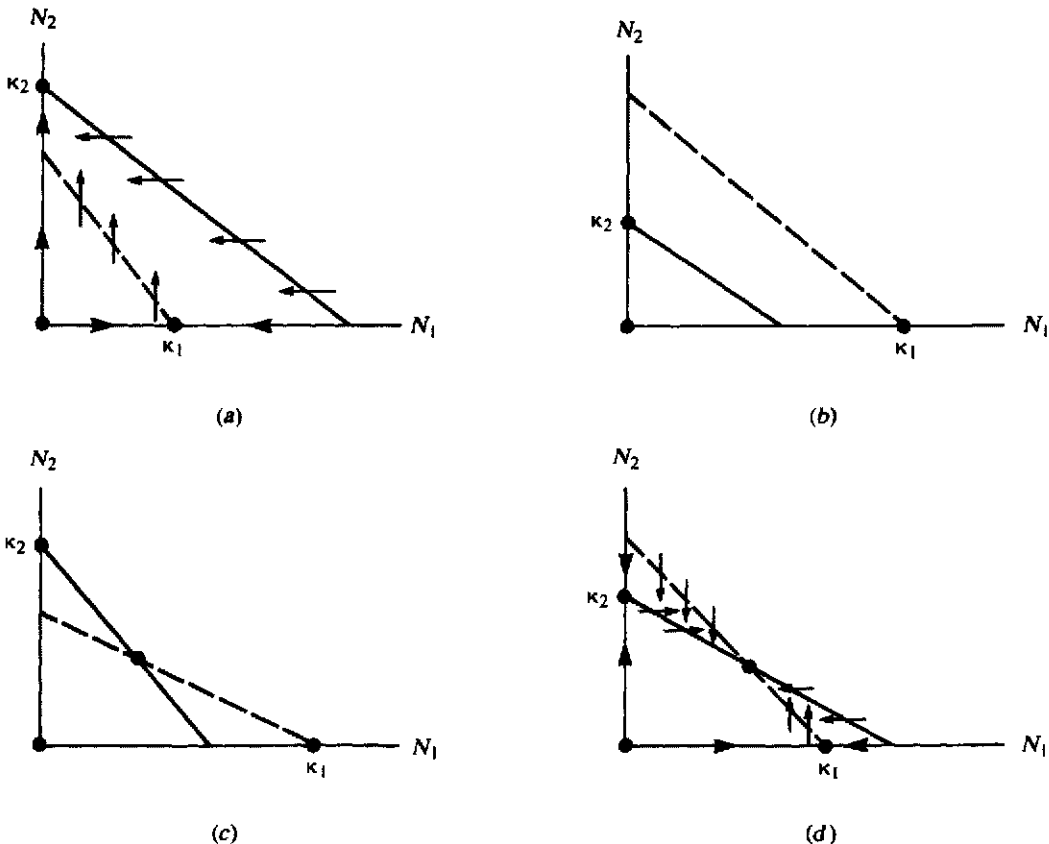


Figure 6.7 Steady states of equation (9) shown as heavy dots at the intersections of the N_1 nullclines (dashed lines) and the N_2 nullclines (solid lines).

(a–d) correspond to cases 1–4 shown in Figure 6.6 and described in text.

As a somewhat optional final step, we can confirm the conjectured flow by determining what happens close to steady-state values, using the linearization procedures outlined in Chapter 5 [see problem 15(c)]. By carrying out this analysis it can be shown that the outcome of competition is as follows:

case 1: only $(0, \kappa_2)$ is stable,

case 2: only $(\kappa_1, 0)$ is stable,

case 3: both $(0, \kappa_1)$ and $(0, \kappa_2)$ are stable,

case 4: only the steady state given by the expression in problem 15(b) is stable.

With the combined information above, the qualitative pictures in Figure 6.8 can be confirmed, and the mathematical steps in understanding the model are complete. It is now necessary to make a biological interpretation of the result. Part of this is left as a problem for the reader. A rather clear prediction is that in three out of the four cases, competition will lead to extinction of one species. Only in case 4 does the interaction result in coexistence, and then at population levels *below* the normal carrying capacities.

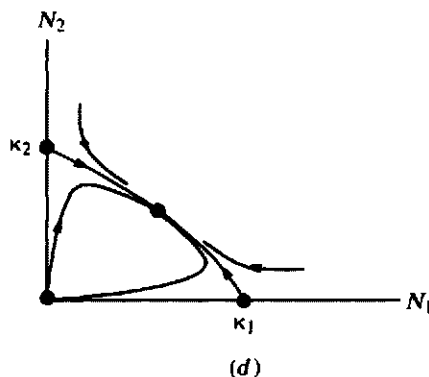
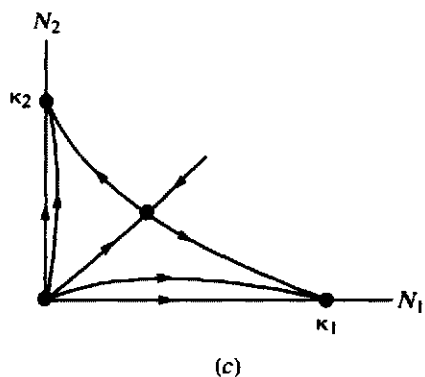
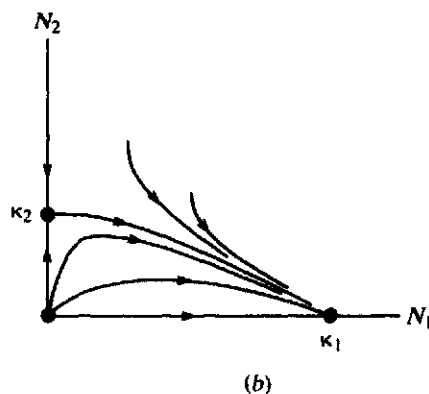
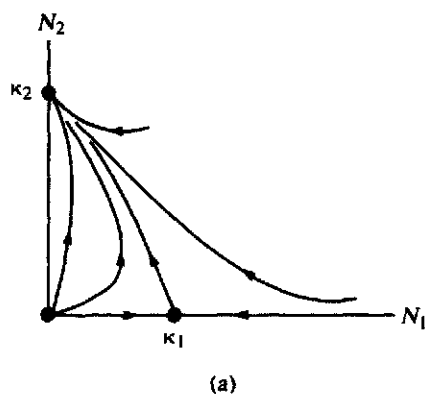


Figure 6.8 The final phase-plane behavior of solutions to equations (9a,b). (a–d) correspond to

cases 1–4 in Figure 6.6. See text for details.

A small change in the format of the inequalities for cases 1 through 4 will reveal how the *intensity of competition*, which is represented by the β parameters, influences the outcome. To make things more transparent, suppose the carrying capacities are equal ($\kappa_1 = \kappa_2$). Conditions 1 to 4 can be written as follows:

1. $\beta_{21} < 1$ and $\beta_{12} > 1$.
2. $\beta_{21} > 1$ and $\beta_{12} < 1$.
3. $\beta_{21} > 1$ and $\beta_{12} > 1$.
4. $\beta_{21} < 1$ and $\beta_{12} < 1$.

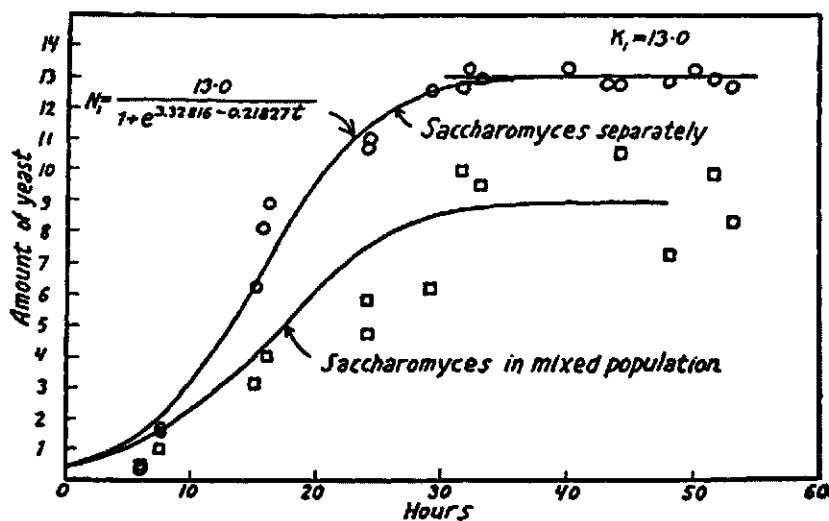
From this, observe that in cases 1, 2, and 3, one or both species are aggressive in competing with their adversary (that is, at least one β is large). In case 4, for which coexistence is obtained, β_{21} and β_{12} are both small, indicating that competition is less intense.

An accepted biological fact is that species very similar in habits, size, and/or feeding preferences tend to compete more strongly for resources when confined to the same habitat (Roughgarden, 1979). For example, species of fish that have similar mouth parts and thus seek the same type of food would overlap in their resource utilization and, thus be more aggressive competitors than those that feed differently. With this observation, a prediction of the model is that similar species in the same habitat will not coexist. (This is a popular version of the principle of competitive exclusion.)

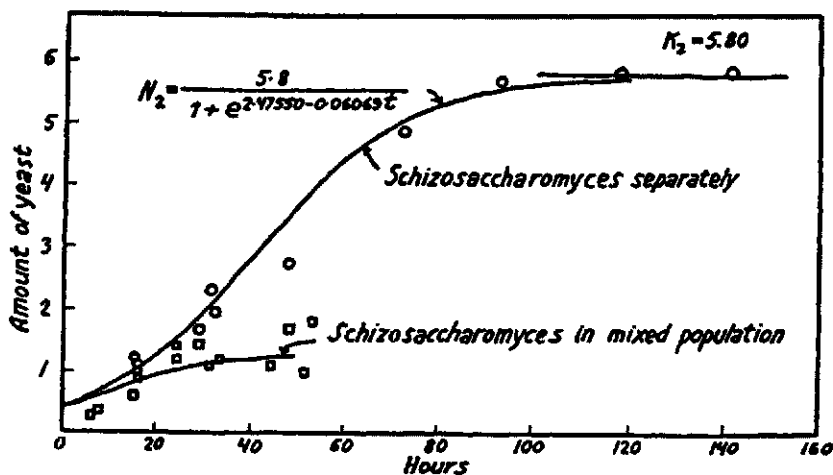
Recent research directions in population biology have focused on questions raised by this principle. Because ecosystems frequently consist of many competitors that appear to vie for common resources, the predictions of this simple model have reshaped some preconceptions about coexistence and species interactions. It has become more challenging to discover the numerous ways competitive exclusion can be foiled.

The model ignores spatial distributions of species and variations in both space and time of the significant quantities as well as many other subtle influences (such as the effects of predation on one of the species). This points to numerous possible effects that could come into play in permitting species to live and share a common habitat. In fact, it is now recognized that species are distributed in a patchy way, rather than uniformly partitioning their habitat so that competition tends to diminish somewhat. A time-sharing arrangement with succession of species or seasonal variability can effect a similar result. Other factors include gradual evolution of differing traits (*character displacement*) to minimize competition, and more complex multi-species interactions in which predation mediates competition. Observations of such special cases are abundant in the current biological literature. Sources for additional readings are Whitaker and Levin (1975) and a forthcoming monograph on theoretical ecology by Simon Levin (Cornell University). Chapter 21 of Roughgarden (1979) also makes for good reading on the competition model and its implications.

There are recent extensions of the competition model to handle n species. Luenberger (1979, sec. 9.5) gives an excellent presentation. A good discussion of the principle of competitive exclusion is given in Armstrong and McGehee (1980). A number of other contributors have included T. G. Hallam, T. C. Gard, R. M. May, H. I. Freedman, P. Waltman, and J. Hofbauer.



(a)



(b)

Figure 6.9 Growth of (a) *Saccharomyces cerevisiae* and (b) *Schizosaccharomyces kephir* in original experiments by Gause (1932). The organisms were grown separately (open circles) as well as in a mixed culture

containing both (open rectangles). From Gause, G. F. (1932), *Experimental studies on the struggle for existence*. I. Mixed population of two species of yeast. *J. Exp. Biol.*, 9, Figures 2 and 3.

Gause: Empirical Tests of the Species-Competition Model

In his book, *The Struggle for Existence*, Gause (1934) describes a series of laboratory experiments in which two yeast species, *Saccharomyces cerevisiae* and *Schizosaccharomyces kephir* were grown separately and then paired in a mixed population (Figure

6.9). Using results shown in Figure 6.9 (a and b), he was able to estimate the following values for parameters in equation (9):

$$\begin{aligned} r_1 &= 0.21827, & K_1 &= 13.0, & \beta_{12} &= 3.15, \\ r_2 &= 0.06069, & K_2 &= 5.8, & \beta_{21} &= 0.439. \end{aligned}$$

See problems 33 and 34 for some details and analysis, and Gause for a very readable summary of these and other experiments.

6.4 MULTIPLE-SPECIES COMMUNITIES AND THE ROUTH-HURWITZ CRITERIA

Now that we have spent some time mastering the techniques of linear stability theory, it seems discouraging to realize that the elegance and simplicity of phase-plane methods apply only to two-species systems. In this section we briefly touch on methods for gaining insight into models for k species interacting in a community, where $k > 2$.

The models we have seen thus far take the form

$$\frac{dN_1}{dt} = f(N_1, N_2), \quad (10a)$$

$$\frac{dN_2}{dt} = g(N_1, N_2). \quad (10b)$$

More generally a system comprised of k species with populations N_1, N_2, \dots, N_k would be governed by k equations:

$$\begin{aligned} \frac{dN_1}{dt} &= f_1(N_1, N_2, \dots, N_k), \\ \frac{dN_2}{dt} &= f_2(N_1, N_2, \dots, N_k), \\ &\vdots \\ \frac{dN_k}{dt} &= f_k(N_1, N_2, \dots, N_k). \end{aligned} \quad (11)$$

Since it is cumbersome to carry this longhand version, one often sees the shorthand notation

$$\frac{dN_i}{dt} = f_i(N_1, N_2, \dots, N_k) \quad (i = 1, 2, \dots, k), \quad (12)$$

or, better still, the vector notation

$$\frac{d\mathbf{N}}{dt} = \mathbf{F}(\mathbf{N}), \quad (13)$$

for $\mathbf{N} = (N_1, N_2, \dots, N_k)$, $\mathbf{F} = (f_1, f_2, \dots, f_k)$, where each of the functions f_1, f_2, \dots, f_k may depend on all or some of the species populations N_1, N_2, \dots, N_k .

We shall now suppose that it is possible to solve the equation (or set of equations)

$$\mathbf{F}(\mathbf{N}) = 0, \quad (14)$$

so as to identify one (or possibly several) steady-state points, $\bar{\mathbf{N}} = (N_1, N_2, \dots, N_k)$, satisfying $\mathbf{F}(\bar{\mathbf{N}}) = 0$. The next step, as a diligent reader might have guessed, would be to determine stability properties of this steady solution. While the idea is essentially identical to previous linear stability analysis, a slightly greater sophistication may be necessary to extract an answer. Let us see why this is true.

In linearizing equation (13) we find, as before, the Jacobian of $\mathbf{F}(\mathbf{N})$. This is often symbolized

$$\mathbf{J} = \frac{\partial \mathbf{F}}{\partial \mathbf{N}}(\bar{\mathbf{N}}). \quad (15)$$

Recall that this really means

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f_1}{\partial N_1} & \frac{\partial f_1}{\partial N_2} & \dots & \frac{\partial f_1}{\partial N_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial N_1} & \frac{\partial f_k}{\partial N_2} & \dots & \frac{\partial f_k}{\partial N_k} \end{pmatrix}_{\bar{\mathbf{N}}} \quad (16)$$

so that \mathbf{J} is now a $k \times k$ matrix. Population biologists frequently refer to \mathbf{J} as the *community matrix* (see Levins, 1968). Eigenvalues λ of this matrix now satisfy

$$\det(\mathbf{J} - \lambda \mathbf{I}) = 0. \quad (17)$$

Thinking of what this means, you should arrive at the conclusion that λ must satisfy a characteristic equation of the form

$$\lambda^k + a_1 \lambda^{k-1} + a_2 \lambda^{k-2} + \dots + a_k = 0. \quad (18)$$

If you find this baffling, you may wish to verify this with a 3×3 matrix, that is, by evaluating

$$\det \begin{pmatrix} a - \lambda & b & c \\ d & e - \lambda & f \\ g & h & i - \lambda \end{pmatrix}.$$

(The result is a cubic polynomial.) In general, the characteristic equation is a *polynomial* whose degree k is equal to the number of species interacting. Although for $k = 2$ the quadratic characteristic equation is easily solved, for $k > 2$ this is no longer true.

While we are unable in principle to *find* all eigenvalues, we can still obtain information about their magnitudes. Suppose $\lambda_1, \lambda_2, \dots, \lambda_k$ are all (known) eigenvalues of the linearized system

$$\frac{d\mathbf{N}}{dt} = \mathbf{J} \cdot \mathbf{N}. \quad (19)$$

What must be true about these eigenvalues so that the steady state \bar{N} would be stable? Recall that they must *all* have negative real parts since close to the steady states each of the species populations can be represented by a sum of exponentials in $\lambda_i t$ as follows:

$$N_i = \bar{N}_i + a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t} + \cdots + a_k e^{\lambda_k t}. \quad (20)$$

(This is a direct generalization of Section 5.6.) If one or more eigenvalues have positive real parts, $N_i - \bar{N}_i$ will be an increasing function of t , meaning that N_i will not return to its equilibrium value \bar{N}_i . Thus the question of stability of a steady state can be settled if it can be determined whether or not all eigenvalues $\lambda_1, \dots, \lambda_k$ have negative real parts. (Contrast this with stability conditions for difference equations.) This can be done without actually solving for these eigenvalues by checking certain criteria. Recall that in the two-species case we derived conditions on quantities β and γ (which were, respectively, the trace and the determinant of the Jacobian) that ensured eigenvalues with negative real parts. For $k > 2$ these conditions are known as the *Routh-Hurwitz criteria* and are summarized in the box.

The Routh-Hurwitz Criteria

Given the characteristic equation (18), define k matrices as follows:

$$\begin{aligned} \mathbf{H}_1 &= (a_1), & \mathbf{H}_2 &= \begin{pmatrix} a_1 & 1 \\ a_3 & a_2 \end{pmatrix}, & \mathbf{H}_3 &= \begin{pmatrix} a_1 & 1 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{pmatrix}, \dots \\ \mathbf{H}_j &= \begin{pmatrix} a_1 & 1 & 0 & 0 & \cdots & 0 \\ a_3 & a_2 & a_1 & 1 & \cdots & 0 \\ a_5 & a_4 & a_3 & a_2 & \cdots & 0 \\ a_{2j-1} & a_{2j-2} & a_{2j-3} & a_{2j-4} & \cdots & a_j \end{pmatrix} \dots \mathbf{H}_k = \begin{pmatrix} a_1 & 1 & 0 & \cdots & 0 \\ a_3 & a_2 & a_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & a_k \end{pmatrix}, \end{aligned}$$

where the (l, m) term in the matrix \mathbf{H}_j is

$$\begin{aligned} a_{2l-m} & \quad \text{for } 0 < 2l - m < k, \\ 1 & \quad \text{for } 2l = m, \\ 0 & \quad \text{for } 2l < m \quad \text{or} \quad 2l > k + m. \end{aligned}$$

Then all eigenvalues have negative real parts; that is, the steady-state \bar{N} is stable if and only if the determinants of all Hurwitz matrices are positive:

$$\det \mathbf{H}_j > 0 \quad (j = 1, 2, \dots, k).$$

See Pielou (1969, chap. 6) for a treatment of the above and further references.

May (1973) summarizes these stability conditions in the cases $k = 2, \dots, 5$. Some of these are given in the small box.

Example 1 illustrates how the Routh-Hurwitz criteria might be applied in a situation in which three species interact.

Routh-Hurwitz Criteria for $k = 2, 3, 4$

$$k = 2: \quad a_1 > 0, \quad a_2 > 0.$$

$$k = 3: \quad a_1 > 0, \quad a_3 > 0; \quad a_1 a_2 > a_3.$$

$$k = 4: \quad a_1 > 0, \quad a_3 > 0; \quad a_4 > 0; \quad a_1 a_2 a_3 > a_3^2 + a_1^2 a_4.$$

Example 1

Suppose x is a predator and y and z are both its prey. z grows logistically in the absence of its predator. x dies out in the absence of prey, and y grows at an exponential rate in the absence of predator. We shall use the Routh-Hurwitz techniques to discover whether these species can coexist in a stable equilibrium.

Step 1

Writing equations for this system we get

$$\frac{dx}{dt} = \alpha(xz) + \beta(xy) - \gamma x, \quad (21a)$$

\swarrow growth from \nwarrow from eating y \nwarrow mortality
 eating z

$$\frac{dy}{dt} = \delta y - \epsilon(xy), \quad (21b)$$

\swarrow growth when no \nwarrow predation
 predator present mortality

$$\frac{dz}{dt} = \mu z(\nu - z) - \chi(xz). \quad (21c)$$

\swarrow logistic \nwarrow predation
 growth mortality

Step 2

Solving for steady state values we get

$$\alpha xz + \beta xy - \gamma x = 0 \Rightarrow \alpha \bar{z} + \beta \bar{y} = \gamma, \text{ or } \bar{x} = 0, \quad (22a)$$

$$\delta y - \epsilon(xy) = 0 \Rightarrow \bar{x} = \frac{\delta}{\epsilon}, \text{ or } \bar{y} = 0, \quad (22b)$$

$$\mu z(\nu - z) - \chi xz = 0 \Rightarrow \mu \nu - \mu \bar{z} - \chi \bar{x} = 0. \quad (22c)$$

From the above we arrive at the nontrivial steady state

$$\bar{x} = \frac{\delta}{\epsilon}, \quad \bar{y} = \gamma - \alpha \bar{z}, \quad \bar{z} = \nu - \frac{\chi}{\mu} \bar{x}.$$

This equilibrium makes sense biologically whenever $\gamma > \alpha \bar{z}$ and $\nu > \chi/\mu \bar{x}$.

Step 3

Calculating the Jacobian of the system, we get

$$J = \begin{pmatrix} \alpha\bar{z} + \beta\bar{y} - \gamma & \beta\bar{x} & \alpha\bar{x} \\ -\epsilon\bar{y} & \delta - \epsilon\bar{x} & 0 \\ -\chi\bar{z} & 0 & \mu\nu - 2\mu\bar{z} - \chi\bar{x} \end{pmatrix}. \quad (23)$$

Using the conclusions of step 2, we notice that terms on the diagonal of J evaluated at steady state lead to particularly simple forms, so that J is

$$J = \begin{pmatrix} 0 & \beta\bar{x} & \alpha\bar{x} \\ -\epsilon\bar{y} & 0 & 0 \\ -\chi\bar{z} & 0 & -\mu\bar{z} \end{pmatrix} \quad (24)$$

Step 4

To find eigenvalues we must set

$$\det(J - \lambda I) = 0.$$

Thus we must evaluate

$$\det \begin{pmatrix} 0 - \lambda & \beta\bar{x} & \alpha\bar{x} \\ -\epsilon\bar{y} & 0 - \lambda & 0 \\ -\chi\bar{z} & 0 & -\mu\bar{z} - \lambda \end{pmatrix}, \quad (25)$$

and set the result equal to zero to get the characteristic equation. Expanding, we get

$$\begin{aligned} & -\lambda \begin{vmatrix} -\lambda & 0 \\ 0 & -\mu\bar{z} - \lambda \end{vmatrix} - (-\epsilon\bar{y}) \det \begin{pmatrix} \beta\bar{x} & \alpha\bar{x} \\ 0 & -\mu\bar{z} - \lambda \end{pmatrix} + (-\chi\bar{z}) \det \begin{pmatrix} \beta\bar{x} & \alpha\bar{x} \\ -\lambda & 0 \end{pmatrix} \\ & = (-\lambda)(-\lambda)(-\mu\bar{z} - \lambda) - (-\epsilon\bar{y})(\beta\bar{x})(-\mu\bar{z} - \lambda) + (-\chi\bar{z})(-\alpha\bar{x})(-\lambda) \\ & = -\lambda^3 - \lambda^2\mu\bar{z} + \lambda(-\epsilon\bar{y}\beta\bar{x} - \chi\bar{z}\alpha\bar{x}) + (-\mu\bar{z}\epsilon\bar{y}\beta\bar{x}) = 0. \end{aligned}$$

Cancelling a factor of -1 we obtain

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0, \quad (26a)$$

where

$$a_1 = \mu\bar{z}, \quad (26b)$$

$$a_2 = \epsilon\beta\bar{x}\bar{y} + \chi\alpha\bar{x}\bar{z}, \quad (26c)$$

$$a_3 = \mu\epsilon\beta\bar{x}\bar{y}\bar{z}. \quad (26d)$$

Step 5

Now we check the three conditions using the Routh-Hurwitz criteria for the case $k = 3$ (three species): The three conditions are

1. $a_1 > 0$,
2. $a_3 > 0$,
3. $a_1a_2 > a_3$.

Condition 1 is true since $a_1 = \mu\bar{z}$ is a positive quantity. Condition 2 is true for the same reason. Looking at condition 3, we note that

$$a_1a_2 = \mu\bar{z}(\epsilon\beta\bar{x}\bar{y} + \chi\alpha\bar{x}\bar{z}).$$

This is clearly bigger than $a_3 = \mu\epsilon\beta\bar{x}\bar{y}\bar{z}$ since the quantity $\chi\alpha\mu\bar{x}\bar{z}^2$ is positive. Thus condition 3 is also satisfied.

We conclude that the steady state is a stable one.

Remark 1

Because calculations of 3×3 (and higher-order) determinants can be particularly cumbersome, it is advisable to express the Jacobian in the simplest possible notation. We do this by leaving entries in terms of \bar{x} , \bar{y} , and \bar{z} except where further simplification can be made (such as along the diagonal of \mathbf{J}). In step 5 we then use the fact that the quantities \bar{x} , \bar{y} , and \bar{z} are positive.

Remark 2

In some situations the *magnitudes* of the steady-state values also enter into the stability conditions. We will see in a later section why this is not the case here in example 1.

6.5 QUALITATIVE STABILITY

The Routh-Hurwitz criteria outlined in Section 6.4 are an exact but cumbersome method for determining stability of a large system. For communities of five or more species, the technique proves so computationally involved that it is of diminishing practical value. Shortcuts, when available, can be quite useful.

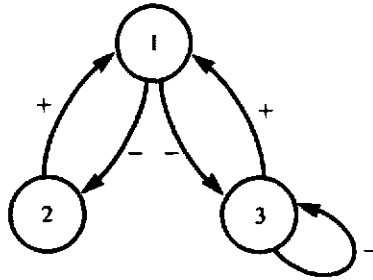
In this section we explore a shortcut method for investigating large systems that needs little if any computation. Because this method is not universally applicable, its importance is viewed as secondary. Furthermore, to understand exactly why the method works requires knowledge of matrices beyond elementary linear algebra. Nevertheless, what makes the technique of *qualitative stability* appealing is that it is easy to explain, easy to test, and thus a refreshing change from intensive computations.

The technique of qualitative stability analysis applies ideally to large complicated systems in which there is no quantitative information about the interrelationship of species or subsystems. Motivation for this method actually came from economics. A paper by the economists Quirk and Ruppert (1965) was followed later by further work and application to ecology by May (1973), Levins (1974), and Jeffries (1974).

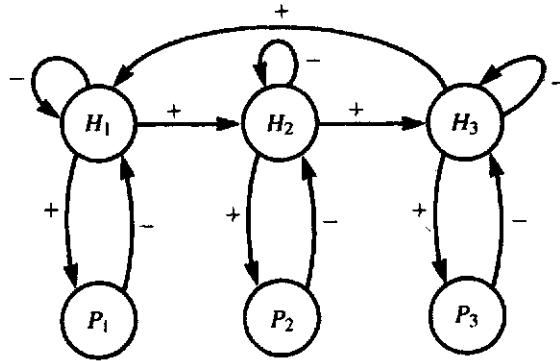
In a complex community composed of many species, numerous interactions take place. The magnitudes of the mutual effects of species on each other are seldom accurately known, but one can establish with greater certainty whether predation, competition, or other influences are present. This means that technically the functions appearing in equations that describe the system [such as equation (11)] are not known. What is known instead is the pattern of signs of partial derivatives of these functions [contained, for example, in the Jacobian of equation (16)]. We encountered a similar problem in the context of a plant-herbivore system (Chapter 3) and of a glucose-insulin model (Chapter 4). Here the problem consists of larger systems in a continuous setting, and even the magnitudes of partial derivatives may not be known.

There are two equivalent ways of representing qualitative information. A more obvious one is to assign the symbols $+$, 0 , and $-$ to the (i, j) th entry of a matrix if the species j has respectively a positive influence, no influence, or a negative

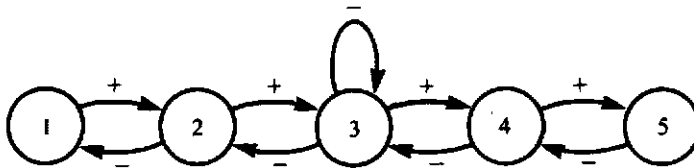
influence on species i . An alternate, visual representation captures the same ideas in a *directed graph* (also called *digraph*) in which nodes represent species and arrows between them represent the mutual interactions, as shown in Figures 6.10 and 6.11. The question is then whether it can be concluded, *from this graph or sign pattern only*, that the system is stable. If so, the system is called *qualitatively stable*.



(a)



(b)



(c)

Figure 6.10 Signed directed graphs (digraphs) can be used to represent species interactions in a complex ecosystem. The graphs shown here are

equivalent to the matrix representation of sign patterns given in the text (a) example 2, (b) example 3, and (c) example 4.

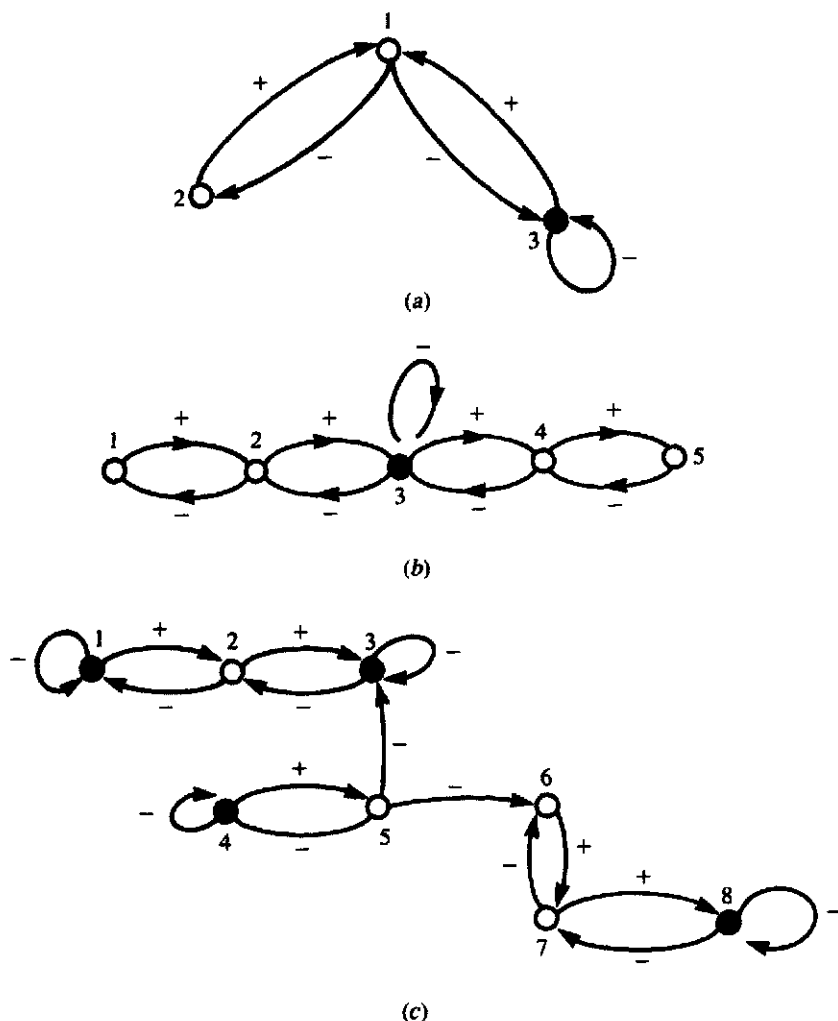


Figure 6.11 Properties of signed directed graphs can be used to deduce whether the system is qualitatively stable (stable regardless of the magnitudes of mutual effects). The Jeffries color test and the Quirk-Ruppert conditions are applied to

these graphs to conclude that (a), which corresponds to example 2, and (c) are stable communities, whereas (b), which corresponds to example 4, is not.

Systems that are qualitatively stable are also stable in the ordinary sense. (The converse is not true.) Systems that are *not* qualitatively stable can still be stable under certain conditions (for example, if the magnitudes of interactions are appropriately balanced.)

Following Quirk and Ruppert (1965), May (1973) outlined five conditions for

Example 2

Here we study the sign pattern of the community described in equations (21a, b, c) of Section 6.4. From Jacobian (24) of the system we obtain the qualitative matrix

$$Q = \text{sign } J = \begin{bmatrix} 0 & + & + \\ - & 0 & 0 \\ - & 0 & - \end{bmatrix}.$$

This means that close to equilibrium, the community can also be represented by the graph in Figure 6.10. Reading entries in Q from left to right, top to bottom:

Species 1 gets positive feedback from species 2 and 3.

Species 2 gets negative feedback from species 1.

Species 3 gets negative feedback from species 1 and from itself.

Example 3 (Levins, 1977)

In a closed community, three predators or parasitoids, labeled P_1 , P_2 , and P_3 , attack three different stages in the life cycle of a host, H_1 , H_2 , and H_3 . The presence of hosts is a positive influence for their predators but predators have a negative influence on their prey. Figure 6.10(b) and the following matrix summarize the interactions:

$$\begin{array}{c} P_1 \\ P_2 \\ P_3 \\ H_1 \\ H_2 \\ H_3 \end{array} \begin{bmatrix} P_1 & P_2 & P_3 & H_1 & H_2 & H_3 \\ 0 & 0 & 0 & + & 0 & 0 \\ 0 & 0 & 0 & 0 & + & 0 \\ 0 & 0 & 0 & 0 & 0 & + \\ - & 0 & 0 & - & 0 & + \\ 0 & - & 0 & + & - & 0 \\ 0 & 0 & - & 0 & + & - \end{bmatrix}.$$

Note that H_1 , H_2 , and H_3 each exert negative feedback on themselves.

Example 4 (Jeffries, 1974)

In a five-species ecosystem, species 2 preys on species 1, species 3 on species 2, and so on in a food chain up to species 5. Species 3 is also self-regulating. A qualitative matrix for this community is

$$Q = \begin{bmatrix} 0 & - & 0 & 0 & 0 \\ + & 0 & - & 0 & 0 \\ 0 & + & - & - & 0 \\ 0 & 0 & + & 0 & - \\ 0 & 0 & 0 & + & 0 \end{bmatrix}.$$

See Figure 6.10(c).

qualitative stability. Suppose a_{ij} is the ij th element of the matrix of signs Q . Then it is necessary for all of the following conditions to hold:

1. $a_{ii} \leq 0$ for all i .
2. $a_{ii} < 0$ for at least one i .
3. $a_{ij}a_{ji} \leq 0$ for all $i \neq j$.
4. $a_{ij}a_{jk} \cdots a_{qr}a_{ri} = 0$ for any sequences of three or more distinct indices i, j, k, \dots, q, r .
5. $\det Q \neq 0$.

These conditions can be interpreted in the following way:

1. No species exerts positive feedback on itself.
2. At least one species is self-regulating.
3. The members of any given pair of interacting species must have opposite effects on each other.
4. There are no closed chains of interactions among three or more species.
5. There is no species that is unaffected by interactions with itself or with other species.

For mathematical proof of these five necessary conditions, consult Quirk and Rupert (1965). May (1973) and Pielou (1969) comment on the biological significance, particularly of conditions 3 and 4. The conditions can be tested by looking at graphs representing the communities. One must check that these graphs have all the following properties:

1. No + loops on any single species (that is, no positive feedback).
2. At least one - loop on some species in the graph.
3. No pair of like arrows connecting a pair of species.
4. No cycles connecting three or more species.
5. No node devoid of input arrows.

These five conditions are equivalent to the original algebraic statement.

Example 5

For examples 2 to 4 we check off the five conditions given earlier:

Condition Number	Example 2	Example 3	Example 4
1	✓	✓	✓
2	✓	✓	✓
3	✓	✓	✓
4	✓	No	✓
5	✓	✓	✓

It was shown by Jeffries (1974) that these five conditions alone cannot distinguish between *neutral stability* (as in the Lotka-Volterra cycles) and *asymptotic stability*, wherein the steady state is a stable node or spiral. (In other words, the conditions are *necessary but not sufficient* to guarantee that the species will coexist in a constant steady state). In example 4 Jeffries notes that pure imaginary eigenvalues can occur, so that even though the five conditions are met, the system will oscillate. To weed out such marginal cases, Jeffries devised an auxiliary set of conditions, which he called the "color test," that *replaces* condition 2. Before describing the color test, it is necessary to define the following:

A *predation link* is a pair of species connected by one + line and one - line.

A *predation community* is a subgraph consisting of all interconnected predation links.

If one defines a species not connected to any other by a predation link as a *trivial* predation community, then it is possible to decompose any graph into a set of distinct predation communities. The systems shown in Figure 6.10 have predation communities as follows: (a) {2, 1, 3}; (b) $\{H_1, P_1\}$, $\{H_2, P_2\}$, $\{H_3, P_3\}$; and (c) {1, 2, 3, 4, 5}. In Figure 6.11(c) there are three predation communities: {1, 2, 3}, {4, 5}, and {6, 7, 8}.

The following color scheme constitutes the test to be made. A predation community is said to *fail the color test* if it is not possible to color each node in the subgraph black or white in such a way that

1. Each self-regulating node is black.
2. There is at least one white point.
3. Each white point is connected by a predation link to at least one other white point.
4. Each black point connected by a predation link to one white node is also connected by a predation link to one other white node.

Jeffries (1974) proved that for asymptotic stability, a community must satisfy the original Quirk-Ruppert conditions 1, 3, 4, and 5, and in addition must have only predation communities that fail the color test.

Example 5 (continued)

Examples 2 and 4 satisfy the original conditions. In Figure 6.11 the color test is applied to their communities. We see that example 2 consists of a single predation community that *fails* part 4 of the color test. Example 4 *satisfies* the test. A final example shown in Figure 6.11(c) has three predation communities, and each one fails the test. We conclude that Figure 6.11(a) and (c) represent systems that have the property of asymptotic stability; that is, these ecosystems consist of species that coexist at a stable fixed steady state without sustained oscillations.

A proof and discussion of the revised conditions is to be found in Jeffries (1974). For other applications and properties of graphs, you are encouraged to peruse Roberts (1976).

6.6 THE POPULATION BIOLOGY OF INFECTIOUS DISEASES

Infectious diseases can be classified into two broad categories: those caused by viruses and bacteria are *microparasitic* diseases, and those due to worms (more commonly found in third-world countries) are *macroparasitic*. Other than the relative sizes of the infecting agents, the main distinction is that microparasites reproduce within their host and are transmitted directly from one host to another. Most macroparasites, on the other hand, have somewhat more complicated life cycles, often with a secondary host or carrier implicated. (Examples of these include malaria and schistosomiasis; see Anderson, 1982 for a review.)

This section briefly summarizes some of the classical models for microparasitic infections. The mathematical techniques required for analyzing the models parallel the techniques applied in Sections 6.2 and 6.3. However, as a general remark, it should be said that the *flavor* of the models differs somewhat from the species-interactions models introduced in this chapter.

With no *a priori* knowledge, suppose we are asked to model the process of infection of a viral disease such as measles or smallpox. In keeping with the style of population models for predation or competition, it would be tempting to start by defining variables for population densities of the host x and infecting agent y . Here is how such a model might proceed:

Primitive Model for a Viral Infection

This model is for illustrative purposes only. Let

x = population of human hosts,

y = viral population.

The assumptions are that

1. There is a constant human birth rate α .
2. Viral infection causes an increased mortality due to disease, so $g(y) > 0$.
3. Reproduction of viral particles depends on human presence.
4. In the absence of human hosts, virus particles "die" or become nonviable at rate γ .

The equations then read:

$$\frac{dx}{dt} = [\alpha - g(y)]x,$$

$$\frac{dy}{dt} = \beta xy - \gamma y.$$

The approach leads us to a modified Lotka-Volterra predation model. This view, to put it simply, is that viruses y are predatory organisms searching for human prey x to consume. The conclusions given in Section 6.2 follow with minor modification.

The philosophical view of disease as a process of predation is an unfortunate and somewhat misleading analogy on several counts. First, no one can reasonably suppose it possible to measure or even estimate total viral population, which may range over several orders of magnitude in individual hosts. Second, a knowledge of this number is at best uninteresting and trivial since it is the distribution of viruses over hosts that determines what percentage of people will actually suffer from the disease. To put it another way, some hosts will harbor the infecting agent while others will not. Finally, in the "primitive" model an underlying hidden assumption is that viruses roam freely in the environment, randomly encountering new hosts. This is rarely true of microparasitic diseases. Rather, diseases are spread by contact or close proximity between infected and healthy individuals. How the disease is spread in the population is an interesting question. This crucial point is omitted and is thus a serious criticism of the model.

A new approach is necessary. At the very least it seems sensible to make a distinction between sick individuals who harbor the disease and those who are as yet healthy. This forms the basis of all microparasitic epidemiological models, which, as we see presently, virtually omit the population of parasites from direct consideration.

Instead, the host population is subdivided into distinct classes according to the health of its members. A typical subdivision consists of susceptibles S , infectives I , and a third, removed class R of individuals who can no longer contract the disease because they have recovered with immunity, have been placed in isolation, or have died. If the disease confers a *temporary immunity* on its victims, individuals can also move from the third class to the first.

Time scales of epidemics can vary greatly from weeks to years. *Vital dynamics* of a population (the normal rates of birth and mortalities in the absence of disease) can have a large influence on the course of an outbreak. Whether or not immunity is conferred on individuals can also have an important impact. Many models using the general approach with variations on the assumptions have been studied. An excellent summary of several is given by Hethcote (1976) and Anderson and May (1979), although different terminology is unfortunately used in each source.

Some of the earliest classic work on the theory of epidemics is due to Kermack and McKendrick (1927). One of the special cases they studied is shown in Figure 6.12(a). The diagram summarizes transition rates between the three classes with the parameter β , the rate of transmission of the disease, and the rate of removal ν . It is assumed that each compartment consists of identically healthy or sick individuals and that no births or deaths occur in the population. (In more current terminology, the situation shown in Figure 6.12(a) would be called an *SIR model without vital dynamics* because the transitions are from class S to I and then to R ; see for example, Hethcote, 1976.)

Figure 6.12 A number of epidemic models that have been studied. The total population N is subdivided into susceptible (S), infective (I) and removed (R) classes. Transitions between compartments depict the course of transmission, recovery, and loss of immunity with rate constants β , ν , and γ . A population with vital dynamics is assumed to be producing new susceptibles at rate δ which is identical to the mortality rate. (a) SIR model; (b, c) SIRS models; and (d) SIS model.

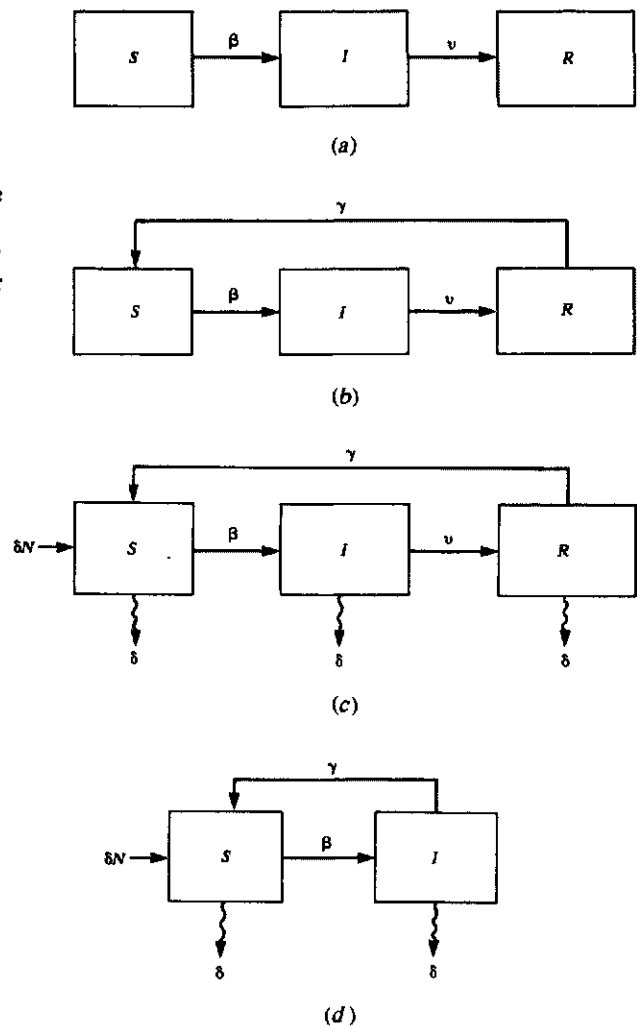


Figure 6.12(a) and those following it are somewhat reminiscent of models we have already studied for the physical flows between well-mixed compartments (for example, the chemostat). A subtle distinction must be made though, since the passage of individuals from the susceptible to the infective class generally occurs as a result of close proximity or contact between healthy and infective individuals. Thus the rate of exchange between S and I has a special character summarized by the following assumption:

Assumption

The rate of transmission of a microparasitic disease is proportional to the rate of encounter of susceptible and infective individuals modelled by the product (βSI) .

The equations due to Kermack and MacKendrick for the disease shown in Figure 6.12(a) are thus

$$\frac{dS}{dt} = -\beta SI, \quad (27a)$$

$$\frac{dI}{dt} = \beta SI - \nu I, \quad (27b)$$

$$\frac{dR}{dt} = \nu I. \quad (27c)$$

It is easily verified that the total population $N = S + I + R$ does not change. Though these equations are nonlinear, Kermack and MacKendrick derived an approximate expression for the rate of removal dR/dt (in their paper called dz/dt) as a function of time. The result is a rather messy expression involving hyperbolic secants; when plotted with the appropriate values given to the parameters it compares rather well with data for death by plague in Bombay during an epidemic in 1906 (see Figure 6.13).

A more instructive approach is to treat the problem by qualitative methods. Now we shall carry out this procedure on a slightly more general case, allowing for a loss of immunity that causes recovered individuals to become susceptible again [Figure 6.12(b)]. It will be assumed that this takes place at a rate proportional to the population in class R , with proportionality constant γ . Thus the equations become

$$\frac{dS}{dt} = -\beta SI + \gamma R, \quad (28a)$$

$$\frac{dI}{dt} = \beta SI - \nu I, \quad (28b)$$

$$\frac{dR}{dt} = \nu I - \gamma R. \quad (28c)$$

This model is called an *SIRS* model since removed individuals can return to class S ($\gamma = 0$ is the special case studied by Kermack and McKendrick). It is readily shown that these equations have two steady states:

$$\bar{S}_1 = N, \quad \bar{I}_1 = 0, \quad \bar{R}_1 = 0; \quad (29a)$$

$$\bar{S}_2 = \frac{\nu}{\beta}, \quad \bar{I}_2 = \gamma \frac{N - \bar{S}_2}{\nu + \gamma}, \quad \bar{R}_2 = \frac{\nu \bar{I}_2}{\gamma}. \quad (29b)$$

In (29a) the whole population is healthy (but susceptible) and disease is eradicated. In (29b) the community consists of some constant proportions of each type provided $(\bar{S}_2, \bar{I}_2, \bar{R}_2)$ are all positive quantities. For \bar{I}_2 to be positive, N must be larger than \bar{S}_2 . Since $\bar{S}_2 = \nu/\beta$, this leads to the following conclusion:

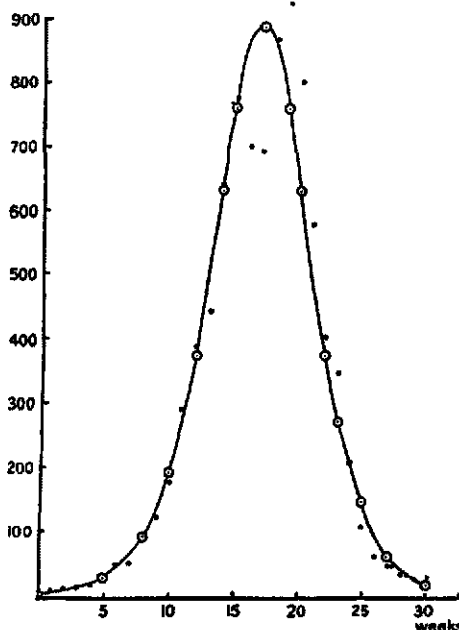
The disease will be established in the population provided the total population N exceeds the level ν/β , that is,

$$\frac{N\beta}{\nu} > 1.$$

This important *threshold effect* was discovered by Kermack and McKendrick; the population must be "large enough" for a disease to become endemic.

Also for the rate at which cases are removed by death or recovery which is the form in which many statistics are given

$$\frac{dz}{dt} = \frac{I^2}{2\alpha\sigma c^2} \sqrt{-g} \operatorname{sech}^2\left(\frac{\sqrt{-g}}{2} t - \phi\right). \quad (\text{thirty-one})$$



The accompanying chart is based upon figures of deaths from plague in the island of Bombay over the period December 17, 1905, to July 31, 1906. The ordinate represents the number of deaths per week, and the abscissa denotes the time in weeks. As at least 80 to 90 per cent. of the cases reported terminate fatally, the ordinate may be taken as approximately representing dz/dt as a function of t . The calculated curve is drawn from the formula

$$\frac{dz}{dt} = 890 \operatorname{sech}^2(0.26 t - 3.4).$$

Figure 6.13 On a page from their original article, Kermack and McKendrick compare predictions of the model given by equations (27a,b,c) with data for the rate of removal by death. Note: dz/dt is

equivalent to dR/dt in equations (27). [Kermack, W. O., and McKendrick, A. G. (1927). A contribution to mathematical theory of epidemics. Roy Stat. Soc. J., 115, 714.]

The ratio of parameters β/ν has a rather meaningful interpretation. Since removal rate from the infective class is ν (in units of 1/time), the average period of infectivity is $1/\nu$. Thus β/ν is the fraction of the population that comes into contact with an infective individual during the period of infectiousness. The quantity $R_0 = N\beta/\nu$ has been called the *infectious contact number*, σ (Hethcote, 1976) and the *intrinsic reproductive rate of the disease* (May, 1983). R_0 represents the average number of secondary infections caused by introducing a single infected individual into a host population of N susceptibles. (In papers by May and Anderson, the threshold result is usually written $R_0 > 1$.)

In further analyzing the model we can take into account the particularly convenient fact that the total population

$$N = S + I + R$$

does not change (see problem 25 for verification). This means that one variable, say R , can always be eliminated so that the model can be given in terms of two equations in two unknowns. In the following analysis this fact is exploited in applying phase-plane methods to the problem.

Qualitative Analysis of a SIRS Model: Epidemic with Temporary Immunity and No Vital Dynamics

Since the total population is constant, we eliminate R from equations (28) by substituting

$$R = N - S - I. \quad (30)$$

The equations for S and I are then

$$\frac{dS}{dt} = -\beta SI + \gamma(N - S - I) \stackrel{\text{def}}{=} F(S, I), \quad (31a)$$

$$\frac{dI}{dt} = \beta SI - \nu I \stackrel{\text{def}}{=} G(S, I). \quad (31b)$$

Nullclines

$$\begin{aligned} I' = 0 & \quad \text{for} \quad I = 0 \quad \text{and} \quad S = \nu/\beta, \\ S' = 0 & \quad \text{for} \quad \beta SI = \gamma(N - S - I). \end{aligned}$$

After rearranging,

$$I = \frac{\gamma(N - S)}{(\beta S + \gamma)}.$$

This curve intersects the axes at $(N, 0)$ and $(0, N)$.

Steady states

$$(\bar{S}_1, \bar{I}_1) = (N, 0); (\bar{S}_2, \bar{I}_2) = \left(\frac{\nu}{\beta}, \frac{N - (\nu/\beta)}{\nu + \gamma} \right).$$

Jacobian

$$\mathbf{J} = \begin{pmatrix} F_S & F_I \\ G_S & G_I \end{pmatrix}_s = \begin{pmatrix} -(\beta \bar{I} + \gamma) & -(\beta \bar{S} + \gamma) \\ \beta \bar{I} & \beta \bar{S} - \nu \end{pmatrix}.$$

Stability

For (\bar{S}_2, \bar{I}_2) ,

$\text{Tr } \mathbf{J} = -(\beta \bar{I}_2 + \gamma)$ is always negative,

$\det \mathbf{J} = \beta \bar{I}_2(\nu + \gamma)$ is always positive.

Thus this steady state is always stable when it exists, namely when the threshold condition is satisfied. It is evident from Figures 6.14 and 6.15(b) and from further analysis that the approach to this steady state can be oscillatory.

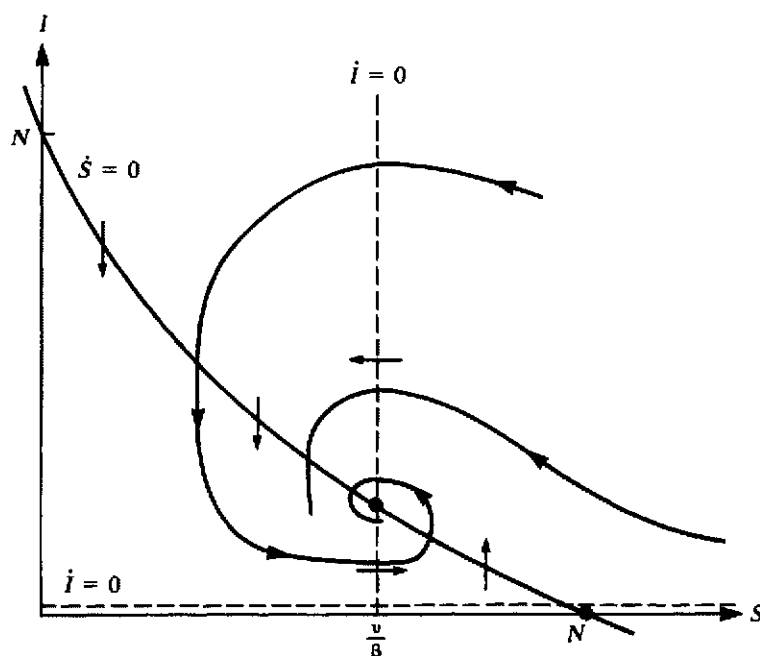


Figure 6.14 Nullclines, steady states, and several trajectories for the SIRS model given by equations

(31a,b), which are equivalent to (28a,b,c).

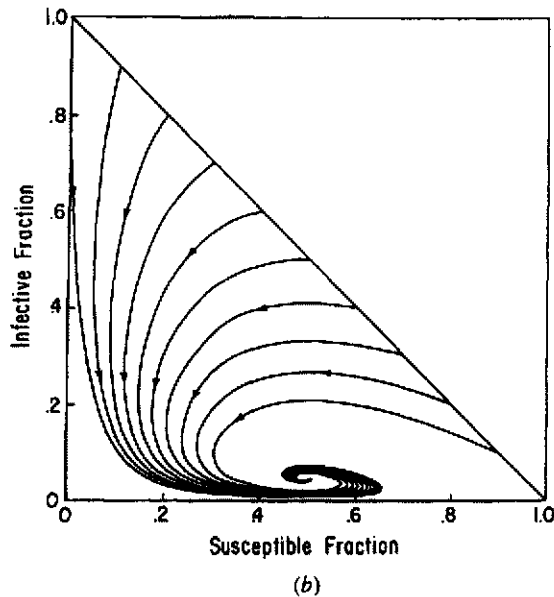
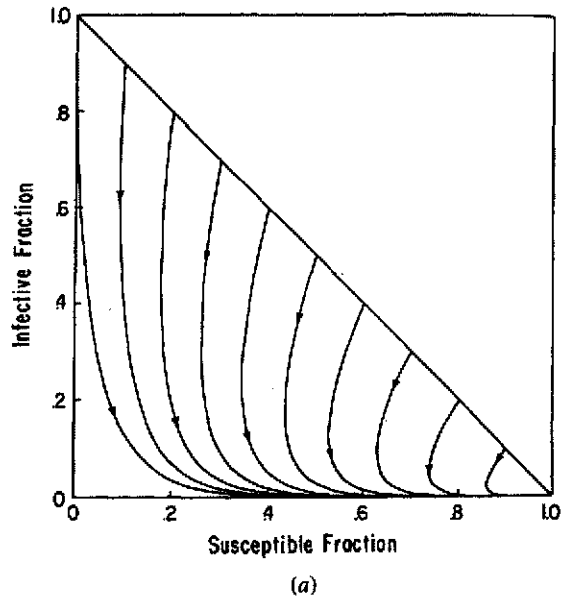
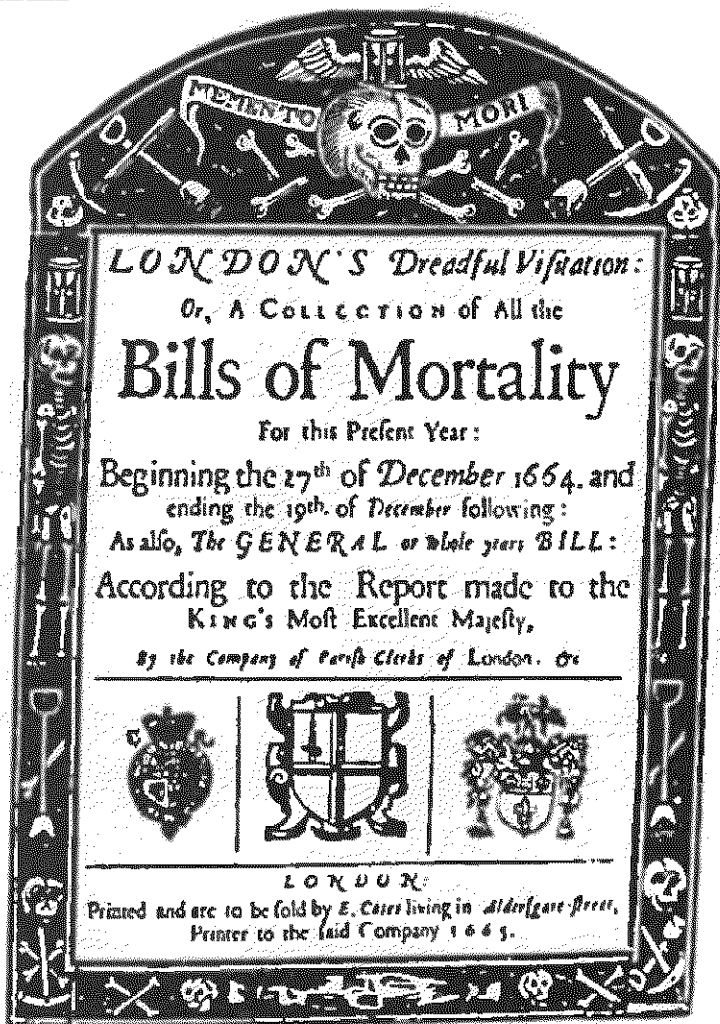


Figure 6.15 Epidemic models are characterized by the magnitude of an infectious contact number σ .
 (a) When $\sigma < 1$, the infective class will disappear.
 (b) When $\sigma > 1$, there is some stable steady state in which both susceptibles and infectives are present. Shown here is an SIRS model with vital

dynamics.) [Reprinted by permission of the publisher from Hethcote, H. W. (1976). Qualitative analyses of communicable disease models. *Math. Biosci.*, 28, 344 and 345. Copyright 1976 by Elsevier Science Publishing Co., Inc.]



Mortality from a variety of afflictions, only some of which were caused by disease, were systematically recorded as early as the 1600s in the Bills of Mortality published in London. Reproduced here is the title page of the London Bills of Mortality for 1665, the year of the great plague. The people of the city followed with anxiety the rise and fall in the number of deaths from the plague, hoping always to see the sharp decline which they knew from past experience indicated that the epidemic was nearing its end. When the decline came the refugees, mostly from the nobility and wealthy merchants, returned to the city, and then for a time

*the mortality rose again as the disease attacked these new arrivals. The plague of 1665 started in June; its peak came in September and its decline in October. The secondary rise occurred in November and cases of the disease were reported as late as March of the following year. [From H. W. Haggard (1957), *Devils, Drugs, Doctors*, Harper & Row, New York.]*

The World of Mathematics, Vol. 3. Copyright ©1956 by James R. Newman; renewed ©1984 by Ruth G. Newman. Reprinted by permission of Simon & Schuster, Inc.

The Diseases, and Casualties this year being 1632.

A Bortive, and Stilborn ..	445	Grief	11
Affrighted	1	Jaundies	43
Aged	628	Jawfain	8
Ague	43	Impostume	74
Apoplex, and Meagrom	17	Kil'd by several accidents..	46
Bit with a mad dog.....	1	King's Evil.....	38
Bleeding	3	Lethargie	2
Bloody flux, scowring, and		Livergrown	87
flux	348	Lunatique	6
Brused, Issues, sores, and		Made away themselves....	15
ulcers,	28	Measles	80
Burnt, and Scalded.....	5	Murthered	7
Burst, and Rupture.....	9	Over-laid, and starved at	
Cancer, and Wolf.....	10	nurse	7
Canker	1	Palsie	25
Childbed	171	Piles.....	1
Chrisomes, and Infants....	2288	Plague.....	8
Cold, and Cough.....	55	Planet	13
Colick, Stone, and Strangury	56	Pleurisie, and Spleen.....	36
Consumption	1797	Purples, and spotted Fever	38
Convulsion	241	Quinsie	7
Cut of the Stone.....	5	Rising of the Lights.....	98
Dead in the street, and		Sciatica	1
starved	6	Scurvey, and Itch.....	9
Dropsie, and Swelling.....	267	Suddenly	62
Drowned	34	Surfet	36
Executed, and prest to death	18	Swine Pox	6
Falling Sicknease.....	7	Teeth	470
Fever	1108	Thrush, and Sore mouth...	40
Fistula	13	Tympany	13
Flocks, and small Pox....	531	Tissick	34
French Pox.....	12	Vomiting	1
Gangrene	5	Worms	27
Gout	4		

Christened { Males....4994 } Buried { Males....4932 } Whereof,
 { Females..4590 } { Females..4603 } of the
 { In all....9584 } { In all....9535 } Plague.8

Increased in the Burials in the 122 Parishes, and at the Pest-house this year..... 993

Decreased of the Plague in the 122 Parishes, and at the Pest-house this year..... 266 [10]

Numerous other cases have been analyzed in detail. Perhaps the best summary is given by Hethcote (1976), in which theoretical results are followed by *biocorollaries* that spell out the biological predictions. His paper was used in drawing up Table 6.1, a composite that describes a number of cases.

One point worth mentioning is the essential difference between models in which the susceptible class is *renewed* (by recovery or loss of immunity) and those in which it is not. [The *SIRS* and *SIS* examples shown in Figure 6.12(b–d) belong to the former category.] These distinct types behave differently when the normal turnover of births and deaths is superimposed on the dynamics of the disease. In the *SIR* type without births, the continual decrease of the susceptible class results in a decline in the effective reproductive rate of the disease. The epidemic stops for want of infectives, not, as it might seem, for want of susceptibles (Hethcote, 1976). On the other hand, if the susceptible class is replenished by births or recoveries, the subpopulation that participates in the disease is maintained, and the disease can persist.

From Table 6.1 we see that *SIR* models are subdivided into those with and without births and deaths. In other models the chief effect of normal birth and mortality at rates δ is to decrease the infectious contact number σ . This means that a smaller population can sustain an endemic disease. Note that the total population N is taken to be constant in all of these models since the number of deaths from all classes is assumed to exactly balance the births of new susceptibles. Among other things, this permits all such models to be analyzed by methods similar to the method used here since one variable can always be eliminated.

A somewhat different philosophical approach was taken by Anderson and May (1979), who were less interested in the dynamics of the disease itself. By analyzing a model in which a disease-free population grows exponentially, rather than being maintained at a constant level, they demonstrated that epidemics increasing host mortality have the potential to regulate population levels (see problem 30). This adds yet another interesting possibility to the list of causes of decelerating growth rates in natural populations. Aside from inter- and intraspecies competition and predation, disease-causing agents (much like parasites) can control the population dynamics of their hosts.

The theory of epidemics has numerous ramifications, some of which are mathematical and some practical. In recent years more advanced mathematical models have been studied to determine the effects of delay factors (such as a waiting time in the infectious class), age structure, migration, and spatial distributions. Many of these models require sophisticated mathematical methods of analysis (see, for example, Busenberg and Cooke, 1978). An excellent survey with detailed references is given by Hethcote et al. (1981).

One theoretical question such papers often address is whether models with particular structures lead to stable (limit-cycle) oscillations. This question is of interest since some diseases are associated with periodic outbreaks with very low endemic periods followed by peak epidemic cycles. In some cases the forces driving such cyclic behavior are related to seasonality and to changes in contact rates. (A good example is childhood diseases, which invariably peak during the school year when contact between their potential hosts is greatest.) However, even in the absence of externally imposed periodicity, models similar to *SIRS* can have an inherent ten-

Table 6.1 *A Summary of Several Epidemic Models*

Type	Immunity	Birth/Death	Significant quantity	Results	Figures
SIS	None	Rate = δ	$\sigma = \frac{\beta}{\gamma + \delta}$	(1) $\sigma > 1$: constant endemic infection (2) $\sigma < 1$: infection disappears	6.9(d)
		Additional disease fatality rate η	σ as above, and $\epsilon = \frac{\eta}{\gamma + \delta}$	Disease always eventually disappears leaving some susceptibles.	
SIR	Yes, recovery gives immunity.	None	$\sigma = \frac{S_0 \beta}{\gamma}$ (S_0 = initial S)	(1) $\sigma > 1$: infection peaks and then disappears (2) $\sigma < 1$: infection disappears	6.9(a)
		Yes, rate = δ	$\sigma = \frac{\beta}{\nu + \delta}$	(1) $\sigma < 1$: susceptibles and infectives approach constant levels (2) $\sigma < 1$: infectives disappear; only S remains	6.9(c) with $\gamma = 0$
(SIR with carriers)	Yes	Yes		Disease always remains endemic.	
SIRS	Temporary, lost at rate γ	Rate = δ	$\sigma = \frac{\beta}{\nu + \delta}$	(1) $\sigma > 1$: same as SIR (1) but higher levels of infectives (2) $\sigma < 1$: same as SIR (2)	6.9(c)

dency to give rise to oscillations. This is particularly true of models with long periods of immunity or some other delaying factor. Hethcote notes that a sequence of at least three removed classes will also achieve the result (for example, $SIR_1 R_2 R_3 S$).

The implications of many aspects of applying mathematical theory to natural populations are eloquently described in numerous papers by May and Anderson. Some questions are of a basic scientific nature. For example, the extent to which diseases and hosts have coevolved is a fascinating topic; a second controversial question is whether or not diseases are in fact a predominant factor in controlling natural populations. Other questions have more immediate medical ramifications. Anderson and May suggest that theory has an important place in illuminating the impact of disease on human populations and the ability to eradicate or control disease. Two applications of the theory to vaccination programs are briefly highlighted in the following section.

6.7 FOR FURTHER STUDY: VACCINATION POLICIES

Models for infectious diseases lead to a better understanding of how vaccination programs affect the control or eradication of the disease. Several popular articles by Anderson and May (1982) and a more detailed mathematical version (1983) are thought-provoking and informative. The full theory that takes into account age structure of the population uses a partial differential equation model (which would be understood more fully after covering Chapter 11). However, a number of rather interesting consequences of the theory can be understood with no further preparation.

Eradicating a Disease

Immunization can reduce or eliminate the incidence of infection, even when only part of the population receives the treatment. Those individuals who have been vaccinated will be protected from acquiring infection (this is the obvious direct effect). A secondary effect is that since vaccinated individuals are essentially removed from participating in transmission of the disease, there will be fewer infectious individuals and thus a decreased likelihood that an unvaccinated susceptible will come in contact with the disease. This indirect effect is known as *herd immunity*.

Administering vaccinations to an entire population can be costly. Some vaccines (for example, some measles and whooping cough vaccines) also carry the risk, though rare, of causing various reactions or neurological damage. Thus, if disease eradication can be achieved by partially vaccinating some fraction p of the population, an advantage is gained.

The fraction to be immunized must be such that the remaining population, $(1 - p)N$, will no longer exceed the threshold level necessary to perpetuate the disease. In the terminology of Anderson and May, the reproductive factor R_0 of the infection is to be reduced below 1. Since $R_0 = N\beta/\nu$, for a given disease this factor can be estimated from epidemiological and population records. Table 6.2 lists several common diseases with their corresponding R_0 factors.

Table 6.2 *Estimates of the intrinsic reproductive rate R_0 for human diseases and the corresponding percentage of the population p that must be protected by immunization to achieve eradication. [Reprinted by permission, American Scientist, journal of Sigma Xi, "Parasitic Infections as Regulator of Animal Populations," by Robert M. May, 71:36-45 (1983).]*

<i>Infection</i>	<i>Location and Time</i>	R_0	<i>Approximate Value of p (%)</i>
Smallpox	Developing countries, before global campaign	3-5	70-80
Measles	England and Wales, 1956-68;	13	92
	U.S., various places, 1910-30	12-13	92
Whooping cough	England and Wales, 1942-50;	17	94
	Maryland, U.S., 1908-17	13	92
German measles	England and Wales, 1979;	6	83
	West Germany, 1972	7	86
Chicken pox	U.S., various places, 1913-21 and 1943	9-10	90
Diphtheria	U.S., various places, 1910-47	4-6	~80
Scarlet fever	U.S., various places, 1910-20	5-7	~80
Mumps	U.S., various places, 1912-16 and 1943	4-7	~80
Poliomyelitis	Holland, 1960; U.S., 1955	6	83

The fraction p to be immunized is then deduced from the following simple calculation:

Intrinsic reproductive rate of disease = R_0 , Fraction immunized = p , Fraction not immunized = $1 - p$, Population participating in disease = $N(1 - p)$, Effective intrinsic reproduction rate of disease (after immunization) = $R'_0 = (1 - p)R_0$.

Thus

$$R'_0 < 1 \Rightarrow (1 - p)R_0 < 1 \Rightarrow p > 1 - \frac{1}{R_0}$$

The percentage of the population to be vaccinated thus depends strongly on the infectiousness of the disease. It is noteworthy that smallpox, a disease essentially eradicated by vaccination, has one of the lowest R_0 values and a correspondingly low required vaccination fraction. By contrast, measles and whooping cough require a much higher percentage of immunization and would be harder to eradicate.

Average Age of Acquiring a Disease

Is it always wise to vaccinate at least some people, even if the disease will not be eradicated? When a disease has different impacts on individuals of different ages, vaccination at a young age can have a somewhat surprising deleterious effect. A case in point is German measles (Rubella). Normally a mild short-lasting infection, Rubella can be particularly devastating when contracted by a pregnant woman, as it results in birth defects to the fetus during the first trimester of pregnancy. Vaccinating *against* Rubella raises the average age at which the disease is first acquired (see box). Thus, rather than incurring the disease on average at age 12, it may be more prevalent at ages 20 to 30, precisely the most dangerous period for women of child-bearing ages.

How Vaccinations Raise the Age of First Acquiring a Disease

Define $\lambda = \beta I$. Called the *per capita force of infection*, λ has units of 1/time and is the rate of acquiring the disease given a population containing I infectives and a transmission constant β .

Let $A = 1/\lambda$. A is the *average age of first infection*, or average waiting time in the susceptible compartment before acquiring the disease. A has units of time.

Now note that a vaccination program tends to reduce the number of infectives I , thus reducing λ and *raising* A .

For other aspects of the topic of vaccinations, epidemiology, and population dynamics, the many excellent sources quoted in the references are highly recommended.

PROBLEMS*

1. (a) Assuming that $N(0) = N_0$, integrate equation (2a) and show that its solution is given by equation (2b). (See problem 5 of Chapter 4 or Braun, 1979; sec. 1.5.)
- (b) Show that the solution given by (2b) has the following properties:
 - (1) $N \rightarrow K$ as $t \rightarrow \infty$.
 - (2) The graph is concave up for $N_0 < N < K/2$.
 - (3) The graph is concave down for $K/2 < N < K$.
 - (4) If $N_0 > K$, the graph is concave up.

2. Consider the model

$$\frac{dN}{dt} = rN(K - N)(N - M) \quad \text{where } r > 0 \text{ and } 0 < M < K$$

- (a) Express the intrinsic growth rate $g(N)$ as a polynomial in N and find the coefficients a_1 , a_2 , a_3 .
 - (b) Show that $N = 0$, $N = K$, and $N = M$ are steady states and determine their stability.
 - (c) Solve the equation and graph the solution.
3. Models that are commonly used in fisheries are

$$\frac{dN}{dt} = Ng(N),$$

where $g(N)$ is given by

$$\text{Ricker model: } g(N) = re^{-\beta N}.$$

$$\text{Beverton-Holt: } g(N) = \frac{r}{\alpha + N}.$$

Analyze the behavior of the solutions to these questions. (Assume α , β , $r > 0$).

4. For single-species populations, which of the following density-dependent growth rates would lead to a decelerating rate of growth as the population increases? Which would result in a stable population size?
 - (a) $g(N) = \frac{\beta}{1 + N}$, $\beta > 0$.
 - (b) $g(N) = \beta - N$, $\beta > 0$.
 - (c) $g(N) = N - e^{\alpha N}$, $\alpha > 0$.
 - (d) $g(N) = \log N$.
5. List the assumptions that underlie the logistic equation (2a). You may wish to think about such factors as environmental or individual variability, reproductive ages, and the effects of the spatial distribution of the population. Which assumptions are not generally valid?

*Problems preceded by an asterisk are especially challenging.

1. Several problems were kindly suggested by C. Biles.

6. The factor $g(N) = r(1 - N/K)$ in equation (2a) is a per capita growth rate. Smith (1963) observed that in cultures of the unicellular alga *Daphnia magna* g decreases at a nonlinear rate as N increases. To account for this fact, Smith suggested that the growth rate depends on the rate at which food is utilized:

$$g(N) = r \frac{T - F}{T}$$

where F is the rate of utilization when the population size is N , and T is the maximal rate, when the population has reached a saturated level. He further assumed that

$$F = c_1 N + c_2 \frac{dN}{dt}, \quad (c_1, c_2 > 0)$$

as long as $dN/dt > 0$.

- (a) Explain this assumption for F .
 (b) Show that the modified logistic equation is then

$$\frac{dN}{dt} = rN \left[\frac{K - N}{K + (\gamma N)} \right],$$

where $\gamma = rc_2/c_1$ and $K = T/c_1$.

- (c) Sketch the expression in square brackets as a function of N .
 (d) What would be the qualitative behavior of this population growth? (For a deeper analysis of this problem see Pielou, 1977.)
7. (a) Show that equations (6a,b,c) for the Gompertz growth law are equivalent. Find κ in terms of α .
 (b) In a tumor the cells without access to nutrients and oxygen stop reproducing and generally die, leaving a *necrotic center*. The Gompertz growth law can be interpreted as a description of this necrosis. Discuss this point and give alternative interpretations of equations (6a,b).
8. Suppose that prey have a refuge from predators into which they can retreat. Assume the refuge can hold a fixed number of prey. How would you model this situation, and what predictions can you make?
9. (a) Suppose a one-time fishing expedition reduced the prey population by 10% of its current level. What does the Lotka-Volterra model predict about the subsequent behavior of the system? (Note: this prediction is one of the most objectionable features of the model and will be dealt with in a later chapter.)
 (b) Now consider the situation in which there is a constant level of fishing in which both prey and predatory fish are caught and removed at rates proportional to their densities, ϕx and ϕy . Compare this to the situation in the absence of fishing, and show what Volterra concluded about d'Ancona's observation. (For one treatment of this problem see Braun, 1979; a more advanced mathematical treatment can be found in Brauer and Soudack, 1979.)

*10. In Section 6.2 we showed that the steady state $(\bar{x}_2, \bar{y}_2) = (c/d, a/b)$ is associ-

ated with pure imaginary eigenvalues. Since the equations are nonlinear, it is necessary to consider the possibility that the steady state is a spiral point.

- (a) Write the system of equations in the form

$$\frac{dy}{dx} = \frac{-cy + dxy}{ax - bxy}.$$

Separate variables and integrate both sides to obtain

$$y^a e^{-by} = Kx^c e^{-dx}$$

(where K is an arbitrary constant).

- (b) On the nullcline $x = c/d$ observe that

$$y^a e^{-by} = \text{constant}.$$

Graph the function $f(y) = y^a e^{-by}$ and use your graph to demonstrate that $f(y) = \text{constant}$ can have at most two solutions for any given constant.

- (c) Conclude that the trajectory cannot be a spiral. (*Hint*: Consider how many times it intersects the line $x = c/d$.)

11. Interpret the assumptions 1 to 4 made in the Kolmogorov equations for a predator-prey system.

12. (a) At the end of Section 6.2 a number of modifications of the Lotka-Volterra equations are described. Explain these modifications, paying particular attention to their predictions for low and high values of the prey population x .

- (b) For each modification you discuss in (a), assume it is the only change made in equations (7a,b) and determine the effect on steady states and their stability properties.

13. Show it is possible, by introducing dimensionless variables, to rewrite the Lotka-Volterra equations as

$$\frac{dv}{dt} = v(1 - e), \quad \frac{de}{dt} = \alpha e(v - 1),$$

where v = victims and e = exploiters.

14. Determine whether the nontrivial steady state of equations (8a,b) is a stable node or a stable spiral.

15. In this problem we attend to several details that arise in the species competition model [equations (9a,b)].

- (a) The four cases shown in Figure 6.6 correspond to four possible sets of inequalities satisfied by the parameters κ_1 , κ_2 , β_{12} , and β_{21} . Give a biological interpretation of these relations.

- (b) Show that a fourth steady state to equations (9a,b) is

$$(\bar{x}, \bar{y}) = \left(\frac{\beta_{21}\kappa_1 - \kappa_2}{\beta_{21}\beta_{12} - 1}, \frac{\beta_{12}\kappa_2 - \kappa_1}{\beta_{21}\beta_{12} - 1} \right),$$

and demonstrate that this steady state has biological relevance only in cases 3 and 4.

- (c) The Jacobian of (9a,b) has off-diagonal elements as follows:

$$J = \begin{pmatrix} ? & \frac{-r_1 \beta_{12} N_1}{\kappa_1} \\ \frac{-r_2 \beta_{21} N_2}{\kappa_2} & ? \end{pmatrix}_{(\bar{N}_1, \bar{N}_2)}.$$

Verify this fact, find the remaining entries, and then compute J for each of the steady states of (9a,b). (Note: one of these involves rather cumbersome expressions.)

- (d) Verify the stability properties of steady states to the species competition model [equations (9a,b)] in the four cases discussed in Section 6.3.
- (e) Give a biological interpretation of cases 1 through 4 in Figure 6.8.
- (1) What is the outcome of competition in case 1? How does this differ from case 3?
 - (2) In cases 1, 2, and 4 the final outcome does not depend on the initial levels of the competing populations. This is not true in case 3. Give a rough rule of thumb for determining which species wins in case 3.
- *16. In this problem you are asked to generalize the competition model of equations (9a,b) to the case of k species, with particular emphasis on $k = 3$.
- (a) Write a set of equations for the populations of species 1, 2, . . . , k that compete pairwise as in equations (9a,b).
 - (b) For $k = 3$ how many steady states are possible? (Hint: in the absence of a third species, each pair behaves according to the original two-species competition model.)
 - (c) The accompanying figure sketches the case roughly corresponding to case 4. Now suppose that the populations are such that (1) species 2 wins

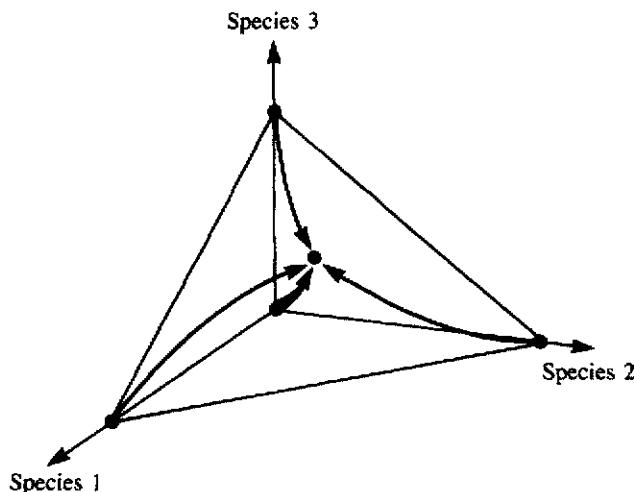


Figure for problem 16(c).

when 3 is absent; (2) species 3 wins when 1 is absent; and (3) species 1 wins when 2 is absent. Sketch the expected dynamics.

17. Species may derive mutual benefit from their association; this type of interaction is known as *mutualism*. May (1976) suggests the following set of equations to describe a possible pair of mutualists:

$$\frac{dN_1}{dt} = rN_1 \frac{1 - N_1}{\kappa_1 + \alpha N_2}, \quad \frac{dN_2}{dt} = rN_2 \frac{1 - N_2}{\kappa_2 + \beta N_1},$$

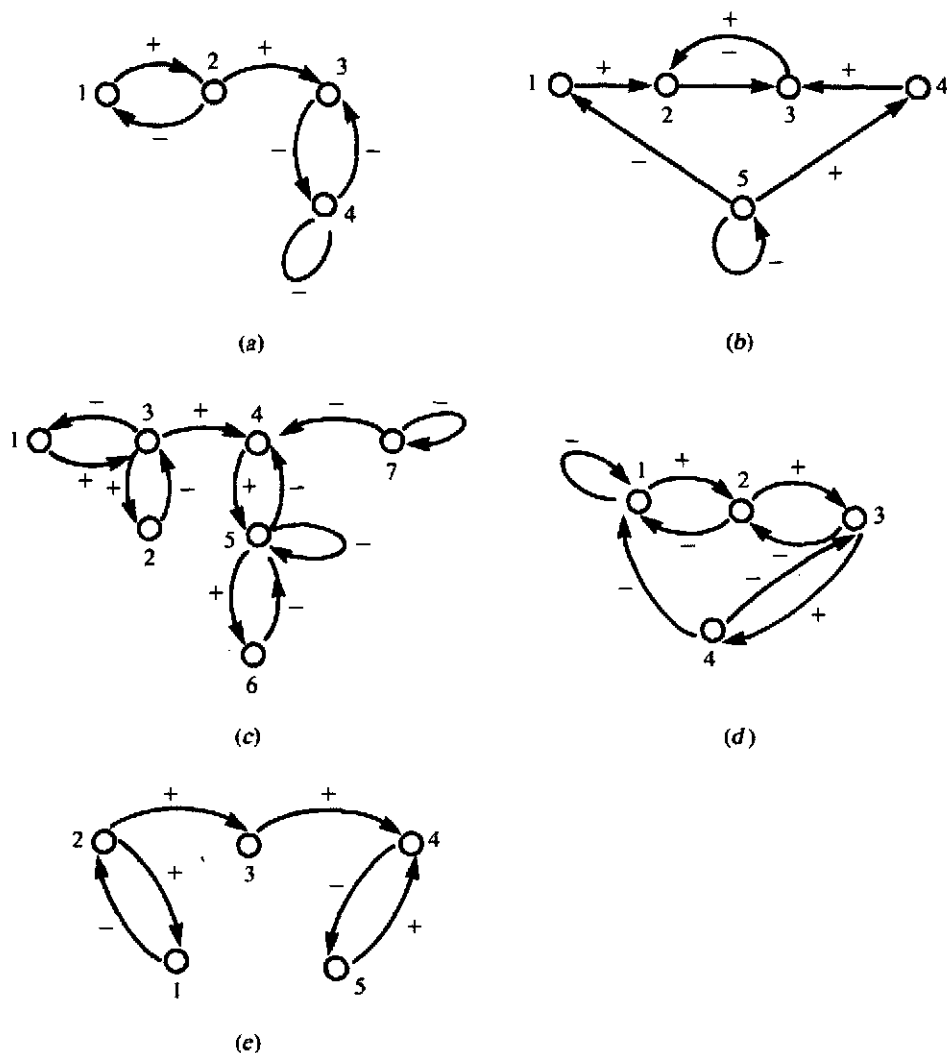
where N_i is the population of the i th species, and $\alpha\beta < 1$.

- Explain why the equations describe a mutualistic interaction.
 - Determine the qualitative behavior of this model by phase-plane and linearization methods.
 - Why is it necessary to assume that $\alpha\beta < 1$?
18. Write down all Routh-Hurwitz matrices H_1 , H_2 , and H_3 for the case of three species. Show that May's conditions are equivalent to the original Routh-Hurwitz criteria by evaluating the determinants of these matrices.
19. Suppose that in the three-species model discussed in example 1 (Section 6.4), species x and z are competitors. How would the model and its conclusions change?
20. Suppose that in the same example species y is also a prey of species z . How would the model and its conclusions change?
21. In the accompanying directional graph, arrows represent positive and negative effects that each of three species exert on each other when they are at equilibrium. For example, the effect of y on x is negative. Use the Routh-Hurwitz criteria to show that this system must be unstable.



Figure for problem 21.

22. Use the Routh-Hurwitz criteria to investigate stability in problem 29 of Chapter 4.
23. Analyze the community structures shown in the accompanying figures in the following way:
- Give the patterns of signs in the qualitative matrix Q that describes the system.
 - Identify predation communities.
 - Determine whether or not the system is qualitatively stable. If not, identify which condition(s) it does not satisfy.
 - Suggest what kind of community might be represented by the graph.



Figures (a-e) for problem 23.

24. Draw directed graphs, identify predation communities, and determine whether the systems depicted in the following qualitative matrices are stable or not.

(a)
$$\begin{bmatrix} 0 & + & 0 & 0 \\ - & + & 0 & - \\ 0 & 0 & 0 & + \\ 0 & + & - & 0 \end{bmatrix}$$

(Levins, 1977: one of two competitors for a common resource is itself preyed on: the other is resistant.)

(b)
$$\begin{bmatrix} 0 & 0 & 0 & + \\ - & 0 & + & + \\ + & - & - & 0 \\ - & - & 0 & - \end{bmatrix}$$

(Levins, 1974: nutrients and organisms in a lake. Two species are algae; the others represent nutrients, one of which is produced by an alga.)

- (c)
$$\begin{bmatrix} - & - & 0 & 0 & + \\ + & - & - & 0 & 0 \\ 0 & + & - & - & 0 \\ 0 & 0 & + & - & - \\ - & 0 & 0 & + & 0 \end{bmatrix}$$
 Unlikely food chain.
- (d)
$$\begin{bmatrix} 0 & - & 0 & 0 \\ + & - & 0 & 0 \\ + & 0 & - & + \\ 0 & + & - & 0 \end{bmatrix}$$
 A larval prey and predator whose roles reverse in adulthood.
- (e)
$$\begin{bmatrix} - & 0 & 0 & 0 & 0 \\ + & - & 0 & 0 & 0 \\ + & 0 & - & 0 & 0 \\ 0 & + & 0 & 0 & - \\ 0 & 0 & + & - & 0 \end{bmatrix}$$
 A species that can exist in two types that compete in adulthood.

25. In this problem you are asked to verify several results quoted in Section 6.6.
- Show that $N = S + I + R$ is constant for the *SIRS* model given by equations (28a,b,c).
 - Verify that the two steady states of (28) are given by (29a,b).
 - Suppose that all members of a population give birth to susceptibles (at rate δ) and die (at rate δ). How would the equations change?
 - Find steady states for part (c), and determine what the infectious contact number is in terms of the parameters.
 - Compare the model with and without the above vital dynamics.
- *26. Analyze an *SIR* model with and without vital dynamics. Verify the results summarized in Table 6.1.
- *27. Show that in an *SIR* model with disease fatality at rate η the disease will always eventually disappear.
- *28. Show that in an *SIR* model with carriers who show no symptoms of the disease, the disease always remains endemic.
29. Capasso and Serio (1978) considered the following model with emigration of susceptibles:

$$\frac{dS}{dt} = -g(I)S - \lambda S,$$

$$\frac{dI}{dt} = g(I)S - \gamma I,$$

$$\frac{dR}{dt} = \lambda S + \gamma I.$$

The function $g(I)$, shown in the accompanying graph, takes into account "psychological" effects. Explain the equations and show that the epidemic will always tend to extinction with respect to both infectives and susceptibles.

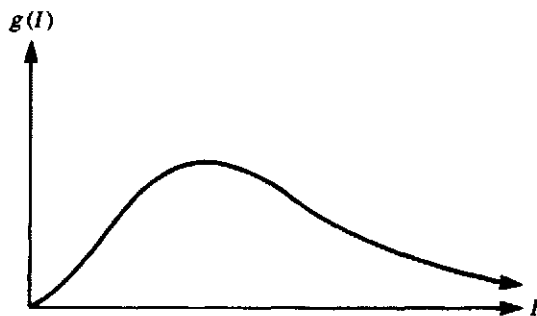


Figure for problem 29.

30. Anderson and May (1979) describe a model in which the natural birth and death rates, a and b , are not necessarily equal so that the disease-free population may grow exponentially:

$$\frac{dN}{dt} = (a - b)N.$$

The disease increases mortality of infected individuals (additional rate of death by infection = α), as shown in the accompanying figure. In their terminology the population consists of the following:

X = susceptible class (= S),

Y = infectious class (= I),

Z = temporarily immune class (= R).

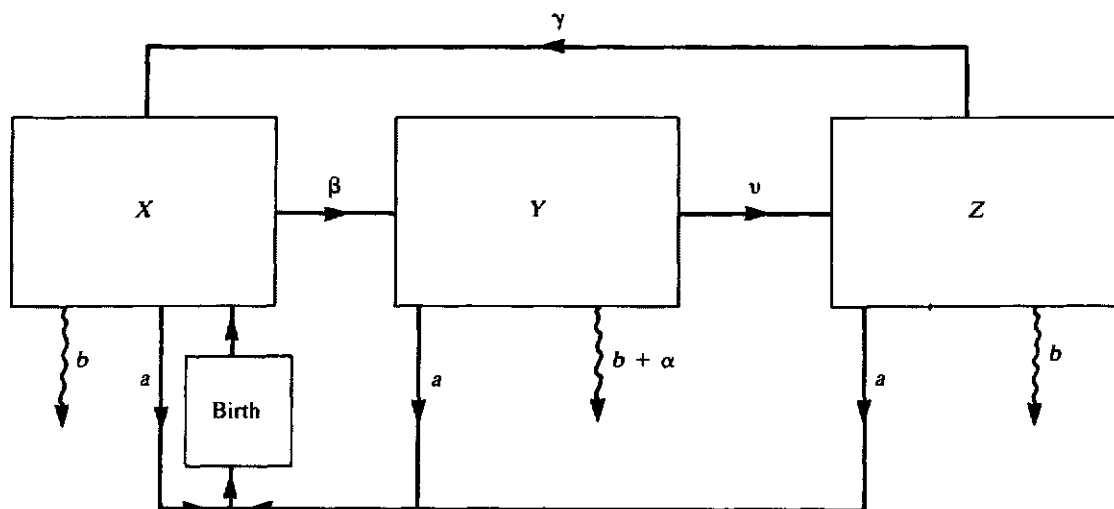


Figure for problem 30.

- Write equations describing the disease.
- Show that the steady-state solution representing the presence of disease is given by

$$\bar{X}_2 = \frac{\alpha + b + \nu}{\beta}, \quad \bar{Y}_2 = \frac{r}{\alpha} N, \quad \bar{Z}_2 = \frac{r}{\alpha} \left(\frac{\nu}{b + \gamma} \right) N_2,$$

where

$$N_2 = \frac{\alpha(\alpha + b + \nu)}{\beta(\alpha - r[1 + \nu/(b + \gamma)])}, \quad \text{and} \quad r = a - b.$$

- (c) The steady state in part (b) makes biological sense whenever

$$\alpha > r \left(1 + \frac{\nu}{b + \gamma} \right).$$

Interpret this inequality.

(Note: Anderson and May conclude that disease can be a regulating influence on the population size.)

31. Consider a lake with some fish attractive to fishermen. We wish to model the fish-fishermen interaction.

Fish Assumptions:

- i. Fish grow logistically in the absence of fishing.
- ii. The presence of fishermen depresses fish growth at a rate jointly proportional to the fish and fishermen populations.

Fishermen Assumptions:

- i. Fishermen are attracted to the lake at a rate directly proportional to the amount of fish in the lake.
 - ii. Fishermen are discouraged from the lake at a rate directly proportional to the number of fishermen already there.
- (a) Formulate, analyze, and interpret a mathematical model for this situation.
- (b) Suppose the department of fish and game decides to stock the lake with fish at a constant rate. Formulate, analyze, and interpret a mathematical model for the situation with stocking included. What effect does stocking have on the fishery?

32. Beddington and May (1982) have proposed the following model to study the interactions between baleen whales and their main food source, *krill* (a small shrimp-like animal), in the southern ocean:

$$\dot{x} = rx \left(1 - \frac{x}{K} \right) - axy$$

$$\dot{y} = sy \left(1 - \frac{y}{bx} \right)$$

Here the whale carrying capacity is not constant but is a function of the krill population:

$$K_{\text{whales}} = bx$$

Analyze this model by determining the steady states and their stability; include a phase-plane diagram.

33. *Estimating parameters in the logistic equation:* We would like to address the curve-fitting problem of how to find the logistic equation of best fit when given appropriate data. First, rearrange equation (2b) and show that

$$N(t) = \frac{K}{1 + [(K - N_0)/N_0]e^{-rt}}.$$

Then conclude that

$$N(t) = \frac{K}{1 + \exp \{-rt + \ln [(K - N_0)/N_0]\}}.$$

Now rearrange the above to demonstrate that

$$\frac{K - N}{N} = \exp \left(-rt + \ln \frac{K - N_0}{N_0} \right).$$

Thus

$$\ln \frac{K - N}{N} = -rt + \ln \frac{K - N_0}{N_0}.$$

Now the quantity $\ln [(K - N)/N]$ is a linear function of time with slope r . The following data is given in Gause (1969) for the growth in volume of each of the yeasts *Saccharomyces* and *Schizosaccharomyces* growing separately. (See Figure 6.9, top curves in each graph.)

Age (h)	(1) <i>Saccharomyces</i>	(2) <i>Schizosaccharomyces</i>
6	0.37	—
16	8.87	1.00
24	10.66	—
29	12.50	1.70
40	13.27	—
48	12.87	2.73
53	12.70	—
72	—	4.87
93	—	5.67
141	—	5.83

- (a) Use the maximal population levels to estimate K_1 and K_2 , the carrying capacities for each of the two species.
- (b) Plot $(K_i - N_i)/N_i$ on a log scale and use this to determine r_1 and r_2 . Do your values agree with those given by Gause in Figure 6.9?
34. Estimating the species-competition parameters.
- (a) Use equations (9a,b) to show that

$$\beta_{12} = \frac{K_1 - (dN_1/dt)K_1 - N_1}{r_1 N_1 N_2},$$

$$\beta_{21} = \frac{K_2 - (dN_2/dt)K_2 - N_2}{r_2 N_1 N_2}.$$

- (b) Suggest how such parameters might be estimated given empirical observations of N_1 and N_2 growing in a mixed population. (See Gause for experimental data.)
 - (c) The values of all parameters determined by Gause are given in Section 6.3, Figure 6.9. Determine whether the two species can coexist in a stable mixed population or if one wins over the other.
35. Populations of lemmings, voles, and other small rodents are known to fluctuate from year to year. Early Scandinavians believed the lemmings to fall down from heaven during stormy weather. Later in history, the legend developed that they migrate periodically into the sea for suicide in order to reduce their numbers. . . . None of these theories, however, was supported by any accurate observations. (H. Dekker, 1975)

An alternate hypothesis was suggested by Dekker to account for rodent population cycles. His theory is based on the idea that the rodents fall into two genotypic classes (Myers and Krebs, 1971) that interact. Type 1 reproduces rapidly, but migrates in response to overcrowding; type 2 is less sensitive to high densities but has a lower reproductive capacity.

The following simple mathematical model was given by Dekker to demonstrate that oscillations could be produced when types 1 and 2 rodents were both present in the population:

$$\begin{aligned}\frac{dn_1}{dt} &= n_1[a_1 - (b_1 - c_1)n_2 - c_1(n_1 + n_2)], \\ \frac{dn_2}{dt} &= n_2[-a_2 + b_2n_1],\end{aligned}$$

where

$$\begin{aligned}n_1 &= \text{density per acre of type 1,} \\ n_2 &= \text{density per acre of type 2.}\end{aligned}$$

The term $b_1 - c_1$ was chosen for convenience in the mathematical calculations rather than for particular biological reasons.

- (a) On the basis of the information, give an interpretation of the individual terms in the equations.
- (b) Using phase-plane methods, determine the qualitative behavior of solutions to Dekker's equations. If there is more than one case, pay particular attention to the case in which oscillations are present. Give conditions on the parameters a_j , b_j , and c_j for which oscillatory behavior will be seen.
- (c) Give a short critique of Dekker's model, indicating whether you would change his assumptions and/or equations. Dekker's article has received a somewhat critical peer review by Nichols et. al. (1979). You may wish to comment on their specific points of contention.

REFERENCES

Single Species Growth

- Lamberson, R., and Biles, C. (1981). Polynomial models of biological growth, *UMAP Journal*, 2 (2), 9–25.
- Malthus, T. R. (1798). An essay on the principle of population, and A summary view of the principle of population. Penguin, Harmondsworth, England.
- Pearl, R., and Reed, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc. Natl. Acad. Sci. USA*, 6, 275–288.
- Slobodkin, L. B. (1954). Population dynamics in *Daphnia obtusa* Kurz. *Ecol. Monogr.*, 24, 69–88.
- Smith, F. E. (1963). Population dynamics in *Daphnia magna*. *Ecology*, 44, 651–663.
- Verhulst, P. F. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathematique et Physique*, 10, 113–121.

Predator-Prey Interactions and Species Competition

- Armstrong, R. A., and McGehee, R. (1980). Competitive exclusion. *Am. Nat.*, 115(2), 151–170.
- Brauer, F., and Soudack, A. C. (1979). Stability regions and transition phenomena for harvested predator-prey systems. *J. Math. Biol.*, 7, 319–337.
- Braun, M. (1979). *Differential Equations and Their Applications*. 3d ed., Springer-Verlag, New York.
- Braun, M. (1983). Chaps. 4, 5, 15, 17 in M. Braun, C. S. Coleman, and D. A. Drew, eds. *Differential Equation Models*. Springer-Verlag, New York.
- Coleman, C. S. (1978). Biological cycles and the fivefold way. In M. Braun, C. S. Coleman, and D. A. Drew, eds., *Differential Equation Models*, Springer-Verlag, New York.
- Gause, G. F. (1932). Experimental studies on the struggle for existence. I. Mixed population of two species of yeast. *J. Exp. Biol.*, 9, 389–402.
- Gause, G. F. (1934). *The Struggle for Existence*. Hafner Publishing, New York. (Reprinted 1964, 1969).
- Holling, C. S. (1965). The functional response of predators to prey density and its role in mimicry and population regulation. *Mem. Entomol. Soc. Can.*, 45, 1–60.
- Ivlev, V. S. (1961). *Experimental Ecology of the Feeding of Fishes*. Yale University Press, New Haven, Conn.
- Kolmogorov, A. (1936). Sulla Teoria di Volterra della Lotta per l'Esistenza. *G. Ist. Ital. Attuari*, 7, 74–80.
- Lotka, A. J. (1925). *Elements of Physical Biology*. Williams & Wilkins, Baltimore.
- May, R. (1973). *Stability and Complexity in Model Ecosystems*. Princeton University Press, Princeton, N. J.
- May, R. (1976). *Theoretical Ecology, Principles and Applications*. Saunders, Philadelphia.
- Odum, E. P. (1953). *Fundamentals in Ecology*. Saunders, Philadelphia.
- Pielou, E. C. (1969). *An Introduction to Mathematical Ecology*. Wiley-Interscience, New York. (Reprinted 1977).
- Rosenzweig, M. L. (1969). Why the prey curve has a hump. *Am. Nat.* 103, 81–87.
- Rosenzweig, M. L. (1971). Paradox of enrichment: Destabilization of exploitation ecosystems in ecological time. *Science*, 171, 385–387.

- Roughgarden, J. (1979). *Theory of Population Genetics and Evolutionary Ecology: An Introduction*. Macmillan, New York.
- Schoener, T. W. (1973). Population growth regulated by intraspecific competition for energy or time. *Theor. Pop. Biol.*, 4, 56–84.
- Takahashi, F. (1964). Reproduction curve with two equilibrium points: A consideration on the fluctuation of insect population. *Res. Pop. Ecol.*, 6, 28–36.
- Van der Vaart, H. R. (1983). Some examples of mathematical models for the dynamics of several-species ecosystems. Chap. 4 in H. Marcus-Roberts and M. Thompson, eds. *Life Science Models*. Springer-Verlag, New York.
- Volterra, V. (1926). Variazioni e fluttuazioni del numero d'individui in specie animal conviventi. *Mem. Acad. Lincei.*, 2, 31–113. Translated as an appendix to Chapman, R. N. (1931). *Animal Ecology*. McGraw-Hill, New York.
- Volterra, V. (1931). *Leçons sur la théorie mathématique de la lutte pour la vie*. Gauthier-Villars, Paris.
- Volterra, V. (1937). Principe de biologie mathématique. *Acta Biotheor.*, 3, 1–36.
- Whitaker, R. H., and Levin, S. A., eds. (1975). *Niche: Theory and Application*. Dowden, Hutchinson & Ross, New York.

Qualitative Stability

- Jeffries, C. (1974). Qualitative stability and digraphs in model ecosystems. *Ecology*, 55, 1415–1419.
- Levins, R. (1974). The qualitative analysis of partially specified systems. *Ann. N. Y. Acad. Sci.*, 231, 123–138.
- Levins, R. (1977). Qualitative analysis of complex systems, in D. E. Matthews, ed., *Mathematics and the Life Sciences. Lecture Notes in Biomathematics*, vol. 18, 153–199. Springer-Verlag, New York.
- Roberts, F. S. (1976). *Discrete Mathematical Models, with Applications to Social, Biological and Environmental Problems*. Prentice-Hall, Englewood Cliffs, N. J.
- Quirk, J., and Ruppert, R. (1965). Qualitative economics and the stability of equilibrium. *Rev. Econ. Stud.*, 32, 311–326.

Mathematical Theory of Epidemics

- Anderson, R. M., ed. (1982). *Population Dynamics of Infectious Diseases, Theory and Applications*. Chapman & Hall, New York.
- Anderson, R. M., and May, R. M. (1979). Population biology of infectious diseases, part I. *Nature*, 280, 361–367; part II, *Nature*, 280, 455–461.
- Busenberg, S. N., and Cooke, K. L. (1978). Periodic solutions of delay differential equations arising in some models of epidemics. In *Proceedings of the Applied Nonlinear Analysis Conference, University of Texas*. Academic Press, New York.
- Capasso, V., and Serio, G. (1978). A generalization of the Kermack-McKendrick deterministic epidemic model. *Math. Biosci.*, 42, 43–61.
- Hethcote, H. W.; Stech, H. W.; and van den Driessche, P. (1981). Periodicity and stability in epidemic models: A Survey. In *Differential Equations and Applications in Ecology, Epidemics, and Population Models*. Academic Press, New York, pp. 65–82.
- Hethcote, H. W. (1976). Qualitative analyses of communicable disease models. *Math. Biosci.*, 28, 335–356.

- Kermack, W. O., and McKendrick, A. G. (1927). Contributions to the mathematical theory of epidemics. *Roy. Stat. Soc. J.*, 115, 700–721.
- May, R. M., (1983). Parasitic infections as regulators of animal populations. *Am. Sci.*, 71, 36–45.

Vaccination Programs

- Anderson, R. M., and May, R. M. (1982). The logic of vaccination. *New Scientist*, (November 18, 1982 issue) pp. 410–415.
- Anderson, R. M., and May, R. M. (1983). Vaccination against rubella and measles: Quantitative investigations of different policies. *J. Hyg. Camb.*, 90, 259–325.
- May, R. M. (1982). Vaccination programmes and herd immunity. *Nature*, 300, 481–483.

Miscellaneous

- Aroesty, J., Lincoln, T., Shapiro, N., and Boccia, G. (1973). Tumor growth and chemotherapy: mathematical methods, computer simulations, and experimental foundations. *Math. Biosci.*, 17, 243–300.
- Beddington, J. R., and May, R. M. (1982). The harvesting of interacting species in a natural ecosystem. *Sci. Am.*, November 1982, pp. 62–69.
- Dekker, H. (1975). A simple mathematical model of rodent population cycles. *J. Math. Biol.*, 2, 57–67.
- Freedman, H. I. (1980). *Deterministic Mathematical Models in Population Ecology*. Marcel Dekker, New York.
- Greenwell, R. (1982). *Whales and Krill: A Mathematical Model*. UMAP module unit 610. COMAP, Lexington, Mass.
- Hofbauer, J. (in press). Permanence and persistence of Lotka-Volterra systems.
- Hutchinson, G. E. (1978). *Introduction to Population Ecology*. Yale University Press, New Haven, Conn.
- Kingsland, S. E. (1985). *Modelling Nature*. University of Chicago Press, Chicago.
- Levins, R. (1968). *Evolution in Changing Environments*. Princeton University Press, Princeton.
- Luenberger, D. G. (1979). *Introduction to Dynamic Systems*. Wiley, New York, pp. 328–331.
- May, R. M., Beddington, J. R., Clark, C. W., Holt, S. J., and Laws, R. M. (1979). Management of multispecies fisheries. *Science*, 205, 267–277.
- Myers, J. H., and Krebs, C. J. (1971). Genetic, behavioral, and reproductive attributes of dispersing field voles *Microtus pennsylvanicus* and *Microtus Ochrogaster*. *Ecol. Monogr.*, 41, 53–78.
- Newton, C. M. (1980). Biomathematics in oncology: modelling of cellular systems. *Ann. Rev. Biophys. Bioeng.*, 9, 541–579.
- Nichols, J. D.; Hestbeck, J. B.; and Conley, W. (1979). Mathematical models and population cycles: A critical evaluation of a recent modelling effort. *J. Math. Biol.*, 8, 259–263.
- Nisbet, R. M., and Gurney, W. S. C. (1982). *Modelling Fluctuating Populations*. Wiley, New York.



DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Quantum Dynamics in Biological Systems

The Harvard community has made this article openly available.
[Please share](#) how this access benefits you. Your story matters.

Citation	Shim, Sangwoo. 2012. Quantum Dynamics in Biological Systems. Doctoral dissertation, Harvard University.
Accessed	May 18, 2016 2:53:43 PM EDT
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:10058477
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

©2012 - Sangwoo Shim

All rights reserved.

Thesis advisor

Author

Alán Aspuru-Guzik

Sangwoo Shim

Quantum Dynamics in Biological Systems

Abstract

In the first part of this dissertation, recent efforts to understand quantum mechanical effects in biological systems are discussed. Especially, long-lived quantum coherences observed during the electronic energy transfer process in the Fenna-Matthews-Olson complex at physiological condition are studied extensively using theories of open quantum systems. In addition to the usual master equation based approaches, the effect of the protein structure is investigated in atomistic detail through the combined application of quantum chemistry and molecular dynamics simulations. To evaluate the thermalized reduced density matrix, a path-integral Monte Carlo method with a novel importance sampling approach is developed for excitons coupled to an arbitrary phonon bath at a finite temperature. In the second part of the thesis, simulations of molecular systems and applications to vibrational spectra are discussed. First, the quantum dynamics of a molecule is simulated by combining semiclassical initial value representation and density functional theory with analytic derivatives. A computationally-tractable approximation to the sum-of-states formalism of Raman spectra is subsequently discussed.

Contents

Title Page	i
Abstract	iii
Table of Contents	iv
Citations to Previously Published Work	vi
Acknowledgments	vii
Dedication	ix
1 Introduction	1
1.1 Review of Theoretical Approaches	4
1.1.1 Basics of Open Quantum Systems	4
1.1.2 Redfield Equation	5
1.1.3 Reduced Hierarchical Equation of Motion	15
1.1.4 Haken-Strobl-Reineker Model	18
I Quantum Coherences in Light Harvesting Systems	25
2 Characterization and quantification of the role of coherence in ultrafast quantum biological experiments	26
2.1 Introduction	26
2.2 The Role of Quantum Coherence	28
2.3 Molecular Dynamics Simulations	38
2.4 Quantum Process Tomography	44
3 Atomistic study of the long-lived quantum coherences in the Fenna-Matthews-Olson complex	56
3.1 Introduction	56
3.2 Methods	59
3.2.1 Molecular Dynamics Simulations	59
3.2.2 Exciton Dynamics	63
3.2.3 Quantum Jump Correction to MD Method (QJC-MD)	64

3.3	Results and Discussion	66
3.3.1	Site Energy Distributions	66
3.3.2	Dephasing Rates	68
3.3.3	Simulated Spectra	69
3.3.4	Population Dynamics and Long-lived Quantum Coherence . .	71
3.3.5	Comparison between MD, QJC-MD, HEOM, and HSR Methods	72
3.3.6	Correlation Functions and Spectral Density	74
3.4	Conclusion	80
4	Path integral Monte Carlo with importance sampling for excitons interacting with arbitrary phonon bath environment	83
4.1	Introduction	83
4.2	Theory	85
4.2.1	Path Integral Formulation of the Reduced Thermal Density Matrix	85
4.2.2	Population-Normalized Estimator and Importance Sampling .	90
4.3	Application	93
4.3.1	Alexander’s 1D Test Model	93
4.3.2	Model of a Chromophore Heterodimer with Displaced Harmonic Oscillators	94
4.4	Conclusion	98
II	Simulations of Molecular Systems and Applications	102
5	First-principles semiclassical initial value representation molecular dynamics	103
5.1	Introduction	103
5.2	First-Principles SC-IVR	105
5.3	Potential Fitting and Grid Calculations	110
5.4	First-Principles SC-IVR Calculations	113
5.5	Conclusions	121
6	Simplified Sum-Over-States Approach for Predicting Resonance Raman Spectra	124
6.1	Introduction	124
6.2	Theory	128
6.3	Resonance Raman Spectra of nucleic acid bases	131
6.4	Conclusion	141
7	Summary and Future Directions	143
	Bibliography	145

Citations to Previously Published Work

Chapters 2, 3, 5 and 6 have, apart from minor changes, appeared as the following publications:

“Characterization and quantification of the role of coherence in ultrafast quantum biological experiments using quantum master equations, atomistic simulations, and quantum process tomography,” Patrick Rebentrost*, Sangwoo Shim*, Joel Yuen-Zhou* and Alán Aspuru-Guzik, *Proceedia Chem.* **3**, 332 (2010).

* : Equal contributions.

“Atomistic study of the long-lived quantum coherences in the Fenna-Matthews-Olson complex,” Sangwoo Shim, Patrick Rebentrost, Stéphanie Valleau and Alán Aspuru-Guzik, *Biophys. J.* **102**, 649 (2012).

“First-principles semiclassical initial value representation molecular dynamics,” Michele Ceotto, Sule Atahan, Sangwoo Shim, Gian Franco Tantardini and Alán Aspuru-Guzik, *Phys. Chem. Chem. Phys.* **11**, 3861 (2009).

“Simplified Sum-Over-States Approach for Predicting Resonance Raman Spectra. Application to Nucleic Acid Bases,” Dmitriy Rappoport, Sangwoo Shim and Alán Aspuru-Guzik, *J. Phys. Chem. Lett.* **2**, 1254 (2011).

Acknowledgments

Working with my advisor Professor Alán Aspuru-Guzik during the last five years has been the most inspiring experience of my life, not only academically but personally as well. I owe my deepest gratitude to Alán for guiding me through the exciting, sometimes intimidating world of scientific research. His enthusiasm, intuition, and scientific diligence are something remarkable and if, by writing this thesis, I have made even a small contribution to science, it was only because I have tried to follow in his footsteps as a scientist. I thank Professors Heller and Shakhnovich for giving invaluable advice during my graduate years and also for serving on the committee for this dissertation. I am especially grateful to Professor Kaxiras for willingly consenting to serve as a replacement thesis committee member.

I was lucky enough to have met and worked with many talented scientists in Alán's group and as collaborators. Discussions with them and listening to their insights have contributed to my academic growth to an extent comparable to Alán's guidance. They have included Joel Yuen-Zhou, Roberto Olivares-Amaya, Dmitriy Rappoport, Patrick Rebentrost, Stéphanie Valleau, Alejandro Perdomo-Ortiz, James Whitfield, Ivan Kassal, Leslie Vogt, David Tempel, Kenta Hongo, Mark Watson, Semion Saikin, John Parkhill, Sule Atahan-Evrenk, Jenny Brookes, Man-Hong Yung, Sarah Mostame, Xavier Andrade, Alex Eisfeld, Johannes Hachmann, Jacob Krich, Takatoshi Fujita, Joonsuk Huh, Jarrod McClean, Jacob Sanders, Ryan Babbush, and Michele Ceotto. I want to convey my sincere thanks to all of them.

I thank my Master's advisor, Professor Chaok Seok at Seoul National University, for introducing me to Monte Carlo and molecular dynamics simulations. These techniques have been my specialty ever since. I thank Professor Sangyoub Lee at Seoul

Acknowledgments

National University for his advice and inspiration. He also taught me many subjects including my first quantum mechanics, which were incredibly helpful in building a strong foundation to pursue academic research.

I also thank my friends Changho Sohn, Jinwoo Chang, Hong Geun Lee, Changhyun Ko, Dann Huh, Joungkeun Lim, Youjin Lee, Sungwoo Park, Junyoep Park, Junyeop Lee, Jaehyuk Choi, Jinyoung Baek, Ji Oon Lee, Eun Gook Moon, Eunmi Chae, Sungkun Hong, Yejin Huh, Joonhyun Lee, Jae Hoon Lee, Sae Kyu Lee, Lia Min, Junwon Choi, Ji-hyun Seo, Sanghyup Kwak, Soyeon Kim, SeungYeon Kang, Jung Ook Hong, Kisun Yoon, Won-Yong Shin, Hyunsung Park, Yun-Ah Jang, Jaeyoung Park, Young Il Cho, Jaehong Jung, Siyeon Song, Dong-In Lee, Hyeongryul Park, Dongwan Ha, Hyewon Yoon, Dainn Wie, Tae Wook Kim, Eun-Suk Lee, Kyung Ryul Park, Yoonha Kim, Jang Ik Cho, Nari Yoon, Hee-Sun Han, Semi Park, Yoonjin Lee, Joonhyuk Choi, Sangmin Lim, Dongeun Lee, Min Ju Sohn, Doory Kim, Jaeyoung Ahn, So Youn Shim, Yongho Park, Wooyoung Hong, and Jeong-Mo Choi for everything. This is only a small sample of the people who have helped and inspired me, and I owe a beer to whomever I have omitted from the list.

I thank my mom, Sooja Son and my dad, Yeongho Shim for their continued support and for wholeheartedly being my fans from the day I was born. I would also like to thank to my sister, Nary Shim, for being such a great sister to me. I especially thank my father-in-law, Jeil Song, and brother-in-law, Jisoo Song, for their love and advice. Last, but not least, I thank my beautiful wife, Jisun, and our soon-to-be-born daughter for always standing beside me.

To my family and friends

Chapter 1

Introduction

Although more than 80 years passed of Paul Dirac’s announcement that “the underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are completely known” [1], tremendous amount of efforts are still being made to achieve computationally-scalable simulations for quantum dynamics and their associated chemical phenomena. The cost of solving the time-dependent Schrödinger equation increases very quickly as the size of the system grows and as the total length of the propagation time gets longer. Even for a modest sized system, exact quantum mechanical dynamics easily becomes untractable with currently available computational resources. Therefore, most of the useful approaches for treating biological systems inevitably involve approximations to some extent. For example, the structure and behavior of protein complexes found in biology are explained well in terms of classical statistical mechanics and molecular dynamics, which are approximations of quantum statistical mechanics and time evolution [2–6]. For a larger system, even classical mechanical calculations are very hard to carry out.

Therefore, many coarse-grained and multiscale simulation methods have been suggested and are still actively being developed [7–9]. Nevertheless, classical mechanics has been the method of choice for studying biological systems in molecular level.

The Fenna-Matthews-Olson complex is a trimeric bacteriochlorophyll protein in the light-harvesting system of green sulfur bacteria [10, 11]. This complex transfers the energy of the photons collected at the photosynthetic antenna complex to the reaction center [12]. Eight bacteriochlorophyll (BChl) molecules each of which acting as a chromophore are embedded in its monomer. Because its high-resolution X-ray structure has been known for a long time, this subsystem has been studied extensively by theoreticians [13, 14] as well as experimentalists [15–18]. Early efforts were mostly focused on evaluating the Hamiltonian relevant to the spectroscopic measurement. BChl molecules were modelled as two-level systems interacting each other through electronic Förster’s dipole-dipole coupling [19]. Each BChl molecule was also assumed to coupled to a harmonic oscillator bath to give the line broadening of the spectroscopic spectra. Within this assumption, the electronic Hamiltonian operator was evaluated by fitting to linear absorption spectra [13, 20], interpreting 2D electronic spectroscopy data [21], calculation based on force fields [14] and density functional theory [22].

Recent 2D nonlinear spectroscopy experiments suggested the existence of long-lived quantum coherences lasting up to several hundreds of femtoseconds during the electronic energy transfer process in certain photosynthetic subsystems, especially within a Fenna-Matthews-Olson (FMO) complex of sulfur bacterium, even under physiological conditions [23–25]. Moreover, the observed quantum coherences are

thought to contribute the energy transfer efficiency [26, 27]. Apparently, this energy transfer dynamics cannot be explained without quantum mechanics. Moreover, traditional master equation with Born-Markov approximation proven to be unable to reproduce this long-lived coherence [28]. Thus, more advanced theories of open quantum systems have been applied to explain the dynamics of excitons in FMO complex with some degree of success [29–33], and still being actively developed.

In the Part I of this dissertation, the efforts we made to understand those long lived quantum coherences in biological systems are presented; Chapter 2 presents a review on three approaches made in our group to characterize quantum effects in the FMO complex. Chapter 3 is about the atomistic simulation to include the effects from the realistic environment to the dynamics of excitons. Equilibrium properties of the reduced density matrix of excitons coupled to an arbitrary bath are explored in Chapter 4 using the path integral Monte Carlo method with importance sampling. Part II features two projects on calculations absorption and resonance Raman spectra based on approximate quantum dynamics of the molecular system, respectively. Chapter 5 introduces an approximate but very accurate real space wavefunction propagation in real time using time-averaged semiclassical initial value representation implemented on top of the *ab initio* molecular dynamics. A simplified and computationally tractable formulation of the resonance Raman scattering cross section using time dependent density functional theory and analytic derivatives is presented in chapter 6.

1.1 Review of Theoretical Approaches

For better understanding of the materials included this dissertation, introductions to basic concepts and relevant theories will be provided in the rest of the current chapter.

1.1.1 Basics of Open Quantum Systems

The density matrix of a closed system evolves according to the quantum Liouville equation. As elaborated in the previous section, explicit evaluation of a quantum mechanical time evolution becomes easily untractable, especially when there exist a large number of degrees of freedom. Fortunately, we are concerned with only a part of the total system in most cases. Consider an electronic energy transfer process in a biomolecular system; the entire system encompassing all electronic and vibrational degrees of freedom of the molecules and solvents should be, in principle, explicitly propagated to obtain the exact dynamics. But we are interested only in the electronic state of chromophores, which exists in a Hilbert space with only a few degrees of freedom. Therefore, if an equation of motion for such a part of the total system can be derived, all the information we need to obtain the solution for the problem can be identified. Theories for treating such a reduced quantum system interacting with a macroscopic environment is referred to theories of open quantum systems. The part of our interest is called *the system*, whereas the rest of the total system is referred as *the bath*. The partitioning is entirely determined by the decision of the physicist, although there may be an obvious choice for the system and the bath in many cases. Given a total density matrix, the system and the bath density matrices can be defined in a

very similar way to a marginal probability density in the probability theory. Given a probability density function of two random variables X and Y is given as $P_{X,Y}(x, y)$, the marginal density for X is given as,

$$P_X(x) = \int dy P_{X,Y}(x, y), \quad (1.1)$$

which is an effective probability density only for X . This density contains every information we need if we are only interested in the distribution of X even though the actual random process produces a random vector (X, Y) . Now a reduced density matrix of the system can be defined as an effective, averaged density matrix over its bath by tracing out the bath degrees of freedom,

$$\rho_S(t) \equiv \text{Tr}_B \rho(t), \quad (1.2)$$

where $\rho(t)$ is the density matrix of the total system and Tr_B is the partial trace operator which traces out the bath degrees of freedom. In the following sections, two types of equations of motion for the reduced density matrix of the system will be introduced based on different sets of approximations.

1.1.2 Redfield Equation

We will discuss the general formulation of the Redfield equation first and then focus on the application on the electronic energy transfer dynamics. The Hamiltonian for the total system can be decomposed as three components:

$$\hat{H}_{total} = \hat{H}_S + \hat{H}_B + \hat{H}_{SB}, \quad (1.3)$$

where the system Hamiltonian \hat{H}_S only acts on the Hilbert space of the system and the bath Hamiltonian \hat{H}_B only act on the Hilbert space of the bath. Rest of the total

Hamiltonian causing the entanglement between the system and the bath is specified as the system-bath Hamiltonian \hat{H}_{SB} . By choosing the interaction picture relative to $\hat{H}_S + \hat{H}_B$ as the zeroth order Hamiltonian, the quantum mechanical equation of the motion for the total density matrix can be obtained:

$$\frac{d\tilde{\rho}(t)}{dt} = \frac{1}{i\hbar} \left[\tilde{H}_{SB}(t), \tilde{\rho}(t) \right], \quad (1.4)$$

where

$$\begin{aligned} \hat{U}_0(0, t) &= e^{-\frac{i}{\hbar} \int_0^t \hat{H}_S(s) + \hat{H}_B(s) ds}, \\ \tilde{\rho}(t) &= \hat{U}_0^\dagger(0, t) \rho(t) \hat{U}_0(0, t), \\ \tilde{H}_{SB}(t) &= \hat{U}_0^\dagger(0, t) \hat{H}_{SB}(t) \hat{U}_0(0, t). \end{aligned} \quad (1.5)$$

Operators with a tilde are in the interaction picture. Closed form for $\tilde{\rho}(t)$ can be obtained by integrating Eq. 1.4:

$$\tilde{\rho}(t) = \tilde{\rho}(0) + \frac{1}{i\hbar} \int_0^t ds \left[\tilde{H}_{SB}(s), \tilde{\rho}(s) \right]. \quad (1.6)$$

Tracing over the bath degrees of freedom and plugging in Eq. 1.4 gives

$$\begin{aligned} \frac{d\tilde{\rho}_S(t)}{dt} &= \frac{1}{i\hbar} \text{Tr}_B \left[\tilde{H}_{SB}(t), \tilde{\rho}(t) \right] \\ &= \frac{1}{i\hbar} \text{Tr}_B \left[\tilde{H}_{SB}(t), \tilde{\rho}(0) \right] - \frac{1}{\hbar^2} \int_0^t ds \text{Tr}_B \left[\tilde{H}_{SB}(t), \left[\tilde{H}_{SB}(s), \tilde{\rho}(s) \right] \right]. \end{aligned} \quad (1.7)$$

Without the loss of generality, $\tilde{H}_{SB}(t)$ can be expanded as a linear combination of factorized operators:

$$\tilde{H}_{SB}(t) = \sum_k \tilde{A}_k(t) \otimes \tilde{B}_k(t) = \sum_k \tilde{A}_k^\dagger(t) \otimes \tilde{B}_k^\dagger(t). \quad (1.8)$$

Note that individual $\tilde{A}_k^\dagger(t)$ and $\tilde{B}_k^\dagger(t)$ might not be Hermitian even though $\tilde{H}_{SB}(t)$ is Hermitian. A series of assumptions needs to be introduced to proceed further. The

first assumption is called the Born approximation, which states that the total density matrix is factorizable at all times, and the bath state is in thermal equilibrium so it does not depend on time:

$$\tilde{\rho}(t) \approx \tilde{\rho}_S(t) \otimes \tilde{\rho}_B, \quad (1.9)$$

$$\tilde{\rho}_B = \rho_B = \frac{\exp(-\beta \hat{H}_B)}{\text{Tr}_B \exp(-\beta \hat{H}_B)}. \quad (1.10)$$

By plugging in Eq. 1.8 and Eq. 1.9 to Eq. 1.7,

$$\begin{aligned} \frac{d\tilde{\rho}_S(t)}{dt} &= \frac{1}{i\hbar} \sum_k \langle \tilde{B}_k(t) \rangle [\tilde{A}_k(t), \tilde{\rho}_S(0)] \\ &\quad - \frac{1}{\hbar^2} \sum_{k,l} \int_0^t ds \langle \tilde{B}_k^\dagger(t) \tilde{B}_l(s) \rangle \tilde{A}_k^\dagger(t) \tilde{A}_l(s) \tilde{\rho}_S(s) \\ &\quad + \frac{1}{\hbar^2} \sum_{k,l} \int_0^t ds \langle \tilde{B}_l(s) \tilde{B}_k^\dagger(t) \rangle \tilde{A}_k^\dagger(t) \tilde{\rho}_S(s) \tilde{A}_l(s) \\ &\quad + \frac{1}{\hbar^2} \sum_{k,l} \int_0^t ds \langle \tilde{B}_k^\dagger(t) \tilde{B}_l(s) \rangle \tilde{A}_l(s) \tilde{\rho}_S(s) \tilde{A}_k^\dagger(t) \\ &\quad - \frac{1}{\hbar^2} \sum_{k,l} \int_0^t ds \langle \tilde{B}_l(s) \tilde{B}_k^\dagger(t) \rangle \tilde{\rho}_S(s) \tilde{A}_l(s) \tilde{A}_k^\dagger(t) \\ &= \frac{1}{i\hbar} \sum_k \langle \tilde{B}_k(t) \rangle [\tilde{A}_k(t), \tilde{\rho}_S(0)] \\ &\quad - \frac{1}{2\hbar^2} \sum_{k,l} \int_0^t ds \langle \{ \tilde{B}_k^\dagger(t), \tilde{B}_l(s) \} \rangle \left([\tilde{A}_k^\dagger(t), [\tilde{A}_l(s), \tilde{\rho}_S(s)]] \right) \\ &\quad - \frac{1}{2\hbar^2} \sum_{k,l} \int_0^t ds \langle [\tilde{B}_k^\dagger(t), \tilde{B}_l(s)] \rangle \left([\tilde{A}_k^\dagger(t), \{ \tilde{A}_l(s), \tilde{\rho}_S(s) \}] \right), \end{aligned} \quad (1.11)$$

where $\langle \tilde{O} \rangle = \text{Tr}_B [\tilde{O} \tilde{\rho}_B]$. Because the bath is assumed to be in thermal equilibrium, the bath correlation function is stationary and only depends on the difference of the two times:

$$c_{kl}(s) = \frac{1}{\hbar} \langle \tilde{B}_k^\dagger(s) \tilde{B}_l(0) \rangle = \frac{1}{\hbar} \langle \tilde{B}_k^\dagger(t) \tilde{B}_l(t-s) \rangle. \quad (1.12)$$

It is convenient to define the symmetrized correlation function $S_{kl}(t)$ and the response function $\chi_{kl}(t)$:

$$S_{kl}(t) = \frac{1}{\hbar} \left\langle \left\{ \tilde{B}_k^\dagger(t), \tilde{B}_l(0) \right\} \right\rangle = c_{kl}(t) + c_{kl}^*(t), \quad (1.13)$$

$$\chi_{kl}(t) = \frac{i}{\hbar} \left\langle \left[\tilde{B}_k^\dagger(t), \tilde{B}_l(0) \right] \right\rangle = i \{c_{kl}(t) - c_{kl}^*(t)\}. \quad (1.14)$$

$S_{kl}(t)$ and $\chi_{kl}(t)$ are often referred to as the noise and dissipation kernels, respectively [34]. The Eq. 1.11 can be rewritten in terms of these two real functions:

$$\begin{aligned} \frac{d\tilde{\rho}_S(t)}{dt} &= \frac{1}{i\hbar} \sum_k \left\langle \tilde{B}_k(t) \right\rangle \left[\tilde{A}_k(t), \tilde{\rho}_S(0) \right] \\ &\quad - \frac{1}{\hbar} \sum_{k,l} \int_0^t ds \frac{1}{2} S_{kl}(t-s) \left[\tilde{A}_k(t), \left[\tilde{A}_l(s), \tilde{\rho}_S(s) \right] \right] \\ &\quad + \frac{1}{\hbar} \sum_{k,l} \int_0^t ds \frac{i}{2} \chi_{kl}(t-s) \left[\tilde{A}_k(t), \left\{ \tilde{A}_l(s), \tilde{\rho}_S(s) \right\} \right]. \end{aligned} \quad (1.15)$$

By changing the integration variable to $t-s$,

$$\begin{aligned} \frac{d\tilde{\rho}_S(t)}{dt} &= \frac{1}{i\hbar} \sum_k \left\langle \tilde{B}_k(t) \right\rangle \left[\tilde{A}_k(t), \tilde{\rho}_S(0) \right] \\ &\quad - \frac{1}{\hbar} \sum_{k,l} \int_0^t ds \frac{1}{2} S_{kl}(s) \left[\tilde{A}_k(t), \left[\tilde{A}_l(t-s), \tilde{\rho}_S(t-s) \right] \right] \\ &\quad + \frac{1}{\hbar} \sum_{k,l} \int_0^t ds \frac{i}{2} \chi_{kl}(s) \left[\tilde{A}_k(t), \left\{ \tilde{A}_l(t-s), \tilde{\rho}_S(t-s) \right\} \right]. \end{aligned} \quad (1.16)$$

Now we introduce the second assumption which states that the bath is stationary and its correlation function decays rapidly:

$$c_{kl}(s) \approx 0 \quad \text{for } s > \tau_c. \quad (1.17)$$

This assumption will let us integrate up to infinite time in the second term of Eq. 1.16.

Moreover, if $\tilde{\rho}_S(t)$ does not change much during the characteristic time τ_c , $\tilde{\rho}_S(t-s)$

in the integrand can be approximated as $\bar{\rho}_S(t)$ and a Markovian equation of motion is obtained:

$$\begin{aligned} \frac{d\tilde{\rho}_S(t)}{dt} \approx & \frac{1}{i\hbar} \sum_k \langle \tilde{B}_k(t) \rangle \left[\tilde{A}_k(t), \tilde{\rho}_S(0) \right] \\ & - \frac{1}{\hbar} \sum_{k,l} \int_0^\infty ds \frac{1}{2} S_{kl}(s) \left[\tilde{A}_k(t), \left[\tilde{A}_l(t-s), \tilde{\rho}_S(t) \right] \right] \\ & + \frac{1}{\hbar} \sum_{k,l} \int_0^\infty ds \frac{i}{2} \chi_{kl}(s) \left[\tilde{A}_k(t), \left\{ \tilde{A}_l(t-s), \tilde{\rho}_S(t) \right\} \right]. \end{aligned} \quad (1.18)$$

Eq. 1.18 is referred as the Redfield equation. When a model of the bath correlation function is given, this equation can be integrated to give a complete Markovian master equation.

The electronic and phonon Hamiltonians of a typical Frenkel exciton can be specified as [29]:

$$\hat{H}_{el} = \sum_n \varepsilon_n |n\rangle \langle n| + \sum_{m \neq n} E_{mn} |m\rangle \langle n|, \quad (1.19)$$

$$\hat{H}_{ph} = \sum_i \frac{\hat{p}_i^2}{2m_i} + \frac{1}{2} m_i \omega_i^2 \hat{q}_i^2 = \sum_i \hbar \omega_i \left(\hat{a}_i^\dagger \hat{a}_i + \frac{1}{2} \right), \quad (1.20)$$

where the lowering and raising operators of the i th mode are

$$\hat{a}_i = \sqrt{\frac{m_i \omega_i}{2\hbar}} \left(\hat{q}_i + \frac{i}{m_i \omega_i} \hat{p}_i \right), \quad (1.21)$$

$$\hat{a}_i^\dagger = \sqrt{\frac{m_i \omega_i}{2\hbar}} \left(\hat{q}_i - \frac{i}{m_i \omega_i} \hat{p}_i \right), \quad (1.22)$$

with the commutation relation $[\hat{a}_i, \hat{a}_j^\dagger] = \delta_{ij}$. Using the displaced harmonic oscillator model, the electronic phonon interaction Hamiltonian can be specified in the following

way:

$$\begin{aligned}
 \hat{H}_{el-ph} &= \sum_n |n\rangle\langle n| \otimes \sum_i \frac{1}{2} m_i \omega_i^2 [(\hat{q}_i - d_{ni})^2 - \hat{q}_i^2] \\
 &= \sum_n \left(\sum_i \frac{1}{2} m_i \omega_i^2 d_{ni}^2 \right) |n\rangle\langle n| + \sum_n |n\rangle\langle n| \otimes \left(- \sum_i m_i \omega_i^2 d_{ni} \hat{q}_i \right) \\
 &= \sum_n \left(\sum_i \frac{1}{2} m_i \omega_i^2 d_{ni}^2 \right) |n\rangle\langle n| + \sum_n |n\rangle\langle n| \otimes \left(- \sum_i m_i \omega_i^2 d_{ni} (\hat{a}_i + \hat{a}_i^\dagger) \sqrt{\frac{\hbar}{2m_i \omega_i}} \right) \\
 &= \sum_n \underbrace{\left(\sum_i \frac{1}{2} m_i \omega_i^2 d_{ni}^2 \right)}_{\lambda_n} |n\rangle\langle n| + \sum_n \underbrace{|n\rangle\langle n|}_{\hat{A}_n} \otimes \underbrace{\left(- \sum_i \sqrt{\frac{\hbar m_i \omega_i^3}{2}} d_{ni} (\hat{a}_i + \hat{a}_i^\dagger) \right)}_{\hat{B}_n} \\
 &= \underbrace{\sum_n \lambda_n |n\rangle\langle n|}_{\hat{H}_{reorg}} + \underbrace{\sum_n \hat{A}_n \otimes \hat{B}_n}_{\hat{H}_{SB}}. \tag{1.23}
 \end{aligned}$$

where ω_i , \hat{q}_i , \hat{p}_i , \hat{a}_i^\dagger and \hat{a}_i are the angular frequency, position operator, momentum operator, raising and lowering operators for the i th normal mode coordinate, respectively. d_{ni} is the displacement of the i th oscillator for the n th exciton and only the Franck-Condon transition is assumed to occur during the dynamics. To apply the Redfield equation, the decomposition of the total system into the system and bath will be done in the following way:

$$\hat{H}_S = \hat{H}_{el} + \hat{H}_{reorg} = \sum_n (\varepsilon_n + \lambda_n) |n\rangle\langle n| + \sum_{m \neq n} E_{mn} |m\rangle\langle n|, \tag{1.24}$$

$$\hat{H}_B = \hat{H}_{ph} = \sum_i \hbar \omega_i \left(\hat{a}_i^\dagger \hat{a}_i + \frac{1}{2} \right), \tag{1.25}$$

$$\hat{H}_{SB} = \hat{H}_{el-ph} - \hat{H}_{reorg} = \sum_n \hat{A}_n \otimes \hat{B}_n. \tag{1.26}$$

Note that the first term of the RHS of Eq. 1.18 vanishes with this decomposition. Thus, evaluating the bath correlation would be enough to obtain the equation of

motion for the reduced density matrix of the system.

$$\begin{aligned}
c_{mn}(t) &= \frac{1}{\hbar} \left\langle \tilde{B}_m^\dagger(t) \tilde{B}_n(0) \right\rangle \\
&= \sum_{i,j} \frac{\sqrt{m_i m_j \omega_i^3 \omega_j^3}}{2} d_{mi} d_{nj} \\
&\quad \times \text{Tr}_B \frac{e^{\beta \hat{H}_B}}{Z(\beta)} e^{i\omega_i t (\hat{a}_i^\dagger \hat{a}_i + \frac{1}{2})} (\hat{a}_i^\dagger + \hat{a}_i) e^{-i\omega_i t (\hat{a}_i^\dagger \hat{a}_i + \frac{1}{2})} (\hat{a}_j^\dagger + \hat{a}_j). \tag{1.27}
\end{aligned}$$

From the commutation relation $[\hat{a}_i^\dagger, \hat{a}_i] = 1$,

$$\hat{a}_i^\dagger e^{-i\omega_i t (\hat{a}_i^\dagger \hat{a}_i + \frac{1}{2})} = e^{-i\omega_i t (\hat{a}_i^\dagger \hat{a}_i - \frac{1}{2})} \hat{a}_i^\dagger, \tag{1.28}$$

$$\hat{a}_i e^{-i\omega_i t (\hat{a}_i^\dagger \hat{a}_i + \frac{1}{2})} = e^{-i\omega_i t (\hat{a}_i^\dagger \hat{a}_i + \frac{3}{2})} \hat{a}_i \tag{1.29}$$

By plugging in Eq. 1.28 and 1.29 to Eq. 1.27, we obtain

$$\begin{aligned}
c_{mn}(t) &= \sum_{i,j} \frac{\sqrt{m_i m_j \omega_i^3 \omega_j^3}}{2} d_{mi} d_{nj} \text{Tr}_B \frac{e^{-\beta \hat{H}_B}}{Z(\beta)} \left(e^{i\omega_i t \hat{a}_i^\dagger} + e^{-i\omega_i t \hat{a}_i} \right) (\hat{a}_j^\dagger + \hat{a}_j) \\
&= \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} \text{Tr}_B \left\{ \frac{e^{-\beta \hat{H}_B}}{Z(\beta)} e^{i\omega_i t \hat{a}_i^\dagger \hat{a}_i} + \frac{e^{-\beta \hat{H}_B}}{Z(\beta)} e^{-i\omega_i t \hat{a}_i \hat{a}_i^\dagger} \right\} \\
&= \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} \text{Tr}_B \left\{ \frac{e^{-\beta \hat{H}_B}}{Z(\beta)} e^{i\omega_i t \hat{a}_i^\dagger \hat{a}_i} + \frac{e^{-\beta \hat{H}_B}}{Z(\beta)} e^{-i\omega_i t (\hat{a}_i^\dagger \hat{a}_i + 1)} \right\} \\
&= \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} [n(\omega_i; \beta) e^{i\omega_i t} + \{n(\omega_i; \beta) + 1\} e^{-i\omega_i t}], \\
c_{mn}^*(t) &= \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} [n(\omega_i; \beta) e^{-i\omega_i t} + \{n(\omega_i; \beta) + 1\} e^{i\omega_i t}], \tag{1.30}
\end{aligned}$$

where $Z(\beta) = \text{Tr}_B e^{-\beta \hat{H}_B}$ is the partition function of the bath, and $n(\omega_i; \beta) = \frac{1}{e^{\beta \hbar \omega} - 1}$ is the Bose-Einstein distribution function at the inverse temperature β . Plugging in

to Eq. 1.13 and 1.14,

$$\begin{aligned}
 S_{kl}(t) &= \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} \{2n(\omega_i; \beta) + 1\} (e^{-i\omega_i t} + e^{i\omega_i t}) \\
 &= \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} \{2n(\omega_i; \beta) + 1\} \{2 \cos(\omega_i t)\} \\
 &= 2 \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} \coth\left(\frac{\beta \hbar \omega_i}{2}\right) \cos(\omega_i t), \tag{1.31}
 \end{aligned}$$

$$\begin{aligned}
 \chi_{kl}(t) &= i \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} (e^{-i\omega_i t} - e^{i\omega_i t}) \\
 &= i \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} \{-2i \sin(\omega_i t)\} \\
 &= 2 \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} \sin(\omega_i t). \tag{1.32}
 \end{aligned}$$

For convenience, we will rewrite the Eq. 1.31 and 1.32 by defining the spectral density of the bath associated with the m th and n th excitons as

$$J_{mn}(\omega) = \sum_i \frac{m_i \omega_i^3 d_{mi} d_{ni}}{2} \delta(\omega - \omega_i). \tag{1.33}$$

For any macroscopic bath with many degrees of freedom, its spectral density is essentially a continuous function. The $S_{kl}(t)$ and $\chi_{kl}(t)$ can now be expressed in terms of the spectral density as integral equations with respect to ω . Because ω_i 's are positive definite, the integration can be done only in the positive region.

$$S_{kl}(t) = 2 \int_0^\infty d\omega J_{mn}(\omega) \coth\left(\frac{\beta \hbar \omega}{2}\right) \cos(\omega t), \tag{1.34}$$

$$\chi_{kl}(t) = 2 \int_0^\infty d\omega J_{mn}(\omega) \sin(\omega t). \tag{1.35}$$

The Markovian master equation for a system coupled to a harmonic oscillator bath with linear coupling, like the displaced oscillator model, can be completely specified by spectral densities. One popular phenomenological model for the spectral density is

an Ohmic spectral density with a Lorentz-Drude cutoff function:

$$J_{mn}(\omega) = \frac{2\lambda_{mn}}{\pi} \omega \frac{\Omega_{mn}}{\Omega_{mn}^2 + \omega^2}, \quad (1.36)$$

where λ_{mn} is the reorganization energy and Ω_{mn} is a high frequency cutoff constant.

With this choice of spectral density, analytic expressions for the noise and dissipation kernels can be obtained:

$$S_{mn}(t) = \frac{4\lambda_{mn}\Omega_{mn}}{\beta\hbar} \left(\frac{e^{-\Omega_{mn}t}}{\Omega_{mn}} + \sum_{k=1}^{\infty} \frac{2\Omega_{mn}e^{-\Omega_{mn}t} - \nu_k e^{-\nu_k t}}{\Omega_{mn}^2 - \nu_k^2} \right), \quad (1.37)$$

$$\chi_{mn}(t) = 2\lambda_{mn}\Omega_{mn}e^{-\Omega_{mn}t}, \quad (1.38)$$

where $\nu_k = \frac{2\pi k}{\beta\hbar}$ are Matsubara frequencies. There exist alternative expressions known to converge faster than the Matsubara series and they are often favored in actual implementations of the formalism[35–37]. For simplicity, only the high temperatures approximation of $S_{mn}(t) \approx \frac{4\lambda_{mn}}{\beta\hbar} e^{-\Omega_{mn}t}$ will be considered.

Under the assumptions of the Redfield equation, $\tilde{A}_k(t-s)$ can be approximated as a Taylor expansion up to the first order in s :

$$\tilde{A}_k(t-s) = \tilde{A}_k(t) - s \frac{d}{dt} \tilde{A}_k(t) = \tilde{A}_k(t) + \frac{s}{i\hbar} \left[\hat{H}_S, \tilde{A}_k(t) \right]. \quad (1.39)$$

Plugging in all above to Eq. 1.18 and integrating gives

$$\begin{aligned}
\frac{d}{dt}\tilde{\rho}_S(t) &\approx -\frac{1}{\hbar} \sum_{k,l} \int_0^\infty ds \frac{2\lambda_{kl}}{\beta\hbar} e^{-\Omega_{mn}s} \left[\tilde{A}_k(t), \left[\tilde{A}_l(t), \tilde{\rho}_S(t) \right] \right] \\
&\quad - \frac{1}{\hbar} \sum_{k,l} \int_0^\infty ds \frac{2\lambda_{kl}}{i\beta\hbar^2} s e^{-\Omega_{mn}s} \left[\tilde{A}_k(t), \left[\left[\hat{H}_S, \tilde{A}_l(t) \right], \tilde{\rho}_S(t) \right] \right] \\
&\quad + \frac{1}{\hbar} \sum_{k,l} \int_0^\infty ds \frac{i\lambda_{mn}\Omega_{mn}}{\hbar} e^{-\Omega_{mn}s} \left[\tilde{A}_k(t), \left\{ \tilde{A}_l(t), \tilde{\rho}_S(t) \right\} \right] \\
&\quad + \frac{1}{\hbar} \sum_{k,l} \int_0^\infty ds \frac{\lambda_{mn}\Omega_{mn}}{\hbar^2} s e^{-\Omega_{mn}s} \left[\tilde{A}_k(t), \left\{ \left[\hat{H}_S, \tilde{A}_l(t) \right], \tilde{\rho}_S(t) \right\} \right] \\
&= -\frac{1}{\hbar} \sum_{k,l} \frac{2\lambda_{kl}}{\beta\hbar\Omega_{mn}} \left[\tilde{A}_k(t), \left[\tilde{A}_l(t), \tilde{\rho}_S(t) \right] \right] \\
&\quad + \frac{1}{\hbar} \sum_{k,l} \frac{2i\lambda_{kl}}{\beta\hbar^2\Omega_{mn}^2} \left[\tilde{A}_k(t), \left[\left[\hat{H}_S, \tilde{A}_l(t) \right], \tilde{\rho}_S(t) \right] \right] \\
&\quad + \frac{1}{\hbar} \sum_{k,l} \frac{i\lambda_{mn}}{\hbar} \left[\tilde{A}_k(t), \left\{ \tilde{A}_l(t), \tilde{\rho}_S(t) \right\} \right] \\
&\quad + \frac{1}{\hbar} \sum_{k,l} \frac{\lambda_{mn}}{\hbar^2\Omega_{mn}} \left[\tilde{A}_k(t), \left\{ \left[\hat{H}_S, \tilde{A}_l(t) \right], \tilde{\rho}_S(t) \right\} \right]. \tag{1.40}
\end{aligned}$$

If expressed in the Schrödinger picture, the generator of the quantum master equation does not depend on time:

$$\begin{aligned}
\frac{d}{dt}\rho_S(t) &= \frac{1}{i\hbar} \left[\hat{H}_S, \rho(t) \right] - \frac{1}{\hbar} \sum_{k,l} \frac{2\lambda_{kl}}{\beta\hbar\Omega_{mn}} \left[\hat{A}_k, \left[\hat{A}_l, \rho_S(t) \right] \right] \\
&\quad + \frac{1}{\hbar} \sum_{k,l} \frac{2i\lambda_{kl}}{\beta\hbar^2\Omega_{mn}^2} \left[\hat{A}_k, \left[\left[\hat{H}_S, \hat{A}_l \right], \rho_S(t) \right] \right] \\
&\quad + \frac{1}{\hbar} \sum_{k,l} \frac{i\lambda_{mn}}{\hbar} \left[\hat{A}_k, \left\{ \hat{A}_l, \rho_S(t) \right\} \right] \\
&\quad + \frac{1}{\hbar} \sum_{k,l} \frac{\lambda_{mn}}{\hbar^2\Omega_{mn}} \left[\hat{A}_k, \left\{ \left[\hat{H}_S, \hat{A}_l \right], \rho_S(t) \right\} \right]. \tag{1.41}
\end{aligned}$$

1.1.3 Reduced Hierarchical Equation of Motion

The Born approximation employed in the Redfield equation leads to the perturbation expansion to the second order. Due to this limitation, the Redfield equation is not applicable when the system-bath interaction is of comparable scale to the site-site coupling [28]. Also, non-Markovian effects cannot be captured because the bath correlation time is assumed to be short. For a harmonic oscillator bath with Ohmic spectral density and Lorentz-Drude cutoff, a non-perturbative and non-Markovian equation of motion can be derived by exploiting the following facts: (1) The noise and dissipation kernels are linear combinations of exponential functions of time and (2) the system-bath interaction Hamiltonian is linear in the position operators of bath oscillators. We now derive this non-Markovian master equation.

Starting from the previous decomposition for the Hamiltonian of Frenkel excitons in Eq. 1.24-1.26 and substituting to Eq. 1.4, we obtain

$$\begin{aligned}\frac{d}{dt}\tilde{\rho}(t) &= \frac{1}{i\hbar} [\tilde{H}_{SB}, \tilde{\rho}(t)] \\ &= \frac{1}{i\hbar} \sum_k \left\{ \tilde{\mathcal{A}}_k(t) \otimes \tilde{\mathcal{B}}_k(t) \right\} \tilde{\rho}(t),\end{aligned}\tag{1.42}$$

where

$$\begin{aligned}\tilde{\mathcal{A}}_k(t)\sigma &= [\tilde{A}_k(t), \sigma], \tilde{\mathcal{A}}_k^\dagger(t)\sigma = [\tilde{A}_k^\dagger(t), \sigma], \\ \tilde{\mathcal{B}}_k(t)\sigma &= [\tilde{B}_k(t), \sigma], \tilde{\mathcal{B}}_k^\dagger(t)\sigma = [\tilde{B}_k^\dagger(t), \sigma].\end{aligned}\tag{1.43}$$

The formal solution of the Eq.1.42 is,

$$\tilde{\rho}(t) = T_{\leftarrow} \exp \left(\frac{1}{i\hbar} \int_0^t ds \sum_k \tilde{\mathcal{A}}_k(s) \otimes \tilde{\mathcal{B}}_k(s) \right) \tilde{\rho}(0),\tag{1.44}$$

where T_{\leftarrow} is the chronological time ordering operator. By assuming that the initial state is factorizable, $\tilde{\rho}(0) = \tilde{\rho}_S(0) \otimes \tilde{\rho}_B$, where the bath is in equilibrium like Eq. 1.10, the reduced density matrix of the system can be written as,

$$\tilde{\rho}_S(t) = T_{\leftarrow} \left\langle \exp \left(\frac{1}{i\hbar} \int_0^t ds \sum_k \tilde{\mathcal{A}}_k(s) \otimes \tilde{\mathcal{B}}_k(s) \right) \right\rangle \tilde{\rho}(0). \quad (1.45)$$

The bath operators $\tilde{B}_k(t)$ are linear in the bath position operators. By applying Kubo's generalized cumulant expansion [38], the equation above can be expressed as,

$$\begin{aligned} \tilde{\rho}_S(t) &= T_{\leftarrow} \left\langle \exp \left(\frac{1}{i\hbar} \int_0^t ds \sum_k \tilde{\mathcal{A}}_k(s) \otimes \tilde{\mathcal{B}}_k(s) \right) \right\rangle \tilde{\rho}(0) \\ &= T_{\leftarrow} \exp \left(-\frac{1}{\hbar^2} \sum_k \int_0^t dt_1 \int_0^{t_1} ds_1 \left\langle \left\{ \tilde{\mathcal{A}}_k^\dagger(t_1) \otimes \tilde{\mathcal{B}}_k^\dagger(t_1) \right\} \left\{ \tilde{\mathcal{A}}_k(s_1) \otimes \tilde{\mathcal{B}}_k(s_1) \right\} \right\rangle \right) \tilde{\rho}(0). \end{aligned} \quad (1.46)$$

The integrand of Eq. 1.46 can be explicitly evaluated by adapting an Ohmic spectral density in Eq. 1.36:

$$-\frac{1}{\hbar^2} \left\langle \left\{ \tilde{\mathcal{A}}_k^\dagger(t_1) \otimes \tilde{\mathcal{B}}_k^\dagger(t_1) \right\} \left\{ \tilde{\mathcal{A}}_k(s_1) \otimes \tilde{\mathcal{B}}_k(s_1) \right\} \right\rangle = \tilde{\mathcal{F}}_k(t_1) e^{-\Omega_k(t_1-s_1)} \tilde{\mathcal{T}}_k(s_1), \quad (1.47)$$

where $\tilde{\mathcal{F}}_k(t)$ and $\tilde{\mathcal{T}}_k(t)$ are superoperators defined as,

$$\begin{aligned} \tilde{\mathcal{F}}_k(t) &= i \left[\tilde{A}_k^\dagger(t), \sigma \right], \\ \tilde{\mathcal{T}}_k(t) &= \frac{2i\lambda_k}{\beta\hbar^2} \left[\tilde{A}_k(t), \sigma \right] + \frac{\lambda_k\Omega_k}{\hbar} \left\{ \tilde{A}_k(t), \sigma \right\}. \end{aligned} \quad (1.48)$$

For algebraic convenience, we assumed that the bath operators coupled to different sites are uncorrelated. Rewriting Eq. 1.46 using these superoperators, the reduced density matrix can be obtained as,

$$\begin{aligned} \tilde{\rho}_S(t) &= T_{\leftarrow} \exp \left(\sum_k \int_0^t dt_1 \tilde{\mathcal{F}}_k(t_1) \int_0^{t_1} ds_1 e^{-\Omega_k(t_1-s_1)} \tilde{\mathcal{T}}_k(s_1) \right) \tilde{\rho}_S(0) \\ &= T_{\leftarrow} \left\{ \prod_k \exp \left(\int_0^t dt_1 \tilde{\mathcal{F}}_k(t_1) \int_0^{t_1} ds_1 e^{-\Omega_k(t_1-s_1)} \tilde{\mathcal{T}}_k(s_1) \right) \right\} \tilde{\rho}_S(0), \end{aligned} \quad (1.49)$$

Differentiating both sides of Eq. 1.49,

$$\begin{aligned}
\frac{d}{dt}\tilde{\rho}_S(t) &= T_{\leftarrow} \sum_l \tilde{\mathcal{F}}_l(t) \left(\int_0^t ds e^{-\Omega_k(t_1-s_1)} \tilde{\mathcal{T}}_l(s) \right) \\
&\quad \times \left\{ \prod_k \exp \left(\int_0^t dt_1 \tilde{\mathcal{F}}_k(t_1) \int_0^{t_1} ds_1 e^{-\Omega_k(t_1-s_1)} \tilde{\mathcal{T}}_k(s_1) \right) \right\} \tilde{\rho}_S(0) \\
&= \sum_l \tilde{\mathcal{F}}_l(t) T_{\leftarrow} \left(\int_0^t ds e^{-\Omega_k(t_1-s_1)} \tilde{\mathcal{T}}_l(s) \right) \\
&\quad \times \left\{ \prod_k \exp \left(\int_0^t dt_1 \tilde{\mathcal{F}}_k(t_1) \int_0^{t_1} ds_1 e^{-\Omega_k(t_1-s_1)} \tilde{\mathcal{T}}_k(s_1) \right) \right\} \tilde{\rho}_S(0) \\
&= \sum_l \tilde{\mathcal{F}}_l(t) \tilde{\sigma}_{\{\dots, n_{l-1}=0, n_l=1, n_{l+1}=0, \dots\}}(t), \tag{1.50}
\end{aligned}$$

where auxiliary matrices $\tilde{\sigma}_{\{n_1, \dots, n_N\}}$ are defined as,

$$\begin{aligned}
\tilde{\sigma}_{\{n_1, \dots, n_N\}}(t) &= T_{\leftarrow} \prod_k \left(\int_0^t ds e^{-\Omega_k(t-s)} \tilde{\mathcal{T}}_k(s) \right)^{n_k} \\
&\quad \times \exp \left(\int_0^t dt_1 \tilde{\mathcal{F}}_k(t_1) \int_0^{t_1} ds_1 e^{-\Omega_k(t_1-s_1)} \tilde{\mathcal{T}}_k(s_1) \right) \tilde{\rho}_S(0), \tag{1.51}
\end{aligned}$$

and it becomes zero if any element in $\{n_k\}$ is negative. Note that $\tilde{\rho}_S(t) = \tilde{\sigma}_{\{0, \dots, 0\}}(t)$ and $\tilde{\sigma}_{\{n_1, \dots, n_N\}}(0) = 0$. The equation of motion of $\tilde{\sigma}_{\{n_1, \dots, n_N\}}(t)$ is,

$$\begin{aligned}
\frac{d}{dt} \tilde{\sigma}_{\{n_1, \dots, n_N\}}(t) &= - \sum_l \Omega_l \tilde{\sigma}_{\{n_1, \dots, n_N\}} \\
&\quad + \sum_l n_l \tilde{\mathcal{T}}_l(t) \tilde{\sigma}_{\{n_1, \dots, n_{l-1}, \dots, n_N\}} + \sum_l \tilde{\mathcal{F}}_l(t) \tilde{\sigma}_{\{n_1, \dots, n_{l+1}, \dots, n_N\}}, \tag{1.52}
\end{aligned}$$

in the interaction picture. Moving to the Schrödinger picture, we obtain a set of hierarchical equations of motion.

$$\begin{aligned}
\frac{d}{dt} \sigma_{\{n_1, \dots, n_N\}}(t) &= \frac{1}{i\hbar} \mathcal{L}_S \sigma_{\{n_1, \dots, n_N\}} - \sum_l \Omega_l \sigma_{\{n_1, \dots, n_N\}} \\
&\quad + \sum_l n_l \mathcal{T}_l(t) \sigma_{\{n_1, \dots, n_{l-1}, \dots, n_N\}} + \sum_l \mathcal{F}_l(t) \sigma_{\{n_1, \dots, n_{l+1}, \dots, n_N\}}. \tag{1.53}
\end{aligned}$$

1.1.4 Haken-Strobl-Reineker Model

First developed by Haken, Strobl and Reineker [39–41], this phenomenological stochastic model describes the coherent and incoherent dynamics of Frenkel excitons at the same time. Instead of decompose the total complex to the system and bath, Haken-Strobl-Reineker (HSR) model mainly focuses on the system Hamiltonian, and the effect of the bath environment is included as time-dependent stochastic terms:

$$\begin{aligned}\hat{H}_{sys} &= \sum_n \varepsilon_n |n\rangle\langle n| + \sum_{m \neq n} E_{mn} |m\rangle\langle n|, \\ \hat{H}_{env}(t) &= \sum_{m,n} h_{mn}(t) |m\rangle\langle n|, \\ \hat{H}_{total}(t) &= \hat{H}_{sys} + \hat{H}_{env}(t).\end{aligned}\tag{1.54}$$

Realized stochastic density matrix $\check{\rho}_S(t)$ can be defined per realization of the stochastic Hamiltonian according to the usual form of the quantum Liouville equation:

$$\begin{aligned}\frac{d\check{\rho}_S(t)}{dt} &= \frac{1}{i\hbar} \left[\hat{H}_{total}, \check{\rho}_S(t) \right] \\ &= \frac{1}{i\hbar} \left[\hat{H}_{sys}, \check{\rho}_S(t) \right] + \frac{1}{i\hbar} \left[\hat{H}_{env}(t), \check{\rho}_S(t) \right].\end{aligned}\tag{1.55}$$

Then the reduced density matrix of the system can be obtained as the expectation of the density matrices over the realized trajectory:

$$\rho_S(t) = \mathbb{E}(\check{\rho}_S(t)).\tag{1.56}$$

Note that $\check{\rho}_S(t)$ is a stochastic process while $\rho_S(t)$ is deterministic. Equivalently, the equation of the motion for the reduced density matrix of the system can be obtained by taking expectation on both sides of Eq. 1.55:

$$\frac{d\rho_S(t)}{dt} = \frac{1}{i\hbar} \left[\hat{H}_{sys}, \rho_S(t) \right] + \frac{1}{i\hbar} \mathbb{E} \left(\left[\hat{H}_{env}(t), \check{\rho}_S(t) \right] \right).\tag{1.57}$$

$h_{mn}(t)$ are assumed to have the correlation functions given by

$$\begin{aligned}\mathbb{E} \{h_{mn}(t)h_{nm}(t')\} &= \frac{\hbar\gamma_{mn}}{\tau_c} e^{-\frac{|t-t'|}{\tau_c}}, \gamma_{mn} = \gamma_{nm}, \\ \mathbb{E} \{h_{mn}(t)h_{mn}(t')\} &= \frac{\hbar\bar{\gamma}_{mn}}{\tau_c} e^{-\frac{|t-t'|}{\tau_c}}, \bar{\gamma}_{mn} = \bar{\gamma}_{nm}^*,\end{aligned}\tag{1.58}$$

where τ_c is the correlation time. The mean values can be set to zero by absorbing any leftover term into the system Hamiltonian. In the original parametrization, these stochastic terms were assumed to have delta function correlation in time. There exist a formulation with exponential correlation function for a two-exciton system [42], but the derivation presented here is generalized to cover any number of excitons. The constant \hbar was introduced for γ_{mn} to have the unit of energy. Although an extension of the HSR model with nonzero intersite correlations exists [43], only the formulation with uncorrelated sites will be discussed in this chapter because of its simplicity and clarity.

To evaluate the second term of Eq. 1.57, we will rewrite the equation using an orthonormal basis set $\{\Omega_k\}$ spanning the density operator space with the following inner product [44]:

$$\langle \Omega_k | \Omega_l \rangle \equiv \text{Tr}(\Omega_k^\dagger \Omega_l) = \delta_{kl}.\tag{1.59}$$

With this orthonormal basis, the commutation relation between operators can be interpreted as a linear operator acting on the aforementioned vector space:

$$\begin{aligned}\mathcal{H}\Omega_l &= [\hat{H}, \Omega_l] \\ &= \sum_k \text{Tr} \left(\Omega_k^\dagger [\hat{H}, \Omega_l] \right) \Omega_k \\ &= \sum_k \text{Tr} \left(\hat{H} [\Omega_l, \Omega_k^\dagger] \right) \Omega_k.\end{aligned}\tag{1.60}$$

Then, Eq. 1.55 can be rewritten as

$$\frac{d}{dt}|\check{\rho}_S(t)\rangle = \frac{1}{i\hbar}\mathcal{H}_{sys}|\check{\rho}_S(t)\rangle + \frac{1}{i\hbar}\mathcal{H}_{env}|\check{\rho}_S(t)\rangle, \quad (1.61)$$

where $|\check{\rho}_S(t)\rangle = \sum_k \text{Tr}(\Omega_k^\dagger \check{\rho}_S) \Omega_k$, $\mathcal{H}_{sys} = \sum_{k,l} \text{Tr} \left(\hat{H}_{sys} \left[\Omega_l, \Omega_k^\dagger \right] \right) |\Omega_k\rangle \langle \Omega_l|$ and $\mathcal{H}_{env} = \sum_{k,l} \text{Tr} \left(\hat{H}_{env} \left[\Omega_l, \Omega_k^\dagger \right] \right) |\Omega_k\rangle \langle \Omega_l|$.

Because the initial state is same for all instances of the trajectories, a closed form of a realization of the density matrix can be obtained by integrating Eq. 1.61:

$$|\check{\rho}_S(t)\rangle = T_{\leftarrow} \exp \left(\frac{1}{i\hbar} \int_0^t ds \mathcal{H}_{sys} \right) T_{\leftarrow} \exp \left(\frac{1}{i\hbar} \int_0^t ds \mathcal{H}_{env}(s) \right) |\rho_S(0)\rangle, \quad (1.62)$$

where T_{\leftarrow} is the chronological time-ordering operator. Taking expectation on both sides, the vector expression for the system density matrix can be obtained:

$$|\rho_S(t)\rangle = T_{\leftarrow} \exp \left(\frac{1}{i\hbar} \int_0^t ds \mathcal{H}_{sys} \right) \mathbb{E} \left\{ T_{\leftarrow} \exp \left(\frac{1}{i\hbar} \int_0^t ds \mathcal{H}_{env}(s) \right) \right\} |\rho_S(0)\rangle. \quad (1.63)$$

Note that $\langle \Omega_k | \mathcal{H}_{env} | \Omega_l \rangle = \text{Tr} \left(\Omega_k^\dagger \left[\hat{H}_{env}, \Omega_l \right] \right)$ is also a stationary Gaussian random process with zero mean and correlation time τ_c because the term is linear in $h_{mn}(t)$'s:

$$\mathbb{E} \{ \mathcal{H}_{env}(t_2) \mathcal{H}_{env}(t_1) \} = \frac{\hbar}{\tau_c} e^{-\frac{|t_2-t_1|}{\tau_c}} \mathcal{E} \quad (1.64)$$

\mathcal{E} is a constant superoperator and has the unit of energy. The expectation of an operator appearing in Eq. 1.63 can be rewritten using only the second moments by Kubo's generalized cumulant expansion [38]:

$$\begin{aligned} \mathbb{E} \left\{ T_{\leftarrow} \exp \left(\frac{1}{i\hbar} \int_0^t ds \mathcal{H}_{env} \right) \right\} &= T_{\leftarrow} \exp \left(-\frac{1}{\hbar^2} \int_0^t dt_2 \int_0^{t_2} dt_1 \mathbb{E} \{ \mathcal{H}_{env}(t_2) \mathcal{H}_{env}(t_1) \} \right) \\ &= T_{\leftarrow} \exp \left(-\frac{1}{\hbar} \int_0^t dt_2 \int_0^{t_2} dt_1 \frac{1}{\tau_c} e^{-\frac{|t_2-t_1|}{\tau_c}} \mathcal{E} \right) \\ &= T_{\leftarrow} \exp \left\{ -\frac{1}{\hbar} \int_0^t dt_2 \left(1 - e^{-\frac{t_2}{\tau_c}} \right) \mathcal{E} \right\}. \end{aligned} \quad (1.65)$$

Plugging in this result to Eq. 1.63 gives

$$|\rho_S(t)\rangle = T_{\leftarrow} \exp \left(\frac{1}{i\hbar} \int_0^t ds \mathcal{H}_{sys} \right) T_{\leftarrow} \exp \left\{ -\frac{1}{\hbar} \int_0^t dt_2 \left(1 - e^{-\frac{t_2}{\tau_c}} \right) \mathcal{E} \right\} |\rho_S(0)\rangle. \quad (1.66)$$

By differentiating the equation above, we can come up with the generator for the vector representation of the density matrix.

$$\frac{d}{dt} |\rho_S(t)\rangle = \frac{1}{i\hbar} \left\{ \mathcal{H}_{sys} - i \left(1 - e^{-\frac{t}{\tau_c}} \right) \mathcal{E} \right\} |\rho_S(t)\rangle. \quad (1.67)$$

For the practical use, we want to obtain the equation of motion for each element of the density matrix of the system. Those equations can be obtained by choosing $\Omega_k = |k_1\rangle\langle k_2|$. k is the collective index for (k_1, k_2) in this case and $|\Omega_k\rangle$ is equivalent to $|k_1, k_2\rangle$.

$$\begin{aligned} \frac{d}{dt} \langle k_1 | \rho_S(t) | k_2 \rangle &= \langle k_1, k_2 | \frac{d}{dt} |\rho_S(t)\rangle \\ &= \frac{1}{i\hbar} \langle k_1, k_2 | \mathcal{H}_{sys} | \rho_S(t) \rangle - \frac{1}{\hbar} \left(1 - e^{-\frac{t}{\tau_c}} \right) \langle k_1, k_2 | \mathcal{E} | \rho_S(t) \rangle. \end{aligned} \quad (1.68)$$

The explicit form of the $\mathbb{E} \{ \mathcal{H}_{env}(t_1) \mathcal{H}_{env}(t_2) \}$ need to be found to evaluate the second term of Eq. 1.68.

$$\mathcal{H}_{env}(t_2) \mathcal{H}_{env}(t_1) = \sum_{k,l} \sum_j \text{Tr} \left(\hat{H}_{env}(t_2) \left[\Omega_j, \Omega_k^\dagger \right] \right) \text{Tr} \left(\hat{H}_{env}(t_1) \left[\Omega_l, \Omega_j^\dagger \right] \right) |\Omega_k\rangle\langle\Omega_l|, \quad (1.69)$$

where the prefactors of the superoperator \mathcal{H}_{env} are

$$\begin{aligned} \text{Tr} \left(\hat{H}_{env}(t) \left[\Omega_l, \Omega_k^\dagger \right] \right) &= \text{Tr} \left\{ \sum_{m,n} h_{mn}(t) (|m\rangle\langle n| l_1 \rangle \langle l_2 | k_2 \rangle \langle k_1 | - |l_1\rangle \langle l_2 | m \rangle \langle n | k_2 \rangle \langle k_1 |) \right\} \\ &= \text{Tr} \left\{ \sum_m \delta_{k_2 l_2} h_{m l_1}(t) |m\rangle \langle k_1| - h_{l_2 k_2}(t) |l_1\rangle \langle k_1| \right\} \\ &= \delta_{k_2 l_2} h_{k_1 l_1}(t) - \delta_{k_1 l_1} h_{l_2 k_2}(t). \end{aligned} \quad (1.70)$$

By plugging in Eq. 1.70 to Eq. 1.69,

$$\begin{aligned}
& \sum_j \text{Tr} \left(\hat{H}_{env}(t_2) \left[\Omega_j, \Omega_k^\dagger \right] \right) \text{Tr} \left(\hat{H}_{env}(t_1) \left[\Omega_l, \Omega_j^\dagger \right] \right) \\
&= \sum_{j_1, j_2} \{ \delta_{k_2 j_2} h_{k_1 j_1}(t_2) - \delta_{k_1 j_1} h_{j_2 k_2}(t_1) \} \{ \delta_{j_2 l_2} h_{j_1 l_1}(t_2) - \delta_{j_1 l_1} h_{l_2 j_2}(t_1) \} \\
&= \sum_{j_1} \delta_{k_2 l_2} h_{k_1 j_1}(t_2) h_{j_1 l_1}(t_1) - h_{k_1 l_1}(t_2) h_{l_2 k_2}(t_1) \\
&\quad - h_{l_2 k_2}(t_2) h_{k_1 l_1}(t_1) + \sum_{j_2} \delta_{k_1 l_1} h_{j_2 k_2}(t_2) h_{l_2 j_2}(t_1) \\
&= -h_{k_1 l_1}(t_2) h_{l_2 k_2}(t_1) - h_{l_2 k_2}(t_2) h_{k_1 l_1}(t_1) \\
&\quad + \sum_j \delta_{k_2 l_2} h_{k_1 j}(t_2) h_{j l_1}(t_1) + \sum_j \delta_{k_1 l_1} h_{j k_2}(t_2) h_{l_2 j}(t_1). \tag{1.71}
\end{aligned}$$

To calculate the second term on the right hand side of Eq. 1.68, the following term should be evaluated first:

$$\begin{aligned}
& \langle k_1, k_2 | \mathbb{E} \{ \mathcal{H}_{env}(t_2) \mathcal{H}_{env}(t_1) \} | \rho_S(t) \rangle \\
&= - \sum_{l_1, l_2} \mathbb{E} \{ h_{k_1 l_1}(t_2) h_{l_2 k_2}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle - \sum_{l_1, l_2} \mathbb{E} \{ h_{l_2 k_2}(t_2) h_{k_1 l_1}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle \\
&\quad + \sum_{l_1, l_2} \sum_j \mathbb{E} \{ \delta_{k_2 l_2} h_{k_1 j}(t_2) h_{j l_1}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle \\
&\quad + \sum_{l_1, l_2} \sum_j \mathbb{E} \{ \delta_{k_1 l_1} h_{j k_2}(t_2) h_{l_2 j}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle. \tag{1.72}
\end{aligned}$$

For the diagonal elements, or for $k_1 = k_2 = k$,

$$\begin{aligned}
& \langle k, k | \mathbb{E} \{ \mathcal{H}_{env}(t_2) \mathcal{H}_{env}(t_1) \} | \rho_S(t) \rangle \\
&= - \sum_{l_1, l_2} \mathbb{E} \{ h_{kl_1}(t_2) h_{l_2k}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle - \sum_{l_1, l_2} \mathbb{E} \{ h_{l_2k}(t_2) h_{kl_1}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle \\
&+ \sum_{l_1, l_2} \sum_j \mathbb{E} \{ \delta_{kl_2} h_{kj}(t_2) h_{jl_1}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle \\
&+ \sum_{l_1, l_2} \sum_j \mathbb{E} \{ \delta_{kl_1} h_{jk}(t_2) h_{l_2j}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle \\
&= - \sum_j \frac{\hbar(\gamma_{kj} + \gamma_{jk})}{\tau_c} e^{-\frac{|t_2-t_1|}{\tau_c}} \langle j | \rho_S(t) | j \rangle + \sum_j \frac{\hbar(\gamma_{kj} + \gamma_{jk})}{\tau_c} e^{-\frac{|t_2-t_1|}{\tau_c}} \langle k | \rho_S(t) | k \rangle.
\end{aligned} \tag{1.73}$$

Therefore,

$$\langle k, k | \mathcal{E} | \rho_S(t) \rangle = - \sum_j (\gamma_{kj} + \gamma_{jk}) \langle j | \rho_S(t) | j \rangle + \sum_j (\gamma_{kj} + \gamma_{jk}) \langle k | \rho_S(t) | k \rangle. \tag{1.74}$$

Plugging in Eq. 1.74 to Eq. 1.68,

$$\begin{aligned}
\frac{d}{dt} \langle k | \rho_S(t) | k \rangle &= \frac{1}{i\hbar} \langle k | \left[\hat{H}_{sys}, \rho_S(t) \right] | k \rangle + \left(1 - e^{-\frac{t}{\tau_c}} \right) \sum_j \frac{\gamma_{kj} + \gamma_{jk}}{\hbar} \langle j | \rho_S(t) | j \rangle \\
&- \left(1 - e^{-\frac{t}{\tau_c}} \right) \sum_j \frac{\gamma_{kj} + \gamma_{jk}}{\hbar} \langle k | \rho_S(t) | k \rangle.
\end{aligned} \tag{1.75}$$

Similarly, for the off-diagonal elements, $k_1 \neq k_2$,

$$\begin{aligned}
& \langle k_1, k_2 | \mathbb{E} \{ \mathcal{H}_{env}(t_2) \mathcal{H}_{env}(t_1) \} | \rho_S(t) \rangle \\
&= - \sum_{l_1, l_2} \mathbb{E} \{ h_{k_1 l_1}(t_2) h_{l_2 k_2}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle - \sum_{l_1, l_2} \mathbb{E} \{ h_{l_2 k_2}(t_2) h_{k_1 l_1}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle \\
&+ \sum_{l_1, l_2} \sum_j \mathbb{E} \{ \delta_{k_2 l_2} h_{k_1 j}(t_2) h_{j l_1}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle \\
&+ \sum_{l_1, l_2} \sum_j \mathbb{E} \{ \delta_{k_1 l_1} h_{j k_2}(t_2) h_{l_2 j}(t_1) \} \langle l_1 | \rho_S(t) | l_2 \rangle \\
&= - \frac{2\hbar \bar{\gamma}_{k_1 k_2}}{\tau_c} e^{-\frac{|t_2 - t_1|}{\tau_c}} \langle k_2 | \rho_S(t) | k_1 \rangle \\
&+ \sum_j \frac{\hbar \gamma_{k_1 j}}{\tau_c} e^{-\frac{|t_2 - t_1|}{\tau_c}} \langle k_1 | \rho_S(t) | k_2 \rangle + \sum_j \frac{\hbar \gamma_{j k_2}}{\tau_c} e^{-\frac{|t_2 - t_1|}{\tau_c}} \langle k_1 | \rho_S(t) | k_2 \rangle. \tag{1.76}
\end{aligned}$$

Therefore,

$$\langle k_1, k_2 | \mathcal{E} | \rho_S(t) \rangle = -2\bar{\gamma}_{k_1 k_2} \langle k_2 | \rho_S(t) | k_1 \rangle + \sum_j (\gamma_{k_1 j} + \gamma_{j k_2}) \langle k_1 | \rho_S(t) | k_2 \rangle. \tag{1.77}$$

Plugging in Eq. 1.77 to Eq. 1.68,

$$\begin{aligned}
\frac{d}{dt} \langle k_1 | \rho_S(t) | k_2 \rangle &= \frac{1}{i\hbar} \langle k_1 | [\hat{H}_{sys}, \rho_S(t)] | k_2 \rangle + \left(1 - e^{-\frac{t}{\tau_c}} \right) \frac{2\bar{\gamma}_{k_1 k_2}}{\hbar} \langle k_2 | \rho_S(t) | k_1 \rangle \\
&- \left(1 - e^{-\frac{t}{\tau_c}} \right) \sum_j \frac{\gamma_{k_1 j} + \gamma_{j k_2}}{\hbar} \langle k_1 | \rho_S(t) | k_2 \rangle. \tag{1.78}
\end{aligned}$$

Note that in the completely memoryless bath limit of $\tau_c \rightarrow 0$, the original Haken-Strobl-Reineker fomulation is restored.

Part I

Quantum Coherences in Light Harvesting Systems

Chapter 2

Characterization and quantification of the role of coherence in ultrafast quantum biological experiments

2.1 Introduction

The initial step in photosynthesis is highly efficient excitonic transport of the energy captured from photons to a reaction center [45]. In most plants and photosynthetic organisms this process occurs in light-harvesting complexes which are interacting chlorophyll molecules embedded in a solvent and a protein environment [46]. Several recent experiments show that excitonic coherence can persist for several hundreds of femtoseconds even at physiological temperature [23–25, 47]. These experiments suggest the hypothesis that quantum coherence is biologically relevant for photosynthesis. The results have motivated a sizeable amount of recent theoretical

work regarding the reasons for the long-lived coherences and their role to the function.

The focus of many studies is on the theoretical models employed. In this context, it is essential to be as realistic as possible and employ the least amount of approximations. Most of the currently-employed methods involve a master equation for the reduced excitonic density operator where the vibrational degrees of freedom (phonons) of the protein and solvent are averaged out. Amongst these simple methods are the Haken-Strobl model and Redfield theory as employed in Refs. 27, 48 and 49 respectively. To interpolate between the usual weak and strong exciton-phonon coupling limits, Ishizaki and Fleming developed a hierarchical equation of motion (HEOM) theory which takes into account non-equilibrium molecular reorganization effects [29]. Jang et al. perform a second order time-convolutionless expansion after a small polaron transformation to include strong coupling effects [50]. Another set of studies focuses on the role of quantum coherence and the phonon environment in terms of transport efficiency or entanglement. It was shown that the transport efficiency is enhanced by the interaction or interplay of the quantum evolution with the phononic environment [27, 48, 49, 51]. Entanglement between molecules is found to persist for long times [52–54].

The ongoing effort can be summarized with two equally important questions: What are the microscopic reasons for the persistence of quantum coherence and what is the relevance of the quantum effect to the biological functionality of the organism under study? In this work, we summarize the recent efforts from our group to approach the problem from several angles. Firstly, we investigate the role of coherences in the exciton transfer process of the Fenna-Matthews-Olson (FMO) complex.

We quantify the amount and the contribution of coherence to the efficient energy transfer process. Secondly, we present our quantum mechanics/molecular mechanics (QM/MM) approach to obtain information about the system at the atomistic level, such as detailed bath dynamics and spectral densities. Finally, we propose a spectroscopic tool that allows for obtaining directly the information of the quantum process via our recent theoretical proposal for the quantum process tomography technique to the ultrafast regime.

2.2 The Role of Quantum Coherence

In this section, we discuss the question about the relevance of quantum effects to the biological function. A negative answer to this question would mean that a particular effect, while being quantum, is not leading to any improvement in the functionality of a biological system, and therefore would be a byproduct of the spatial and temporal scales and physical properties of the problem. For example, in energy transfer (ET) quantum coherence could arise from the closely packed arrangement of the chromophores in a protein scaffold but it could, in principle, represent a byproduct of that arrangement and not a relevant feature. Another example, it may be true that the human eye can detect a single photon, but it is not clear if this quantum effect is relevant to the biological function, which usually operates at much larger photon fluxes. If, on the other hand, the above yes-no question of the relevance is answered positively for a particular effect in a biological system, it would present a major step towards establishing the relevance or importance of a quantum biological phenomenon. A natural follow-up questions is: How important *quantitatively* is a

particular quantum effect?

Both of these questions should preferably be studied by experimental means. An experiment would have to be designed in a way that tests for the biological relevance of quantum coherence. Possible experiments could involve quantum measurements on mutated samples. In the FMO complex that acts as a molecular ET wire the efficiency of the transport event is most likely a good quantifier for biological function. One would need a way to experimentally quantify this efficiency and extract the relevance of quantum coherence to the efficiency. This can be hard in practice. Yet, as we will discuss in this work, quantum process tomography is able to obtain detailed information about quantum coherence and the phonon environment and might thus lead to progress in this area.

In the case when experimental access to an observable that involves the biological relevance is hard or impossible, a theoretical treatment can provide insight. It is illustrative to analyze a model of the particular biological process in terms of a quantifier for the success of the process. An example is the aforementioned efficiency of energy transport. In bird vision, the quantum yield of a chemical reaction is a relevant measure [55]. Once a detailed model and a success criterion is established, one needs to quantify the contribution of quantum coherence to the success criterion. For this step, one can proceed in two distinct pathways. The first pathway is a comparison to a classical reference point; the success criterion is computed for the actual system/model and a classical reference model that does not include quantum correlations. The difference of these two values is attributed to quantum mechanics and can be considered the quantum mechanical contribution to the success of the process. For

example, the energy transfer dynamics of a sophisticated quantum mechanical model such as [29] could be compared to a semi-classical Förster treatment that leads to a hopping description. In general, this comparison strategy has the drawback that one has to invoke a classical, and in some cases very artificial, model.

Our work has been mainly concentrated on a second theoretical pathway in answering the relevance question, which overcomes this issue. It is based on just the quantum mechanical model and the success quantifier. No other, for example classical, model is invoked. The actual model will contain dynamical processes that are quantum coherent and others that are incoherent. The non-trivial task is to deconstruct how the various processes contribute to the performance criterion. This can be done by decomposing the performance criterion into a sum of contributions, each associated with a particular process. The terms in this sum related to quantum mechanical processes will then give a theoretical answer to the overall relevance of the particular process and will quantify this relevance. This line of thought was developed and discussed in Ref. 26 for energy transfer in the FMO complex and provided insight into both questions "Is a quantum effect relevant?" and "If yes, how much?", at least from a theoretical standpoint within the approximations of the model under consideration. In this section, we extend this idea to include the effect of the initial conditions and compare the results to a total integrated coherence, or concurrence, measure. We utilize secular Redfield theory and the hierarchy equation of motion approach.

The Hamiltonian describing a single exciton is given by:

$$H_e = \sum_m (\epsilon_m + \lambda) |m\rangle\langle m| + \sum_{m < n} J_{mn} (|m\rangle\langle n| + |n\rangle\langle m|). \quad (2.1)$$

where the site energies ϵ_m , and couplings J_{mn} are usually obtained from detailed quantum chemistry studies and/or fitting of experimental spectra. The reorganization energy λ , which we assume to be the same for each site, is the energy difference of the non-equilibrium phonon state after Franck-Condon excitation and the excited-state equilibrium phonon state. The set of states $|m\rangle$ is called the site basis and the set of states $|\alpha\rangle$ with $H_e|\alpha\rangle = E_\alpha|\alpha\rangle$ is called the exciton basis. We now briefly introduce the secular Redfield master equation in the weak exciton-phonon (or system-bath) coupling limit and the non-perturbative hierarchy equation of motion approach. In both approaches, the dynamics of a single exciton is governed by a master equation, which is schematically given by:

$$\frac{\partial}{\partial t}\rho(t) = \mathcal{M}\rho(t) = (\mathcal{M}_H + \mathcal{M}_{\text{decoherence}} + \mathcal{M}_{\text{trap}} + \mathcal{M}_{\text{loss}})\rho(t). \quad (2.2)$$

The master equation consists of the superoperator \mathcal{M} , which is divided into several components. First, coherent evolution with the excitonic Hamiltonian H_e is described by the superoperator $\mathcal{M}_H = -i[H_e, \cdot]$. In addition, decoherence due to the interaction with the phonon bath is incorporated by $\mathcal{M}_{\text{decoherence}}$. $\mathcal{M}_{\text{decoherence}}$ depends on the spectral density, which models the coupling strengths of the phonon modes to the system. Finally, one has the processes for trapping to a reaction center $\mathcal{M}_{\text{trap}}$ and exciton loss $\mathcal{M}_{\text{loss}}$ due to spontaneous emission. Associated with these processes are the trapping rate κ and the loss rate Γ . Details about the trapping and exciton loss processes can be found in [26, 56].

The secular Redfield theory is valid in the regime of weak system-bath coupling. The superoperator $\mathcal{M}_{\text{decoherence}}$ is of Lindblad form with Lindblad operators for relaxation in the exciton basis and for dephasing of excitonic superpositions. The

relaxation rates depend on the spectral density evaluated at the particular excitonic transition frequencies, satisfy detailed balance, and depend on temperature through the Bose-Einstein distribution. The dephasing rates are linear in temperature. We use the same Ohmic spectral density as in [29], i.e. $J(\omega) = 2\lambda\gamma\omega/\pi(\omega^2 + \gamma^2)$, where $1/\gamma$ is the bath correlation time. For $1/\gamma = 50$ fs, this spectral density shows only modest differences to the spectral density used in [26]. Further details about the Lindblad model can be found in [26].

The hierarchy equation of motion approach [29] consistently interpolates between weak and strong system bath coupling. The assumption that the fluctuations are Gaussian makes the second-order cumulant expansion exact. The resulting equation of motion can be expressed as an infinite hierarchy of system, i.e. $\rho(t)$, and connected auxiliary density operators $\{\sigma_i\}$, arranged in tiers. For numerical simulation, "far-away" tiers in the hierarchy are truncated in a sensible manner. The hierarchy equation of motion can also be written as in Eq. (2.2) when we make the replacement $\rho(t) \rightarrow (\rho(t), \sigma_1, \sigma_2, \dots)$ and use the hierarchical structure discussed in [29] for the decoherence superoperator $\mathcal{M}_{\text{decoherence}}$. For simulations of the Fenna-Matthews-Olson complex, we use the scaled hierarchy approach developed in [57]. It was shown recently that four tiers of auxiliary density operators are enough for accurate room temperature simulations [58], which enables the rapid computation of efficiency and total coherences. The trapping and exciton loss processes are naturally extended to the auxiliary systems.

In our previous work [26], we developed a method to quantify the role of quantum coherence to the transfer efficiency. The energy transfer efficiency (ETE) is given by

the integrated probability of leaving the system from the sites that are connected to the trap instead to being lost to the environment. That is, $\eta = \int_0^\infty dt \text{Tr}\{\mathcal{M}_{\text{trap}}\rho(t)\}$. It was shown that the ETE can be partitioned into $\eta = \eta_{\text{H}} + \eta_{\text{decoherence}}$, where the efficiency due to the coherent dynamics with the excitonic Hamiltonian is given by:

$$\eta_{\text{H}} = \text{Tr}\{\mathcal{M}_{\text{trap}}(\mathcal{M}_{\text{trap}} + \mathcal{M}_{\text{loss}})^{-1}\mathcal{M}_{\text{H}}\mathcal{M}^{-1}\rho(0)\}. \quad (2.3)$$

The ETE contribution $\eta_{\text{decoherence}}$ involves $\mathcal{M}_{\text{decoherence}}$, i.e.,

$$\eta_{\text{decoherence}} = \text{Tr}\{\mathcal{M}_{\text{trap}}(\mathcal{M}_{\text{trap}} + \mathcal{M}_{\text{loss}})^{-1}\mathcal{M}_{\text{decoherence}}\mathcal{M}^{-1}\rho(0)\}.$$

In this work, we extend our ETE contribution method to quantify the role of the initial state to the ETE. We obtain a separation of the coherent contribution, $\eta_{\text{H}} = \eta_{\text{init}} + \eta_{\text{dyn}}$, where the efficiency η_{init} can be ascribed to the initial state. The η_{dyn} is defined by $\eta_{\text{dyn}} = \eta_{\text{H}} - \eta_{\text{init}}$ and can be interpreted as dynamical part of the coherence contribution arising during the time evolution. For the computation of η_{init} , we note that one can always express the ensemble described by the system density matrix as $\rho(t) = p_{\text{init}}(t)|\psi_{\text{init}}(t)\rangle\langle\psi_{\text{init}}(t)| + \sum_k p_k(t)\rho_k(t)$. Here, $p_{\text{init}}(t)$ is the probability of the quantum system being in the (Hamiltonian time-evolved) initial state $|\psi_{\text{init}}(t)\rangle$, where $p_{\text{init}}(0) = 1$. The $p_k(t)$ are the probabilities of being in some other ensemble state $\rho_k(t)$, where $p_{\text{init}}(t) + \sum_k p_k(t) = 1$. The probability $p_{\text{init}}(t)$ is reduced by the interaction with the environment and readily computed for Markovian Lindblad dynamics by considering the damped no-jump evolution due to the decoherence superoperator $\mathcal{M}_{\text{decoherence}}$ [56, 59, 60]. Therefore, we can compute the efficiency pertaining to the initial state by $\eta_{\text{init}} = \int_0^\infty dt \text{Tr}\{\mathcal{M}_{\text{trap}}p_{\text{init}}(t)|\psi_{\text{init}}(t)\rangle\langle\psi_{\text{init}}(t)|\}$. Together with Equation (2.3), this obtains the desired separation $\eta_{\text{H}} = \eta_{\text{init}} + \eta_{\text{dyn}}$.

Additionally, we employ another measure for the role of coherence by straightforwardly integrating over time all the coherence elements of the density matrix. That is:

$$C(\lambda) = \sum_{m \neq n} \int_0^\infty dt |\langle m | \rho(t) | n \rangle|. \quad (2.4)$$

We normalize with respect to the case of coherent evolution at $\lambda = 0.0/\text{cm}$, i.e. $\tilde{C}(\lambda) = C(\lambda)/C(0)$. Based on the discussion in [52], the quantity \tilde{C} can be considered as the (normalized) integrated entanglement (concurrence) that is present before the exciton is trapped in the reaction center or lost to the environment. We note that the total coherence measure \tilde{C} is similar in spirit to a measure of the first kind discussed above. This is because the normalization essentially performs a comparison of the actual model at a certain λ with an artificial model at $\lambda = 0$. (For the numerical evaluation, the integral in Eq. (2.4) is computed until $\text{Tr}\{\rho(t)\} \leq 10^{-3}$.)

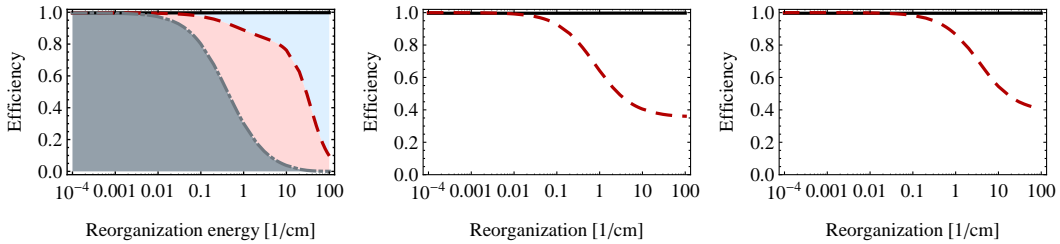


Figure 2.1: (Left panel) Efficiency η (solid black) and contributions of initial state η_{init} (dash-dotted gray) and coherent evolution $\eta_{\text{init}} + \eta_{\text{dyn}}$ (dashed red) for a dimer that is based on the strongly coupled sites 1 and 2 of the Fenna-Matthews-Olson complex using the secular Redfield model. The initial state is at site 1 and the target is site 2. At a physiological value of around $\lambda = 35/\text{cm}$, one finds $\eta_{\text{init}} = 0.0$ and $\eta_{\text{dyn}} = 0.43$. (Center panel) Efficiency and integrated coherence \tilde{C} for the dimer with the secular Redfield approach. At $\lambda = 35/\text{cm}$ there is $\tilde{C} = 0.37$. (Right panel) Same quantities as in the center panel for the dimer using the hierarchy equation of motion approach with 15 tiers of auxiliary systems. At $\lambda = 35/\text{cm}$, one finds $\tilde{C} = 0.44$. The parameters are $1/\kappa = 1$ ps, $1/\Gamma = 1$ ns, and $1/\gamma = 50$ fs for all panels.

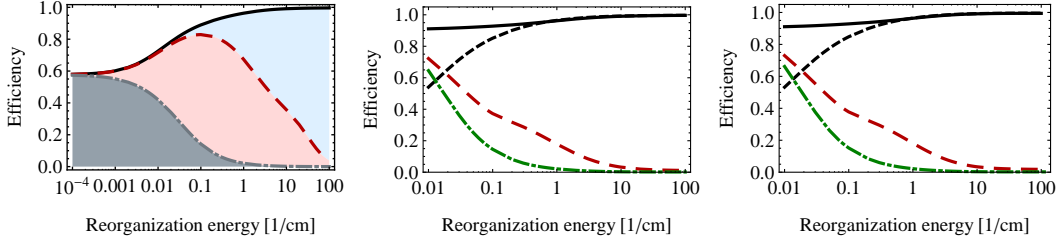


Figure 2.2: (Left panel) Efficiency η (solid black) and contributions of initial state η_{init} (dash-dotted gray) and coherent evolution $\eta_{\text{init}} + \eta_{\text{dyn}}$ (dashed red) for the Fenna-Matthews-Olson complex using the secular Redfield model. The initial state is a classical mixture of site 1 and 6 and the target site for trapping is site 3. The actual system has a reorganization energy of around $\lambda = 35/\text{cm}$, where $\eta_{\text{init}} = 0.0$ and $\eta_{\text{dyn}} = 0.17$. (Center panel) Efficiency for initial site 1 (solid black) and initial site 6 (dashed black) and integrated coherence \tilde{C} for initial site 1 (dashed red) and initial site 6 (dash-dotted green) for the Fenna-Matthews-Olson complex with the secular Redfield approach. At $\lambda = 35/\text{cm}$ there is $\tilde{C} = 0.0151$ (initial site 1) and $\tilde{C} = 0.0017$ (initial site 6). (Right panel) Same quantities as in the center panel for the FMO complex using the scaled hierarchy equation of motion approach with four tiers of auxiliary systems. At $\lambda = 35/\text{cm}$, one finds $\tilde{C} = 0.020$ (initial site 1) and $\tilde{C} = 0.0022$ (initial site 6). The parameters are $1/\kappa = 1$ ps, $1/\Gamma = 1$ ns, and $1/\gamma = 50$ fs for all plots.

In Fig. 2.1, we present the two measures of coherence for a dimer system. For the dimer, we take the sites 1 and 2 of the FMO complex with $\epsilon_1 = 0/\text{cm}$, $\epsilon_2 = 120/\text{cm}$, and $J = -87.7/\text{cm}$, see [14], and room temperature. This system will also be the focus of the following sections on the atomistic detail simulations and quantum process tomography. Here, for studying the role of quantum coherence, we assume that the task is defined by the exciton initially being at the lower energy site 1 and the target site being site 2. In the left panel of Fig. 2.1 we show the efficiency η , the contribution η_{H} from Eq. (2.3), and η_{init} for the secular Redfield model. In the present small system, environment-assisted transport is relatively unimportant, with the efficiency as a function of the reorganization energy being close to unity everywhere. The underlying contributions show a transition from a regime dominated by coherent evolution to a regime dominated by incoherent Lindblad jumps. At $\lambda = 35/\text{cm}$, we find $\eta_{\text{init}} = 0\%$ and $\eta_{\text{H}} = 43\%$. In Fig. 2.1 (center panel), we find that the total coherence measure \tilde{C} for the dimer is around 0.37 for $\lambda = 35/\text{cm}$. In Fig. 2.1 (right panel), the total coherence is plotted for the dimer in the hierarchy equation of motion approach. We use 15 tiers of auxiliary systems. At $\lambda = 35/\text{cm}$, we find $\tilde{C} = 0.44$; because of the sluggish, non-equilibrium bath there is more coherence than in the secular Redfield model.

In Fig. 2.2 (left panel), we present the coherent, decoherent, and initial state contribution to the ETE for the Fenna-Matthews-Olson complex as a function of the reorganization energy for the secular Redfield model at room temperature. We use the Hamiltonian given in [14] and the contribution measures given in Equation (2.3) and by η_{init} . The initial state is a classical mixture of site 1 and 6. For small

reorganization energy, the efficiency is around $\eta = 60\%$ and for larger reorganization energies we observe environment-assisted quantum transport (ENAQT) [27], with the efficiency rising up to almost $\eta = 100\%$ for the physiological value of $\lambda = 35/\text{cm}$. The contributions measures η_{dyn} and η_{init} reveal the underlying dynamics. The quantum dynamical contribution η_{dyn} is around 17% at $\lambda = 35/\text{cm}$ ¹. In our model, this part is due to an interplay of the Hamiltonian dynamics and the trapping/loss dynamics, which both have their preferred basis being the site basis. The main part of the efficiency at $\lambda = 35/\text{cm}$ is due to incoherent Lindblad jumps, having a value of $\eta_{\text{decoherence}} = 83\%$. The initial state contribution is relevant only at small values of the reorganization energy.

In Fig. 2.2 (center and right panel), we compare the efficiency and the coherence measure \tilde{C} for the secular Redfield and the hierarchy equation of motion approach [29] for the Fenna-Matthews-Olson complex. The initial state is either localized at site 1 or at site 6. Four tiers of auxiliary systems were used in the computation, which already lead to a good agreement with [29] for the dynamics at $\lambda = 35\text{cm}^{-1}$, $1/\gamma = 50$ fs, and room temperature. In Fig. 2.2 (right panel), ENAQT is observed with increasing reorganization energy also in the hierarchy approach, with the efficiency rising up to almost $\eta = 100\%$ at $\lambda = 35/\text{cm}$. In Fig. 2.2 (center and right panel), it is observed that the normalized total coherences of the density matrix decrease with increasing reorganization energy. For the secular Redfield case, we obtain $\tilde{C}(\lambda = 35\text{cm}^{-1}) = 0.0151$ for the initial site 1 and $\tilde{C}(\lambda = 35\text{cm}^{-1}) = 0.0017$ for the initial site 6. For the hierarchy case, we obtain more coherence, i.e. $\tilde{C}(\lambda = 35\text{cm}^{-1}) = 0.020$ for the initial

¹In Ref. 26, we found the value $\eta_{\text{H}} = 10\%$ for a different Hamiltonian and a different spectral density.

site 1 and $\tilde{C}(\lambda = 35\text{cm}^{-1}) = 0.0022$ for the initial site 6. In both models, coherence is more important for the rugged energy landscape of the pathway from site 1 than for the funnel-type energy landscape of the pathway from site 6.

Master equation approaches, such as the ones discussed in this section suffer from various drawbacks. Redfield theory is only applicable in the limit of weak system bath coupling and does not take into account non-equilibrium molecular reorganization effects. The hierarchy equation of motion approach assumes Gaussian fluctuations and Ohmic Drude-Lorentz spectral densities. The detailed atomistic structure of the protein and the chlorophylls is not taken into account in these approaches. The results thus provide a general indication of the behavior of the actual system but not a conclusive and detailed theoretical proof. In the next section, we will present a first step toward such a detailed study with our combined molecular dynamics/quantum chemistry method. The atomistic structure is included and realistic spectral densities can be obtained. We also present a straightforward method to simulate exciton dynamics beyond master equations. We thus address the second question of the microscopic origins of the long-lived quantum coherence.

2.3 Molecular Dynamics Simulations

Among many other biologically functional components, protein complexes are essential components of the photosynthetic system. Proteins remain as one of the main topics of biophysical research due to their diverse and unidentified structure-function relationship. Many biological units are highly optimized and efficient, so that even a point mutation of a single amino acid in conserved region often results in the

loss of the functionality [61–63]. Have the photosynthetic system adopted quantum mechanics to improve its efficiency in its course of evolution? To answer this question, careful characterization of the protein environment to the atomistic detail is necessary to identify the microscopic origin of the long-lived quantum coherence. As explained in the previous section, the contribution of the quantum coherence to the energy transfer efficiency in biological systems have been successfully carried out, yet a more detailed description of the bath in atomic detail would be desirable to investigate the structure-function relationship of the protein complex and to test validity of the assumptions used in popular models of the photosynthetic system.

The site energy of a chromophore is a complex function of the configuration of the chromophore molecule, and the relative orientation of the molecule to that of the embedding protein and that of other chromophore molecules. Factors affecting site energies have intractably large degrees of freedom, so it is reasonable to treat those degrees of freedom as the bath of an open quantum system. The state of the system is assumed to be restricted to the single exciton manifold. To construct a system-bath relationship with atomistic detail of the bath, we start from the total Hamiltonian operator, and decomposed the operator in such a way that the system-bath Hamiltonian is not assumed to be any specific functional form:

$$\begin{aligned}
 H_{total} = & \sum_m \epsilon_m(\mathbf{R}_{ch}, \mathbf{R}_{prot}) |m\rangle \langle m| + \sum_{m,n} \{J_{mn}(\mathbf{R}_{ch}, \mathbf{R}_{prot}) |m\rangle \langle n| + c.c.\} \\
 & + T_{ch} + T_{prot} + V_{ch}(\boldsymbol{\sigma}, \mathbf{R}_{ch}, \mathbf{R}_{prot}) + V_{prot}(\mathbf{R}_{ch}, \mathbf{R}_{prot}).
 \end{aligned} \tag{2.5}$$

ϵ_m represents the site energy of m th site, J_{mn} is the coupling constant between m th and n th sites. $\boldsymbol{\sigma}$ denotes the excitonic state of chromophores, \mathbf{R}_{ch} corresponds to the

nuclear coordinates of chromophore molecules, and \mathbf{R}_{prot} are the nuclear coordinates of the remaining protein and enclosing water molecules. T and V are the corresponding kinetic and potential energy operators for the chromophores and proteins respectively under Born-Oppenheimer approximation. The potential energy term for chromophores depends on the exciton state of the system, because dynamics of a molecule will be governed by different Born-Oppenheimer surface when its excitonic state changes. However, as a first approximation, we assumed that the change of Born-Oppenheimer surfaces does not affect the bath dynamics significantly. With this assumption, we can ignore the dependence of the excitonic state in the V_{ch} term and the system-bath Hamiltonian only contains the one way influence from the bath to the system:

$$\begin{aligned}
 H_{total} &\approx \sum_m \epsilon_m(\mathbf{R}_{ch}, \mathbf{R}_{prot}) |m\rangle\langle m| + \sum_{m,n} J_{mn}(\mathbf{R}_{ch}, \mathbf{R}_{prot}) |m\rangle\langle n| \\
 &+ \sum_m \epsilon_m(\mathbf{R}_{ch}, \mathbf{R}_{prot}) |m\rangle\langle m| + T_{ch} + T_{prot} + V_{ch}(\mathbf{R}_{ch}, \mathbf{R}_{prot}) + V_{prot}(\mathbf{R}_{ch}, \mathbf{R}_{prot}) \\
 &= \underbrace{\sum_m \bar{\epsilon}_m |m\rangle\langle m| + \sum_{m,n} \bar{J}_{mn} |m\rangle\langle n|}_{H_S} \\
 &+ \underbrace{\sum_m \{\epsilon_m(\mathbf{R}_{ch}, \mathbf{R}_{prot}) - \bar{\epsilon}_m\} |m\rangle\langle m| + \sum_{m,n} \{J_{mn}(\mathbf{R}_{ch}, \mathbf{R}_{prot}) - \bar{J}_{mn}\} |m\rangle\langle n|}_{H_{SB}} \\
 &+ \underbrace{T_{ch} + T_{prot} + V_{ch}(\mathbf{R}_{ch}, \mathbf{R}_{prot}) + V_{prot}(\mathbf{R}_{ch}, \mathbf{R}_{prot})}_{H_B}. \tag{2.6}
 \end{aligned}$$

Based on this decomposition of the total Hamiltonian, we set up a model of the FMO complex in atomistic detail with the AMBER force field [64, 65] and approximate the propagation of the entire complex by classical mechanics. Molecular dynamics simulations were conducted at 77K and 300K with an isothermal-isobaric (NPT)

ensemble. The parameters for the system and the system-bath Hamiltonian were calculated using quantum chemistry methods along the trajectory from the molecular dynamics simulation. ϵ_m was calculated using the Q-Chem quantum chemistry package [66]. The electronic excitations were modeled using the time-dependent density functional theory using the Tamm-Dancoff approximation. The density functional employed was BLYP and the basis set employed was 3-21G*. External charges from the force field were included in the calculation as the electrostatic external potential. The coupling terms, J_{mn} , were obtained from the Hamiltonian presented in Ref. 14 and considered to be constant in time. $\bar{\epsilon}_m$ was chosen as time averaged site energy for the m th site to minimize the magnitude of the system-bath Hamiltonian. In this work, only site 1 and site 2 were considered for the exciton dynamics. However, the methodology can be applied for the exciton dynamic of all seven chromophores.

To obtain a closed-form equation for the reduced density matrix, we applied mean-field approximation [67]; because no feedback from the system to the bath was assumed, the state of the bath is not affected by the state of the system. Therefore, the total density matrix, $W(t)$, can be factorized into the reduced density matrix $\rho(t)$, and $B(t)$ which is defined only in the Hilbert space of the bath. With additional assumption that the bath is in thermal equilibrium, we can obtain the closed equation for the reduced density matrix.

$$\begin{aligned}
 \frac{\partial}{\partial t} \rho(t) &= -\frac{i}{\hbar} [H_S, \rho(t)] - \frac{i}{\hbar} \text{Tr} \{ [H_{SB}, W(t)] \} \\
 &\approx -\frac{i}{\hbar} [H_S, \rho(t)] - \frac{i}{\hbar} [\text{Tr} \{ H_{SB} B(t) \}, \rho(t)] \\
 &\approx -\frac{i}{\hbar} [H_S, \rho(t)] - \frac{i}{\hbar} [\text{Tr} \{ H_{SB} B_{eq}(t) \}, \rho(t)].
 \end{aligned} \tag{2.7}$$

Thermal equilibrium of the bath was ensured by the thermostat of the molecular

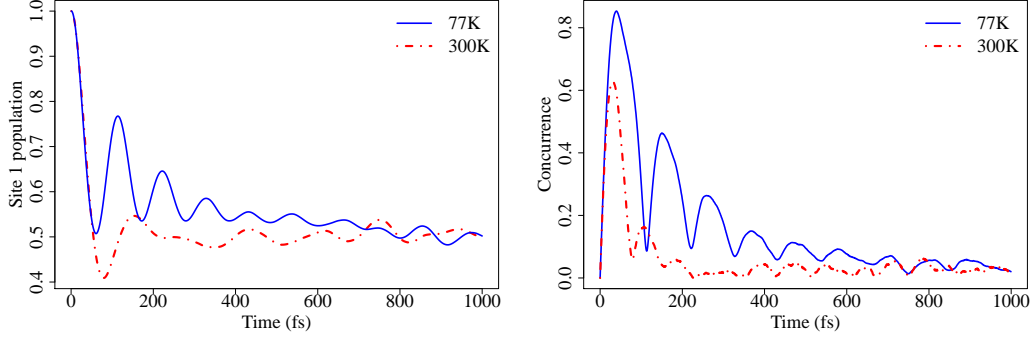


Figure 2.3: (Left panel) Time evolution of the exciton population at the site 1 (ρ_{11}) based on the strongly coupled site 1 and 2 of the FMO complex at 77K and 300K. The initial pure state $\rho = |1\rangle\langle 1|$ was propagated using Monte Carlo integration of unitary evolutions, where the time-dependent site energies are obtained from a combined molecular dynamics/quantum chemistry approach. The asymptotic distribution does not follow a Boltzmann distribution because relaxation of the system to the bath is not considered. (Right panel) The concurrence between site 1 and 2 at 77K and 300K. Quantum coherence lives longer at a lower temperature.

dynamics simulation. Thus, the reduced density matrix was obtained by Monte Carlo integration of 4000 independent instances of unitary quantum evolution with respect to the thermally equilibrated bath. Each instance was propagated by integrating the Schrödinger equation with the simple exponential integrator.

Fig. 2.3 shows the change of the population of the site 1, ρ_{11} , and the concurrence between site 1 and 2. The population is evenly distributed between the two sites because relaxation was not considered. The concurrence, $2|\rho_{12}|$, is an indicator of pairwise entanglement for the system [52]. Note that the coherence builds up during the first ~ 100 fs, and then decreases subsequently due to the decoherence from the bath.

Fig. 2.4 shows the spectral density of the first chromophore. Although the spectral density of the bath from molecular dynamics simulation shows characteristic frequen-

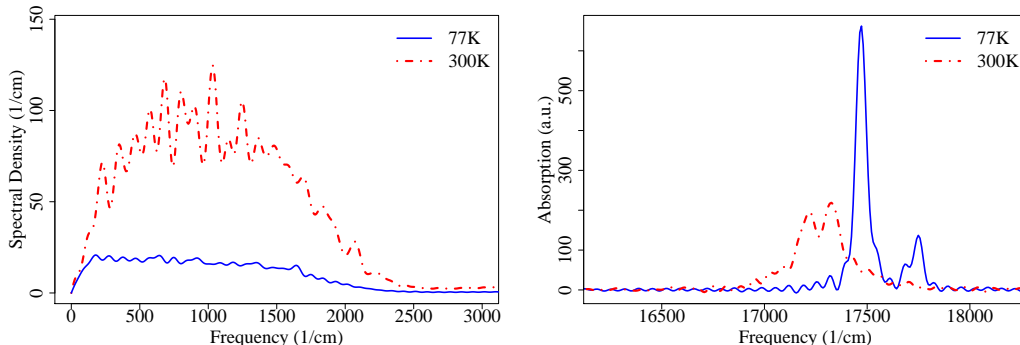


Figure 2.4: (Left panel) Spectral density from the autocorrelation function of the site 1 of the FMO complex from the molecular dynamics simulation at 77K and 300K. While the spectral density reflects the characteristic vibrational modes of the protein and the chromophore molecule, high-frequency modes are overpopulated due to the limitation of the Newtonian mechanics. (Right panel) Absorption spectrum of site 1 and 2 at 77K and 300K.

cies related to the actual protein environment and the bacteriochlorophyll molecule, high-frequency modes are overpopulated due to the limitation of the classical mechanics. There are efforts to incorporate quantum effects into the classical molecular dynamics simulation in a slightly different context [68–70], and we are investigating the possibilities of applying these corrections.

Another simplification employed was the omission of the feedback from exciton states. When the exciton state of a bacteriochlorophyll is changed, the Born-Oppenheimer surface which governs the dynamics of the chromophore molecule should be also changed. The current scheme only propagates the protein complex on the electronically ground-state surface. Incorporating the feedback could lead to the different characteristics of the protein bath. There exist several schemes for mixed quantum-classical dynamics [71–73] which potentially resolve the problem at the additional computational cost of simultaneously propagating excitons and protein bath.

Calculations are underway to carry out the full seven-site simulation of the FMO complex at different temperatures to compare with experimental temperature-dependent results [25].

In the following final section, we will describe our quantum process tomography scheme, which is a spectroscopic technique associated with a computational procedure for direct extraction of the parameters related to the quantum evolution of the system, in terms of *quantum process maps*.

2.4 Quantum Process Tomography

So far, we have delved into several theoretical models to characterize quantum coherence in the entire FMO complex and in a dimer subsystem of it. Experimentally, however, a clear characterization of this coherence is still elusive. Signatures of long lived quantum superpositions between excitonic states in multichromophoric systems are potentially monitored through four wave-mixing techniques [23, 74, 75]. However, a transparent description of the evolving quantum state of the probed system is not necessarily obtained from a single realization of such experiments. In these, a series of three weak incoming ultrashort pulses sent from a noncollinear setup induce a macroscopic third order polarization in the sample. The latter manifests in a time dependent spatial grating which emits a macroscopic polarization that interferes with a fourth pulse, called the local oscillator. From an operational standpoint, this last pulse selects the spatial Fourier component of the polarization which corresponds to its wavevector (heterodyne detection), hence earning the name of four wave-mixing for this technique (FWM) [75]. Extracting specific Fourier components of the induced

polarization allows for the selection of a particular set of processes in the density matrix of the probed system, as each wavevector is associated with a carrier frequency of the pulse. These processes can be intuitively understood by keeping track of the dual Feynman diagrams that account for the perturbations that the pulses induce on the bra or ket sides of the density matrix of the probed system. Whereas the analysis of these experiments is naturally carried out in the density matrix formalism, an important question is whether the density matrix itself can be imaged via these experiments, a problem known as quantum state tomography (QST) [76]. If this were possible, quantum process tomography (QPT) could also be carried out, therefore providing a complete characterization of excited state dynamics [77]. In a previous study, we showed that a series of two-color heterodyned rephasing photon-echo (PE) experiments repeated in different polarization configurations yields the necessary information to carry out QST and QPT of the single-exciton manifold of a coupled heterodimer [78]. In the present article, we adapt our previous theory to extract this information from two-dimensional spectra, similar to those employed in current experiments. An comprehensive study of this possibility has been presented in [79]. Here, we shall highlight some key features of the method.

We begin by reviewing some basic aspects of QPT. Under very general assumptions, the evolution of an open quantum system can be described by a linear transformation [80]:

$$\rho_{ab}(T) = \sum_{cd} \chi_{abcd}(T) \rho_{cd}(0), \quad (2.8)$$

where $\rho_{ab}(T)$ is the element ab of the reduced density matrix ρ of the system at time T . Equation (2.8) is remarkable in that $\chi(T)$ is independent of the initial

state. Knowledge of $\chi(T)$ implies a complete characterization of the dynamics of the reduced system and, in fact, QPT can be operationally defined as the procedure to obtain $\chi(T)$. Conceptually, it is straightforward to recognize that, due to linearity, $\chi(T)$ can be inverted by preparing a complete set of inputs, evolving them for time T , and detecting the outputs along a complete basis. In the context of nonlinear optical spectroscopy, this is exactly the strategy we shall follow, with a few caveats due to experimental constraints.

To place the discussion in context, we shall be again concerned with the subsystem composed of the excitonic dimer between sites 1 and 2 of the FMO complex. For simplicity, we ignore the rest of the sites in this theoretical study. We only need to be concerned with four eigenstates of this model system: The ground state $|g\rangle$, the delocalized single-excitons $|\alpha\rangle$ and $|\beta\rangle$, and the biexciton $|f\rangle$, which in the photosynthetic system can be safely assumed to be the direct sum of the single-excitons without significant interactions between them. Therefore, the biexciton energy level is just $\omega_f = \omega_\alpha + \omega_\beta$. We label the delocalized excitons so that $|\alpha\rangle$ is the higher energy eigenstate compared to $|\beta\rangle$. Denoting the transition energies between the i -th and the j -th states by $\omega_{ij} = \omega_i - \omega_j$, it follows that $\omega_{\alpha g} = \omega_{f\beta}$ and $\omega_{\beta g} = \omega_{f\alpha}$ [81]. The excitonic system is not isolated, and in fact, it interacts with a phonon and photon bath which induces relaxation and dephasing processes in it.

The experimental technique we consider is photon-echo (PE) spectroscopy, which is a particular subset of FWM techniques where the wavevector of the fourth pulse corresponds to the phase-matching condition $\mathbf{k}_{PE} = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$, with \mathbf{k}_i being the wavevector corresponding to the i -th pulse. Here, the labeling of the pulses cor-

responds to the order in which the fields interact with the sample. Typically, the ultrashort pulses employed to study these excitonic systems possess an optical carrier frequency, therefore allowing transitions which are resonant with the frequency components $\pm\omega_{\beta g}$ and $\pm\omega_{\alpha g}$. In PE experiments, the first pulse centered at t_1 creates an optical coherence beating at a frequency $\omega_{g\alpha}$ or $\omega_{g\beta}$. At $t_2 = t_1 + \tau$, the second pulse creates a coherence or a population in the single exciton manifold. At $t_3 = t_2 + T$, the third pulse generates another optical coherence, but this time, beating at the frequencies opposite to the ones in the first interval, that is, at frequencies $\omega_{\alpha g}$ or $\omega_{\beta g}$, causing a rephasing echo of the signal. The heterodyne detection of the nonlinear polarization signal $P_{PE}(\tau, T, t)$ occurs at time $t_4 = t_3 + t$. Borrowing from NMR jargon, the intervals (t_1, t_2) , (t_2, t_3) , and (t_3, t_4) are traditionally referred to as *coherence*, *waiting*, and *echo* times, and their durations are τ , T , and t , respectively. This nomenclature should not be taken literally. For example, in most cases, coherences do not only evolve in the coherence time, but in the waiting and echo times. Similarly, the waiting time is often referred to as population time, which hosts dynamics of both populations and coherences. For a historical perspective on this vocabulary, we refer the reader to any comprehensive NMR treatise such as [82].

The experiment is systematically repeated for many durations for each interval. In order to 'watch' single-exciton dynamics, it is convenient to isolate the changes on the signal due to the waiting time T . This exercise is accomplished by performing a double Fourier transform of the signal along the τ and t axes, which yields a 2D spectra that evolves in T [83–85]:

$$S(\omega_\tau, T, \omega_T) = \int_0^\infty d\tau \int_0^\infty dt P_{PE}(\tau, T, t) e^{-i\omega_\tau \tau + i\omega_T T} \quad (2.9)$$

In order to map a PE experiment to a QPT, we identify the coherence interval as the preparation step and the echo interval as the detection step. This assumption implies that the optical coherence intervals have well characterized dynamics. This hypothesis is reasonable due to a separation of timescales where optical coherences will presumably decay exponentially due to pure dephasing and not due to intricate phonon-induced processes. Therefore, the 2D spectrum consists of four Lorentzian peaks centered about $(\omega_\tau, \omega_t) = (\omega_{\alpha g}, \omega_{\alpha g}), (\omega_{\alpha g}, \omega_{\beta g}), (\omega_{\beta g}, \omega_{\alpha g}), (\omega_{\beta g}, \omega_{\beta g})$. In this discussion, we shall ignore inhomogeneous broadening, noting that it can always be accounted for as a convolution of the signal with the distribution of inhomogeneity. The width of these Lorentzians can be directly related to the dephasing rates of the optical coherences. Loosely speaking, a particular value on the ω_τ axis of the spectrum indicates a specific type of state preparation, whereas the ω_t axis is related to a particular detection. More precisely, a peak in the 2D spectrum displays the correlations between the frequency beats from the coherence and echo intervals. A crucial realization is that the amplitude of these peaks can be written as a linear combination of elements of the time evolving excitonic density matrix stemming from

different initial states, that is, of elements of $\chi(T)$ itself [79]:

$$\begin{aligned}
 \tilde{S}(\omega_{\alpha g}, T, \omega_{\alpha g}) = & -C_{\omega_1}^{\alpha} C_{\omega_2}^{\alpha} (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_1) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_2) \\
 & \times \{C_{\omega_3}^{\alpha} [(\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4) (\chi_{gg\alpha\alpha}(T) - 1 - \chi_{\alpha\alpha\alpha\alpha}(T)) \\
 & + (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) \chi_{\beta\beta\alpha\alpha}(T)] \\
 & + C_{\omega_3}^{\beta} [(\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4)) \chi_{\alpha\beta\alpha\alpha}(T)]\} \\
 & - C_{\omega_1}^{\alpha} C_{\omega_2}^{\beta} (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_1) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_2) \\
 & \times \{C_{\omega_3}^{\alpha} [(\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4) (\chi_{gg\beta\alpha}(T) - \chi_{\alpha\alpha\beta\alpha}(T)) \\
 & + (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) \chi_{\beta\beta\beta\alpha}(T)] \\
 & + C_{\omega_3}^{\beta} [((\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4)) \chi_{\alpha\beta\beta\alpha}(T)]\},
 \end{aligned} \tag{2.10}$$

$$\begin{aligned}
 \tilde{S}(\omega_{\alpha g}, T, \omega_{\beta g}) = & -C_{\omega_1}^{\alpha} C_{\omega_2}^{\alpha} (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_1) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_2) \\
 & \times \{C_{\omega_3}^{\beta} [(\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_4) (\chi_{gg\alpha\alpha}(T) - 1 - \chi_{\beta\beta\alpha\alpha}(T)) \\
 & + (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_4) \chi_{\alpha\alpha\alpha\alpha}(T)] \\
 & + C_{\omega_3}^{\alpha} [((\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_4)) \chi_{\beta\alpha\alpha\alpha}(T)]\} \\
 & - C_{\omega_1}^{\alpha} C_{\omega_2}^{\beta} (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_1) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_2) \\
 & \times \{C_{\omega_3}^{\beta} [(\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_4) (\chi_{gg\beta\alpha}(T) - \chi_{\beta\beta\beta\alpha}(T)) \\
 & + (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_4) \chi_{\alpha\alpha\beta\alpha}(T)] \\
 & + C_{\omega_3}^{\alpha} [((\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_4)) \chi_{\beta\alpha\beta\alpha}(T)]\},
 \end{aligned} \tag{2.11}$$

$$\begin{aligned}
\tilde{S}(\omega_{\beta g}, T, \omega_{\alpha g}) = & -C_{\omega_1}^{\beta} C_{\omega_2}^{\beta} (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_1) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_2) \\
& \times \{C_{\omega_3}^{\alpha} [(\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4) (\chi_{gg\beta\beta}(T) - 1 - \chi_{\alpha\alpha\beta\beta}(T)) \\
& + (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) \chi_{\beta\beta\beta\beta}(T)] \\
& + C_{\omega_3}^{\beta} [(\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4) \chi_{\alpha\beta\beta\beta}(T)]\} \\
& - C_{\omega_1}^{\beta} C_{\omega_2}^{\alpha} (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_1) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_2) \\
& \times \{C_{\omega_3}^{\alpha} [(\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4) (\chi_{gg\alpha\beta}(T) - \chi_{\alpha\alpha\alpha\beta}(T)) \\
& + (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) \chi_{\beta\beta\alpha\beta}(T)] \\
& + C_{\omega_3}^{\beta} [((\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4)) \chi_{\alpha\beta\alpha\beta}(T)]\},
\end{aligned} \tag{2.12}$$

$$\begin{aligned}
\tilde{S}(\omega_{\beta g}, T, \omega_{\beta g}) = & -C_{\omega_1}^{\beta} C_{\omega_2}^{\beta} (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_1) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_2) \\
& \times \{C_{\omega_3}^{\beta} [(\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_4) (\chi_{gg\beta\beta}(T) - 1 - \chi_{\beta\beta\beta\beta}(T)) \\
& + (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_4) \chi_{\alpha\alpha\beta\beta}(T)] \\
& + C_{\omega_3}^{\alpha} [((\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_4)) \chi_{\beta\alpha\beta\beta}(T)]\} \\
& - C_{\omega_1}^{\beta} C_{\omega_2}^{\alpha} (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_1) (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_2) \\
& \times \{C_{\omega_3}^{\beta} [(\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_4) (\chi_{gg\alpha\beta}(T) - \chi_{\beta\beta\alpha\beta}(T)) \\
& + (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_4) \chi_{\alpha\alpha\alpha\beta}(T)] \\
& + C_{\omega_3}^{\alpha} [((\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_3) (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_4)) \chi_{\beta\alpha\alpha\beta}(T)]\}.
\end{aligned} \tag{2.13}$$

Here, the expressions have been obtained using the rotating-wave approximation, as

well as the assumption of no overlap between pulses. $\boldsymbol{\mu}_{pq} = \boldsymbol{\mu}_{qp}$ is the transition dipole moment between states $p, q \in \{g, \alpha, \beta, f\}$. We have rescaled the spectra amplitudes to eliminate the details of the lineshape by multiplying them by the dephasing rates of the optical coherences in the coherence and echo intervals,

$$\tilde{S}(\omega_{pg}, T, \omega_{qg}) = \Gamma_{gp}\Gamma_{qg}S(\omega_{pg}, T, \omega_{qg}). \quad (2.14)$$

The coefficient $C_{\omega_i}^p$ is the amplitude of the i -th pulse at the frequency ω_{pg} ,

$$C_{\omega_i}^p = -\frac{\Lambda}{i}\sqrt{2\pi\sigma^2}e^{-\sigma^2(\omega_{pg}-\omega_i)^2/2}, \quad (2.15)$$

with Λ being the strength of the pulse and σ the width of the Gaussian pulse in time domain. Also, \mathbf{e}_i is the polarization of the i -th pulse. Both $C_{\omega_i}^p$ and \mathbf{e}_i are experimentally tunable parameters for the pulses.

Whereas Equations (14) and (15) presented in [78] correspond to a single value of τ and t , Equations (2.10), (2.11), (2.12), and (2.13) stem from Fourier transform of data collected at many τ and t times (see Ref. 79). Therefore, in principle, a 2D spectrum provides a more robust source of information from which to invert $\chi(T)$ than in the suggested 1D experiment. The displayed equations, albeit lengthy, are easy to interpret. For instance, consider the term which is proportional to $\chi_{\alpha\beta\alpha\alpha}(T)$ in Equation (2.10), which stems from the Feynman diagram depicted in Fig. 2.5. As expected, it consists of a waiting time where the initially prepared population $|\alpha\rangle\langle\alpha|$ is transferred to the coherence $|\alpha\rangle\langle\beta|$. This waiting time is escorted by a coherence $|g\rangle\langle\alpha|$ oscillating as $e^{(-i\omega_{g\alpha}-\Gamma_{g\alpha})\tau}$ which evolves during the coherence time and another set of coherences $|f\rangle\langle\beta|$ and $|\alpha\rangle\langle g|$ which evolve during the echo time as $e^{(-i\omega_{f\beta}-\Gamma_{f\beta})t} = e^{(-i\omega_{\alpha g}-\Gamma_{\alpha g})t}$. These two intervals correspond to the diagonal peak

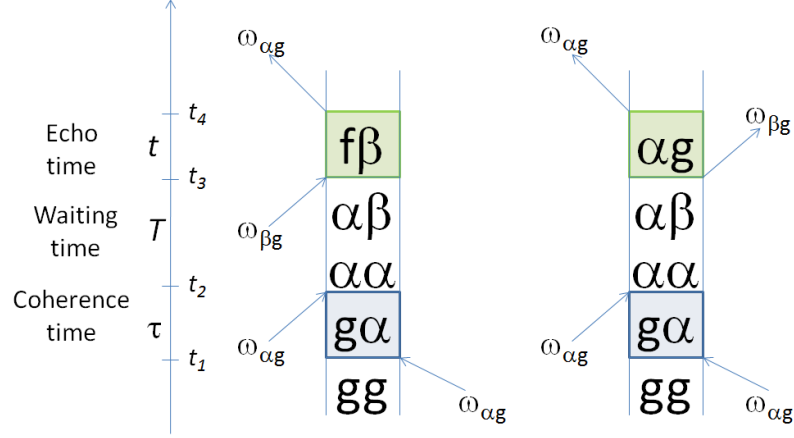


Figure 2.5: Dual Feynman diagrams that account for the population to coherence transfer terms $\chi_{\alpha\beta\alpha\alpha}(T)$ in quantum process tomography.

located at $(\omega_{\alpha g}, \omega_{\alpha g})$. Other processes that exhibit oscillations at those two respective frequencies appear as additional terms in the equation corresponding to that particular peak.

In Ref. 78, we showed that there are sixteen real valued parameters of $\chi(T)$ which need to be determined at every value of T in order to carry out QPT of the single exciton manifold of a heterodimer. For an illustration, we shall describe how to obtain the elements $\chi_{ij\alpha\alpha}(T)$. These quantities are shown in Fig. 2.6 and have been computed using the Ishizaki-Fleming model, with a bath correlation time of 150 fs [29]. They display rich and nontrivial phonon-induced behavior, such as the spontaneous generation of coherence from a population in an eigenstate of the excitonic Hamiltonian, and therefore, is a very good example of how QPT provides access to this nontrivial information via the repetition of a series of 2D PE experiments. For this particular set of $\chi(T)$ elements, we shall exploit the waveform of the pulses but not their polarizations, and for simplicity we will assume the polarization configuration

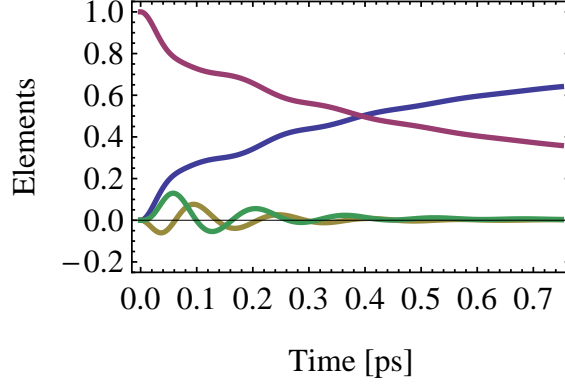


Figure 2.6: Transfer of population in eigenstate $|\alpha\rangle\langle\alpha|$ to other populations and coherences in the eigenbasis of the single exciton Hamiltonian. The hierarchy equation of motion approach is used for a dimer system based on the parameters of the site 1 and site 2 subsystem of the Fenna-Matthews-Olson complex. Population in $|\alpha\rangle\langle\alpha|$ decreases ($\chi_{\alpha\alpha\alpha\alpha}(T)$, purple) and is transferred to $|\beta\rangle\langle\beta|$ ($\chi_{\beta\beta\alpha\alpha}(T)$, blue). Emergence of coherence from the initial population occurs in this model ($\Re\{\chi_{\alpha\beta\alpha\alpha}(T)\}$, yellow and $\Im\{\chi_{\alpha\beta\alpha\alpha}(T)\}$, green).

$xxxx$ for each of the pulses including the heterodyning.

Consider the possibility of using pulses with carrier frequencies centered about $\omega_{\alpha g}$ and $\omega_{\beta g}$ respectively, and such that their bandwidth is narrow enough that the pulse centered about $\omega_{\alpha g}$ has negligible component at $\omega_{\beta g}$ and vice versa. Then, we can carry out an experiment such that $\frac{|C_{\omega_1}^\alpha|}{|C_{\omega_1}^\beta|}, \frac{|C_{\omega_2}^\alpha|}{|C_{\omega_2}^\beta|}, \frac{|C_{\omega_3}^\beta|}{|C_{\omega_3}^\alpha|} \gg 1$ (experiment 1) for all i and notice that the diagonal peak at $(\omega_{\alpha g}, \omega_{\alpha g})$ reduces to,

$$\begin{aligned} \langle \tilde{S}(\omega_\alpha, T, \omega_\alpha) \rangle_{xxxx} &= -C_{\omega_1}^\alpha C_{\omega_2}^\alpha C_{\omega_3}^\beta \\ &\times \langle (\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_1)(\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_2)[(\boldsymbol{\mu}_{f\alpha} \cdot \mathbf{e}_3)(\boldsymbol{\mu}_{f\beta} \cdot \mathbf{e}_4) - (\boldsymbol{\mu}_{\beta g} \cdot \mathbf{e}_3)(\boldsymbol{\mu}_{\alpha g} \cdot \mathbf{e}_4)] \rangle_{xxxx} \chi_{\alpha\beta\alpha\alpha}(T), \end{aligned} \quad (2.16)$$

which implies that its evolution with respect to T directly monitors the transfer of the population prepared at $|\alpha\rangle\langle\alpha|$ to the coherence at $|\alpha\rangle\langle\beta|$. Here, $\langle \cdot \rangle_{xxxx}$ denotes an isotropic average of the experiments performed with the $xxxx$ polarization configura-

tion. $\chi_{\alpha\beta\alpha\alpha}(T)$ can be directly obtained if information of the dipole moments is known in advance. As can be checked easily, $\chi_{\alpha\beta\alpha\alpha}(T) = (\chi_{\beta\alpha\alpha\alpha}(T))^*$ can, in principle, be also obtained directly from an experiment where $\frac{|C_{\omega_i}^\alpha|}{|C_{\omega_i}^\beta|} \gg 1$ for all i (experiment 2) and monitoring $\langle \tilde{S}(\omega_\alpha, T, \omega_\beta) \rangle_{xxxx}$. Redundant measurements can be used as ways of effectively constraining the QPT.

Similarly, the transfer from $|\alpha\rangle\langle\alpha|$ to other populations can be extracted by monitoring $\langle \tilde{S}(\omega_\alpha, T, \omega_\alpha) \rangle_{xxxx}$ in experiment 2 and $\langle \tilde{S}(\omega_\alpha, T, \omega_\beta) \rangle_{xxxx}$ in experiment 1. These two linearly independent conditions are enough to extract $\chi_{gg\alpha\alpha}(T)$, $\chi_{\alpha\alpha\alpha\alpha}(T)$, and $\chi_{\beta\beta\alpha\alpha}(T)$, since there is a third independent condition based on trace preservation which reads $\chi_{gg\alpha\alpha}(T) + \chi_{\alpha\alpha\alpha\alpha}(T) + \chi_{\beta\beta\alpha\alpha}(T) = 1$.

It is now important to verify whether the suggested experiments are feasible. In order to ensure conditions of the form $\frac{|C_{\omega_i}^\alpha|}{|C_{\omega_i}^\beta|} \gg 1$, we need $\sigma \sim \frac{3}{\omega_{\alpha g} - \omega_{\beta g}} \sim 75$ fs, that is, the pulse needs to be long enough to guarantee the narrow band condition. This requirement is very reasonable, as it is not too long to obscure the decoherence processes that we want to witness. In the case where the length of the pulse were of similar length as the dynamical events that one is interested in, it is not necessary to use very narrowband pulses either. The only essential requirement is a toolbox of two different waveforms for the pulses, for instance, a set of pulses centered about $\omega_{\alpha g}$ and $\omega_{\beta g}$ respectively, but having $\sigma \sim 30$ fs, for instance. By carrying out 8 experiments alternating the two waveforms in each of the three pulses, each of the terms in Equations (2.10), (2.11), (2.12), and (2.13) which are proportional to $C_{\omega_1}^i C_{\omega_2}^j C_{\omega_3}^k$ for $i, j, k \in \{\alpha, \beta\}$ may be inverted to yield the block diagonal set of equations discussed above.

In summary, we have presented three different tools for unraveling the role of quantum coherence in biological systems: a) techniques for obtaining the contribution of quantum coherences to biological processes; b) a microscopic simulation approach to explore the dynamics of these systems by direct simulation; and finally c) a new theoretical proposal for an experimental procedure that provides detailed information about the quantum process associated with energy transfer in the ultrafast regime. We believe that ultimately, a combination of these three techniques and tools from other groups will be collectively required to make definitive conclusions about the role of quantum coherence in photosynthetic complexes.

Chapter 3

Atomistic study of the long-lived quantum coherences in the Fenna-Matthews-Olson complex

3.1 Introduction

Recent experiments suggest the existence of long-lived quantum coherence during the electronic energy transfer process in photosynthetic light-harvesting complexes under physiological conditions [23–25]. This has stimulated many researchers to seek for the physical origin of such a phenomenon. The role and implication of quantum coherence during the energy transfer have been explored in terms of the theory of open quantum systems [27–29, 48, 50, 51, 60, 86–91], and also in the context of quantum information and entanglement [52–54, 92]. However, the characteristics of the protein environment, and especially its thermal vibrations or phonons, have not

been fully investigated from the molecular viewpoint. A more detailed description of the bath in atomic detail is desirable; to investigate the structure-function relationship of the protein complex and to go beyond the assumptions used in popular models of photosynthetic systems.

Protein complexes constitute one of the most essential components in every biological organism. They remain one of the major targets of biophysical research due to their tremendously diverse and, in some cases, still unidentified structure-function relationship. Many biological units have been optimized through evolution and the presence of certain amino acids rather than others is fundamental for functionality [61–63]. In photosynthesis, one of the most well-characterized pigment-protein complexes is the Fenna-Matthews-Olson (FMO) complex which is a light-harvesting complex found in green sulphur bacteria. It functions as an intermediate conductor for exciton transport located between the antenna complex where light is initially absorbed and the reaction center. Since the resolution of its crystal structure over 30 years ago [12], the FMO trimer, composed of 3 units each comprising 8 bacteriochlorophylls has been extensively studied both experimentally [15–18] and theoretically [13, 14]. For instance regarding the structure-function relationship, it has been shown [93] that amino acid residues cause considerable shifts in the site energies of bacteriochlorophyll *a* (BChl) molecules of the FMO complex and in turn causes changes to the energy transfer properties.

Have photosynthetic systems adopted interesting quantum effects to improve their efficiency in the course of evolution, as suggested by the experiments? In this article, we provide a first step to answer this question by characterizing the protein

environment of the FMO photosynthetic system to identify the microscopic origin of the long-lived quantum coherence. We investigate the quantum energy transfer of a molecular excitation (exciton) by incorporating an all-atom molecular dynamics (MD) simulation. The molecular energies are computed with time-dependent density functional theory (TDDFT) along the MD trajectory. The evolution of the excitonic density matrix is obtained as a statistical ensemble of unitary evolutions by a time-dependent Schrödinger equation. Thus, this work is in contrast to many studies based on quantum master equations in that it includes atomistic detail of the protein environment into the dynamical description of the exciton. We also introduce a novel approach to add quantum corrections to the dynamics. Furthermore, a quantitative comparison to the hierarchical equation of motion and the Haken-Strobl-Reineker method is presented. As the main result, the time evolution of coherences and populations shows characteristic beatings on the time scale of the experiments. Surprisingly, we observe that the cross-correlation of site energies does not play a significant role in the energy transfer dynamics.

The paper is structured as follows: In the first part we present the methods employed and in the second part the results followed by conclusions. In particular, the partitioning of the system and bath Hamiltonian in classical and quantum degrees of freedom and details of the MD simulations and calculation of site energies are discussed in Section 3.2.1. The exciton dynamics of the system under the bath fluctuations is then presented in Section 3.2.2. In Section 3.2.3 we introduce a quantum correction to the previous exciton dynamics. Using the discussed methods we evaluated site energies and their distribution at 77 and 300K in Section 3.3.1 and we also

computed the linear absorption spectrum of the FMO complex in Section 3.3.3. The site basis dephasing rates are discussed in Section 3.3.2. From the exciton dynamics of the system we obtained populations and coherences and compared to the QJC-MD approach in Section 3.3.4. We then compare the MD and quantum corrected MD methods to the hierarchical equation of motion (HEOM) and Haken-Strobl-Reineker (HSR) methods in Section 3.3.5. In Section 3.3.6 we determined the spectral density for each site from the energy time bath-correlator and studied the effect of auto and cross-correlations on the exciton dynamics by introducing a comparison to first-order autoregressive processes. We conclude in Section 3.4 by summarizing our results.

3.2 Methods

3.2.1 Molecular Dynamics Simulations

A computer simulation of the quantum evolution of the entire FMO complex is certainly unfeasible with the currently available computational resources. However, we are only interested in the electronic energy transfer dynamics among BChl molecules embedded in the protein support. This suggests a decomposition of the total system Hamiltonian operator into three parts: the relevant system, the bath of vibrational modes, and the system-bath interaction Hamiltonians. The system Hamiltonian operates on the excitonic system alone which is defined by a set of two-level systems. Each two-level system represents the ground and first excited electronic state of a BChl molecule. In addition, the quantum mechanical state of the exciton is assumed to be restricted to the single-exciton manifold because the exciton density is low.

On the other hand, factors affecting the system site energies have intractably large degrees of freedom, so it is reasonable to treat all those degrees of freedom as the bath of an open quantum system.

More formally, to describe the system-bath interplay by including atomistic detail of the bath, we start from the total Hamiltonian operator and decompose it in a general way such that no assumptions on the functional form of the system-bath Hamiltonian are necessary [67]:

$$\begin{aligned}\hat{H}_{total} = & \sum_m \int d\mathbf{R} \epsilon_m(\mathbf{R}) |m\rangle\langle m| \otimes |\mathbf{R}\rangle\langle \mathbf{R}| \\ & + \sum_{m,n} \int d\mathbf{R} \{J_{mn}(\mathbf{R}) |m\rangle\langle n| \otimes |\mathbf{R}\rangle\langle \mathbf{R}| + c.c.\} \\ & + |\mathbf{1}\rangle\langle \mathbf{1}| \otimes \hat{T}_{\mathbf{R}} + \sum_m \int d\mathbf{R} V_m(\mathbf{R}) |m\rangle\langle m| \otimes |\mathbf{R}\rangle\langle \mathbf{R}|.\end{aligned}\quad (3.1)$$

Here, \mathbf{R} corresponds to the nuclear coordinates of the FMO complex including both BChl molecules, protein, and enclosing water molecules. The set of states $|m\rangle \otimes |\mathbf{R}\rangle$ denote the presences of the exciton at site m given that the FMO complex is in the configuration \mathbf{R} , $\epsilon_m(\mathbf{R})$ represents the site energy of the m th site and $J_{mn}(\mathbf{R})$ is the coupling constant between the m th and n th sites. Note that the site energies and coupling terms can be modulated by \mathbf{R} . $|\mathbf{1}\rangle\langle \mathbf{1}|$ is the identity operator in the excitonic subspace, $\hat{T}_{\mathbf{R}}$ is the kinetic operator for the nuclear coordinates of the FMO complex, and $V_m(\mathbf{R})$ is the potential energy surface for the complex when the exciton at site m under Born-Oppenheimer approximation. Given multiple Born-Oppenheimer surfaces, one would need to carry out a coupled nonadiabatic propagation. However, as a first approximation, we assume that the change of Born-Oppenheimer surfaces does not affect the bath dynamics significantly. This approximation becomes better

at small reorganization energies. Indeed, BChl molecules have significantly smaller reorganization energies than other chromophores [94]. With this assumption, we can ignore the dependence on the excitonic state in the V term, thus the system-bath Hamiltonian only contains the one-way influence from the bath to the system. We also adopted Condon approximation so that the J terms do not depend on \mathbf{R} :

$$\begin{aligned}
 H_S &= \sum_m \int d\mathbf{R} \bar{\epsilon}_m |m\rangle\langle m| \otimes |\mathbf{R}\rangle\langle \mathbf{R}| + \sum_{m,n} \int d\mathbf{R} \{J_{mn}(\mathbf{R}) |m\rangle\langle n| \otimes |\mathbf{R}\rangle\langle \mathbf{R}| + c.c.\}, \\
 &\approx \sum_m \int d\mathbf{R} \bar{\epsilon}_m |m\rangle\langle m| \otimes |\mathbf{R}\rangle\langle \mathbf{R}| + \sum_{m,n} \int d\mathbf{R} \{\bar{J}_{mn} |m\rangle\langle n| \otimes |\mathbf{R}\rangle\langle \mathbf{R}| + c.c.\}, \\
 H_B &= |\mathbf{1}\rangle\langle \mathbf{1}| \otimes \hat{T}_{\mathbf{R}} + \sum_m \int d\mathbf{R} V_m(\mathbf{R}) |m\rangle\langle m| \otimes |\mathbf{R}\rangle\langle \mathbf{R}|, \\
 &\approx |\mathbf{1}\rangle\langle \mathbf{1}| \otimes \hat{T}_{\mathbf{R}} + \int d\mathbf{R} V_{ground}(\mathbf{R}) |\mathbf{1}\rangle\langle \mathbf{1}| \otimes |\mathbf{R}\rangle\langle \mathbf{R}|, \\
 H_{SB} &= \sum_m \int d\mathbf{R} \{\epsilon_m(\mathbf{R}) - \bar{\epsilon}_m\} |m\rangle\langle m| \otimes |\mathbf{R}\rangle\langle \mathbf{R}|, \\
 H_{total} &= H_S + H_B + H_{SB}.
 \end{aligned} \tag{3.2}$$

Based on this decomposition of the total Hamiltonian, we set up a model of the FMO complex with the AMBER 99 force field [64, 65] and approximate the dynamics of the protein complex bath by classical mechanics. The initial configuration of the MD simulation was taken from the x-ray crystal structure of the FMO complex of *Prosthecochloris aestuarii* (PDB ID: 3EOJ.). Shake constraints were used for all bonds containing hydrogen and the cutoff distance for the long range interaction was chosen to be 12 Å. After a 2ns long equilibration run, the production run was obtained for a total time of 40ps with a 2fs timestep. For the calculation of the optical gap, snapshots were taken every 4fs. Two separate simulations at 77K and 300K were carried out with an isothermal-isobaric (NPT) ensemble to investigate the

temperature dependence of the bath environment. Then, parameters for the system and the system-bath Hamiltonian were calculated using quantum chemistry methods along the trajectory obtained from the MD simulations.

We chose not to include the newly resolved eighth BChl molecule [93] in our simulations because up to now, the large majority of the scientific community has focused on the seven site system which is therefore a better benchmark to compare our calculations to previous work. It is important to note however that this eighth site may have an important role on the dynamics. In particular, as suggested in [95, 96] this eighth site is considered to be the primary entering point for the exciton in the FMO complex and its position dictates a preferential exciton transport pathway rather than two independent ones. Also when starting with an exciton on this eighth site, the oscillations in the coherences are largely suppressed.

The time-dependent site energy ϵ_m was evaluated as the excitation energy of the Q_y transition of the corresponding BChl molecule. We employed the time-dependent density functional theory (TDDFT) with BLYP functional within the Tamm-Dancoff approximation (TDA) using the Q-Chem quantum chemistry package [66]. The basis set was chosen to be 3-21G after considering a trade-off between accuracy and computational cost. The Q_y transition was identified as the excitation with the highest oscillator strength among the first 10 singlet excited states. Then, the transition dipole of the selected state was verified to be parallel to the y molecular axis. Every atom which did not belong to the TDDFT target molecule was incorporated as a classical point charge to generate the external electric field for the QM/MM calculation. Given that the separation between BChl molecules and the protein matrix is

quite clear, employing this simple QM/MM method with classical external charges to calculate the site energies is a good approximation. The external charges were taken from the partial charges of the AMBER force field [64, 65]. The coupling terms, J_{mn} , can also be obtained from quantum chemical approaches like transition density cube or fragment-excitation difference methods [97, 98]. However, in this case we employed the MEAD values of the couplings of the Hamiltonian presented in the literature [14] and considered them to be constant in time. $\bar{\epsilon}_m$ was straightforwardly chosen as the time averaged site energy for the m th site.

3.2.2 Exciton Dynamics

In this section, we describe the method for the dynamics of the excitonic reduced density matrix within our molecular dynamic simulation framework. It is based on a simplified version of the quantum-classical hybrid method (Ehrenfest) described in [67]. The additional assumption on Hamiltonian (3.2) is that the bath coordinate \mathbf{R} is a classical variable, denoted by a superscript “cl”. As discussed above, the time-dependence of these variables arises from the Newtonian MD simulations. The additional force on the nuclei due to the electron-phonon coupling [67] is neglected. Hence, the Schrödinger equation for the excitonic system is given by:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle \approx \{H_S + H_{SB}(\mathbf{R}^{cl}(t))\} |\psi(t)\rangle. \quad (3.3)$$

The system-environment coupling leads to an effective time-dependent Hamiltonian $H_{eff}(t) = H_S + H_{SB}(\mathbf{R}^{cl}(t))$. This equation suggests a way to propagate the reduced density matrix as an average of unitary evolutions given by Eq. (3.3). First, short MD trajectories (in our case 1 ps long) are uniformly sampled from the full MD trajectory

(40 ps). Then, for each short MD trajectory, the excitonic system can be propagated under unitary evolution with a simple time-discretized exponential integrator. The density matrix is the classical average of these unitary evolutions:

$$\rho_S(t) = \frac{1}{M} \sum_{i=1}^M |\psi_i(t)\rangle \langle \psi_i(t)|, \quad (3.4)$$

where M is the number of sample short trajectories. Each trajectory is subject to different time-dependent fluctuations from the bath, which manifests itself as decoherence when averaged to the statistical ensemble. Compared to many methods based on the stochastic unraveling of the master equation, e.g. [59, 99], our formalism directly utilizes the fluctuations generated by the MD simulation. Therefore, the detailed interaction between system and bath is captured. The temperature of the bath is set by the thermostat of the MD simulation, thus no further explicit temperature dependence is required in the overall dynamics. The dynamics obtained by this numerical integration of the Schrödinger equation will also be compared to the HEOM approach. The HEOM is briefly described in the Supporting Material along with a discussion on the differences respect to the MD-method.

3.2.3 Quantum Jump Correction to MD Method (QJC-MD)

The MD/TDDFT simulation above leads to crucial insights into the exciton dynamics. However, it does not capture quantum properties of the vibrational environment such as zero-point fluctuations. At zero temperature all the atoms in the MD simulation are completely frozen. Moreover, similarly to an infinite-temperature model, at long times of the quantum dynamical simulation the exciton is evenly distributed among all molecules, as we will see below. In order to obtain a more

realistic description, we modify the stochastic simulation by introducing quantum jumps derived from the zero-point (zp) fluctuations of the modes in the vibrational environment. We refer to this corrected version of the MD propagation as QJC-MD.

Introducing harmonic bath modes explicitly we reformulate the system-bath Hamiltonian as:

$$H_{SB} = \sum_m |m\rangle\langle m| \sum_{\xi} g_{\xi}^m R_{\xi}. \quad (3.5)$$

Here, each g_{ξ}^m represents the coupling strength of a site m to a particular mode ξ and R_{ξ} is the dimensionless position operator for that mode. We now formulate our correction by separating the bath operators into two parts, $R_{\xi} = R_{\xi}^{zp} + R_{\xi}^{MD}$, the first part is due to zero-point fluctuations and the second comes from our MD simulations. As above, the MD part is replaced by the classical time-dependent variables, $R_{\xi}^{MD} \rightarrow R_{\xi}^{cl}(t)$. The zero-point operator is expressed by creation and annihilation operators, $R_{\xi}^{zp} = b_{\xi}^{zp} + b_{\xi}^{zp,\dagger}$, which satisfy the usual commutation relations $[b_{\xi}^{zp}, b_{\xi'}^{zp,\dagger}] = \delta_{\xi\xi'}$. By construction, for the zp-fluctuations one has $\langle b_{\xi}^{zp,\dagger} b_{\xi}^{zp} \rangle = 0$.

The zp-fluctuations can only induce excitonic transitions from higher to lower exciton states in the instantaneous eigenbasis of the Hamiltonian, thus leading to relaxation of the excitonic system. The evolution of the populations P_M of the instantaneous eigenstates $|M\rangle(t)$ due to the zero-point correction is expressed by a Pauli master equation as:

$$\left(\dot{P}_M \right)_{zpc} = - \sum_N \gamma(\omega_{MN}) P_M + \sum_N \gamma(\omega_{NM}) P_N, \quad (3.6)$$

and for the coherences as:

$$\left(\dot{C}_{MN} \right)_{zpc} = -\frac{1}{2} \gamma(|\omega_{MN}|) C_{MN}. \quad (3.7)$$

The associated rate can be derived from a secular Markovian Redfield theory [34] to be $\gamma(\omega_{MN}) = 2\pi J(\omega_{MN}) \sum_m |c_m(M)|^2 |c_m(N)|^2$, where the spectral density $J(\omega)$ is only non-zero for positive transition frequencies $\omega_{MN} = E_M - E_N$ and taken to be as in [87]. The coefficients $c_m(M)$ translate from site to energy basis. The time evolution given by Equations (3.6) and (3.7) is included in the dynamics simulation by introducing quantum jumps as in the Monte-Carlo wavefunction (MCWF) method [99]. We thus arrive at a hybrid classically averaged $H(t)$ simulation with additional quantum transitions induced by the vacuum fluctuations of the vibrational modes.

3.3 Results and Discussion

3.3.1 Site Energy Distributions

Using the coupled QM/MD simulations, site energies were obtained for each BChl molecule. These energies and their fluctuations are reported in Figure 3.1. We note that the magnitude of the fluctuations are of the order of hundreds of cm^{-1} . Although the order of the site energies does not perfectly match previously reported results [14, 20], the overall trend does not deviate much, especially considering that our result is purely based on *ab initio* calculations without fitting to the experimental result. The Q_y transition energies calculated by TDDFT are known to be systematically blue-shifted with respect to the experiment [100]. However, the scale of the fluctuations remains reasonable. Therefore, the comparison in Fig. 3.1 was made after shifting the overall mean energy to zero for each method.

The excitation energy using TDDFT does not always converge when the config-

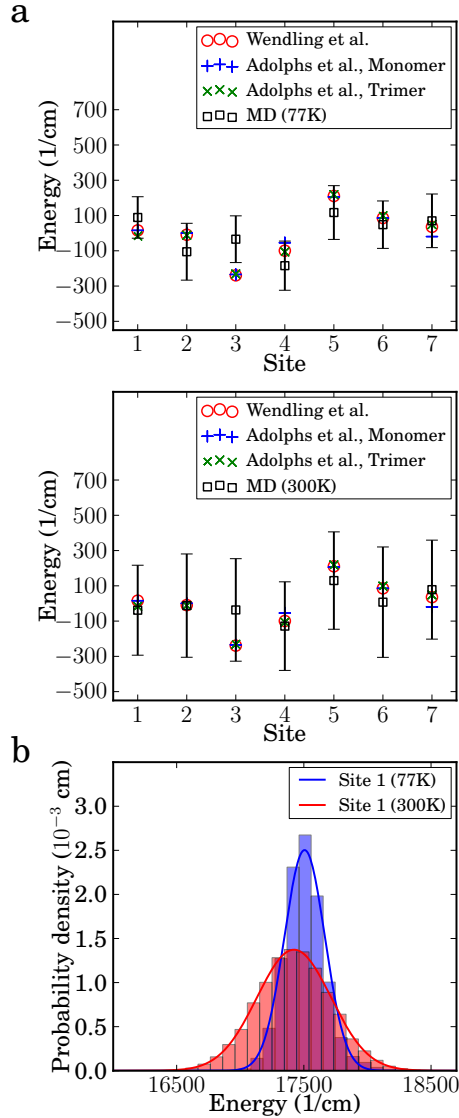


Figure 3.1: Panel **a**: Comparison of the calculated site energies for each BChl molecule to the previous works by Wendling et al. and Adolphs et al. [14, 20]. Our calculation, labeled as MD, was obtained using QM/MM calculations with the TDDFT/TDA method at 77K and 300K. Vertical bars represent the standard deviation for each site. Panel **b**: Marginal distribution of site 1 energy at 77K and 300K. Histograms represent the original data, and solid lines correspond to the estimated Gaussian distribution.

uration of the molecule deviates significantly from its ground state structure. The number of points which failed to converge was on average less than 4% for configurations at 300K, and less than 2% at 77K. We interpolated the original time series to obtain smaller time steps and recover the missing points. Interpolation could lead to severe distortion of the marginal distribution when the number of available points is too small. However, in our case, the distributions virtually remained the same with and without interpolation.

3.3.2 Dephasing Rates

In the Markovian approximation and assuming an exponentially decaying autocorrelation function, the dephasing rate γ_ϕ is proportional to the variance of the site energy σ_ϵ^2 [34]:

$$\gamma_\phi = \frac{2}{\hbar} \sigma_\epsilon^2 \tau, \quad (3.8)$$

where τ is a time decay parameter which we estimated through a comparison to first order autoregressive processes, as described in Section 3.3.6. The dependence on the variance is clearly justified: states associated with large site energy fluctuations tend to undergo faster dephasing. Figure 3.2, panel a), presents the approximate site basis dephasing rates for each site with $\tau \approx 5$ fs. The averaged value of the slopes is $0.485 \text{ cm}^{-1} \text{ K}^{-1}$, which is in good agreement with the experimentally measured value of $0.52 \text{ cm}^{-1} \text{ K}^{-1}$ obtained from a closely related species *Chlorobium tepidum* in the exciton basis [25]. From this plot we note the presence of a positive correlation between temperature and dephasing rate. This correlation is plausible: as temperature increases so does the energy disorder, hence the coherences should decay faster. In

fact, in the Markovian approximation, dephasing rates increase linearly with temperature [34, 101]. Calculations at other temperatures are underway to verify this and to obtain more information on the precise temperature dependence of the dephasing rates.

3.3.3 Simulated Spectra

The absorption, linear dichroism (LD), and circular dichroism (CD) spectra can be obtained from the Fourier transform of the corresponding response functions. The spectra can be evaluated for the seven BChl molecules using the following expressions [102, 103]:

$$\begin{aligned}
 I_{Abs}(\omega) &\propto \text{Re} \int_0^\infty dt e^{i\omega t} \sum_{m,n=1}^7 \langle \vec{d}_m \cdot \vec{d}_n \rangle \{ \langle U_{mn}(t,0) \rangle - \langle U_{mn}^*(t,0) \rangle \}, \\
 I_{LD}(\omega) &\propto \text{Re} \int_0^\infty dt e^{i\omega t} \sum_{m,n=1}^7 \langle 3(\vec{d}_m \cdot \hat{r})(\vec{d}_n \cdot \hat{r}) - \vec{d}_m \cdot \vec{d}_n \rangle \{ \langle U_{mn}(t,0) \rangle - \langle U_{mn}^*(t,0) \rangle \}, \\
 I_{CD}(\omega) &\propto \text{Re} \int_0^\infty dt e^{i\omega t} \sum_{m,n=1}^7 \langle \vec{\epsilon}_m(\vec{R}_m - \vec{R}_n) \cdot (\vec{d}_m \times \vec{d}_n) \rangle \{ \langle U_{mn}(t,0) \rangle - \langle U_{mn}^*(t,0) \rangle \},
 \end{aligned} \tag{3.9}$$

where m and n are indices for the BChl molecules in the complex, \vec{d}_m is the transition dipole moment of the m th site, $U_{mn}(t,0)$ is the (m,n) element of the propagator in the site basis, \hat{r} is the unit vector in the direction of the rotational symmetry axis, \vec{R}_m is the coordinate vector of the site m , and $\langle \cdots \rangle$ indicates an ensemble average. The ensemble average was evaluated by sampling and averaging over 4000 trajectories. We applied a low-pass filter to smooth out the noise originated from truncating the integration and due to the finite number of trajectories. Figure 3.2 panel b) and

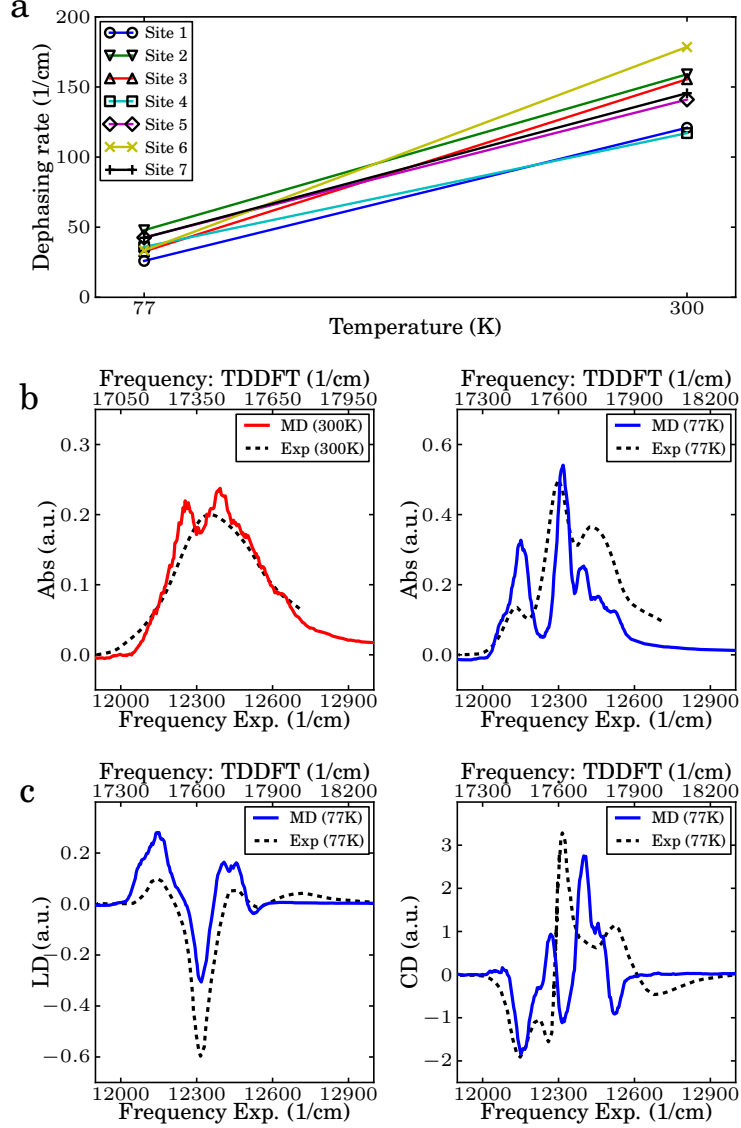


Figure 3.2: Panel **a** shows the calculated dephasing rate for each site at 77K and 300K. Panel **b** shows the simulated linear absorption spectra of the FMO complex at 77K and 300K. They were shifted to be compared to the experimental spectra as obtained by Engel through personal communication. Panel **c** shows the simulated linear dichroism (LD) and circular dichroism (CD) spectra at 77K. Experimental spectra were obtained from Wendling et al.[20] Although TDDFT-calculated spectra shows systematically overestimated site energies, the width and overall shape of the spectra is in good agreement.

c) show direct comparison of the calculated and experimental spectra at 77K and 300K. As discussed in Section 3.3.1, TDDFT tends to systematically overestimate the excitation energy of the Q_y transition [104] yet the fluctuation widths of the site energies are reasonable. In fact, the width and overall shape of the calculated spectrum is in good agreement with the experimental spectrum at each temperature. Calculated LD and CD spectra also reproduce well the experimental measurements, considering that no calibration to experiments was carried out. Since both LD and CD spectra are sensitive to the molecular structure it appears that our microscopic model correctly captures these details.

3.3.4 Population Dynamics and Long-lived Quantum Coherence

The MD method is based on minimal assumptions and directly evaluates the dynamics of the reduced density matrix from the total density matrix as described in Section 3.2. The reduced density matrix was obtained after averaging over 4000 trajectories. Figure 3.3 shows the population and coherence dynamics of each of the seven sites according to the dephasing induced by the nuclear motion of the FMO complex. In particular, the populations and the absolute value of the pairwise coherences, as defined in [52] ($2 \cdot |\rho_{12}(t)|$ and $2 \cdot |\rho_{56}(t)|$) are plotted at both 77 and 300K starting with an initial state in site 1 (first three panels) and then in site 6 (last three panels). Until very recently [95, 96] site 1 and 6 have been thought as the entry point of an exciton in the FMO complex, therefore most of the previous literature chose the initial reduced density matrix to be either pure states $|1\rangle\langle 1|$ or

$|6\rangle\langle 6|$ [29, 51, 105]. However, our method could be applied to any mixed initial state without modification. We note that coherent beatings last for about 400fs at 77K and 200fs at 300K. These timescales are in agreement with those reported for FMO [25, 29] and with what was found in Section 3.3.2. Although quite accurate in the short time limit, the MD method populations do not reach thermal equilibrium at long times. This was verified by propagating the dynamics to twice the time shown in Figure 3.3. This final classical equal distribution is similar to the HSR model result. The three central panels of Figure 3.3 show the same populations and coherences obtained from the QJC-MD method. As discussed in Section 3.2, this method includes a zero point correction through relaxation transitions and predicts a more realistic thermal distribution at 77K. At 300K the quantum correction is less important in the dynamics because the Hamiltonian fluctuations dominate over the zero temperature quantum fluctuations.

3.3.5 Comparison between MD, QJC-MD, HEOM, and HSR Methods

Figure 3.4 shows a direct comparison of the population dynamics of site 1 calculated using the HEOM method discussed by Ishizaki and Fleming [29, 58], our MD and quantum corrected methods at 77K and 300K, and the HSR model [39, 40] with dephasing rates obtained from Eq. (3.8). We observe that the short-time dynamics and dephasing characteristics are surprisingly similar, considering that the methods originate from very different assumptions. Atomistic detail can allow for differentiation of the system-environment coupling for different chromophores. For example, at

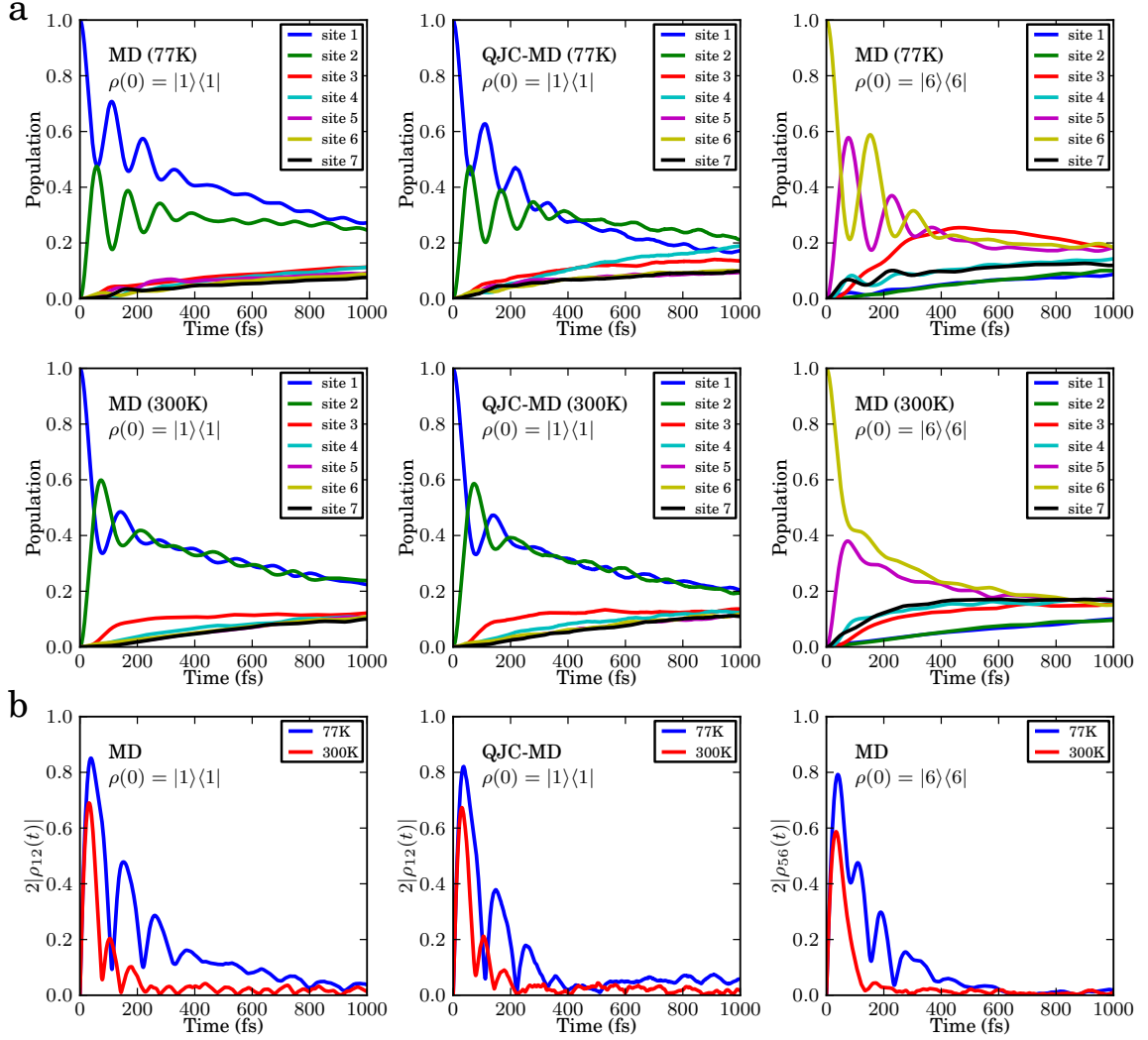


Figure 3.3: Panel **a**: Time evolution of the exciton population of each chromophore in the FMO complex at 77K and 300K. Panel **b**: Change of the pairwise coherence, or concurrence in time. Initial pure states, $\rho_S(0) = |1\rangle\langle 1|$ for the top and center panels were propagated using the two formulations developed in this article, MD and QJC-MD, to utilize the atomistic model of the protein complex bath from the MD/TDDFT calculation. Panel **c** The initial state was set to $|6\rangle\langle 6|$ and propagated using the MD method.

both temperatures (right panels), the MD populations of site 6 undergo faster decoherence than the corresponding HEOM results. We attribute this to the difference in energy gap fluctuations of site energy between site 1 and 6 obtained from the MD simulation as can be seen in 3.1. On one hand, in the HEOM method, site energy fluctuations are considered to be identical across all sites, on the other, in our method the fluctuations of each site are obtained from the MD simulation in which each site is associated with a different chromophore-protein coupling. Nevertheless, the fact that we obtain qualitatively similar results to the HEOM approach (at least when starting in $\rho(0) = |1\rangle\langle 1|$) without considering non equilibrium reorganization processes suggests that such processes might not be dominant in the FMO. The quantum correction results (QJC-MD), for every temperature and initial state, are in between the HEOM and MD results. This is due to the induced relaxation from zero-point fluctuations of the bath environment, which are not included in the MD method but included in the QJC-MD and HEOM methods.

The HSR results take into account the site-dependence of the dephasing rates based on Eq. (3.8). The method is briefly described in the supplementary material. Due to the Markovian assumption, this model shows slightly less coherence than the HEOM method and similarly to the MD method it converges to an equal classical mixture of all sites in the long time limit.

3.3.6 Correlation Functions and Spectral Density

The bath autocorrelation function and its spectral density contain information on interactions between the excitonic system and the bath. The bath correlation

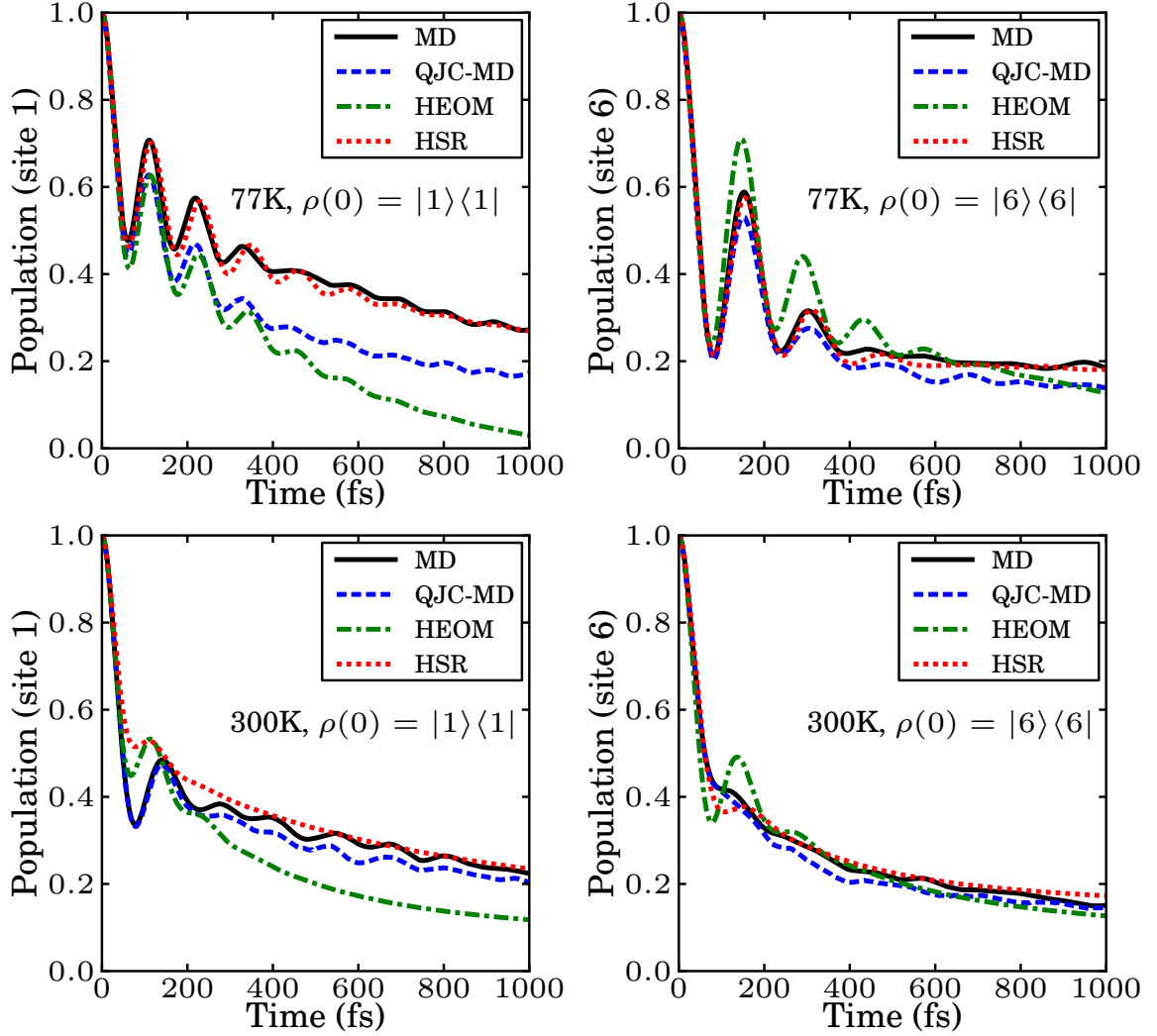


Figure 3.4: Comparison of the population dynamics obtained by using the MD method, the corrected MD, the hierarchy equation of motion approach and the Haken-Strobl-Reineker model at 77K and 300K. Panels on the right correspond to the initial state in site 1 and those on the left to an initial state in site 6. All methods show similar short-time dynamics and dephasing, while the long time dynamics is different and the different increases as relaxation is incorporated in the various methods.

function is defined as $C(t) = \langle \delta\epsilon(t)\delta\epsilon(0) \rangle$ with $\delta\epsilon = \epsilon(t) - \bar{\epsilon}$. For the MD simulation, $C(t)$ is shown in Fig. 3.5 a) for the two temperatures.

To study the effect of the decay rate of the autocorrelation function on the population dynamics, we modeled site energies using first-order autoregressive (AR(1)) processes [106]. The marginal distribution of each process was tuned to have the same mean and variance as for the MD simulation. The autocorrelation function of the AR(1) process is an exponentially decaying function:

$$C(t) \propto \exp(-t/\tau). \quad (3.10)$$

We generated three AR(1) processes with different time constants τ and propagated the reduced density matrix using the Hamiltonian corresponding to each process. As can be seen in Fig. 3.5, panel a), the autocorrelation function of the AR(1) process with $\tau \approx 5\text{fs}$ has a similar initial decay rate to that of the MD simulation at both temperatures. Therefore, as shown in the last three horizontal panels, its spectral density is in good agreement with the MD simulation result in the low frequency region, i.e up to 600cm^{-1} . Modes in this region are known to be the most important in the dynamics and in determining the decoherence rate. Also, as panels b) and c) show, that same AR(1) process with $\tau \approx 5\text{fs}$ exhibits similar population beatings and concurrences to those of the MD simulation. The relation of this 5fs time scale to others reported in [21, 29] is presently unclear. We suspect that the discrepancy between the two results should decrease when one propagates the MD in the excited state. Work in this direction is in progress in our group.

The spectral density can be evaluated as the reweighted cosine transform of the

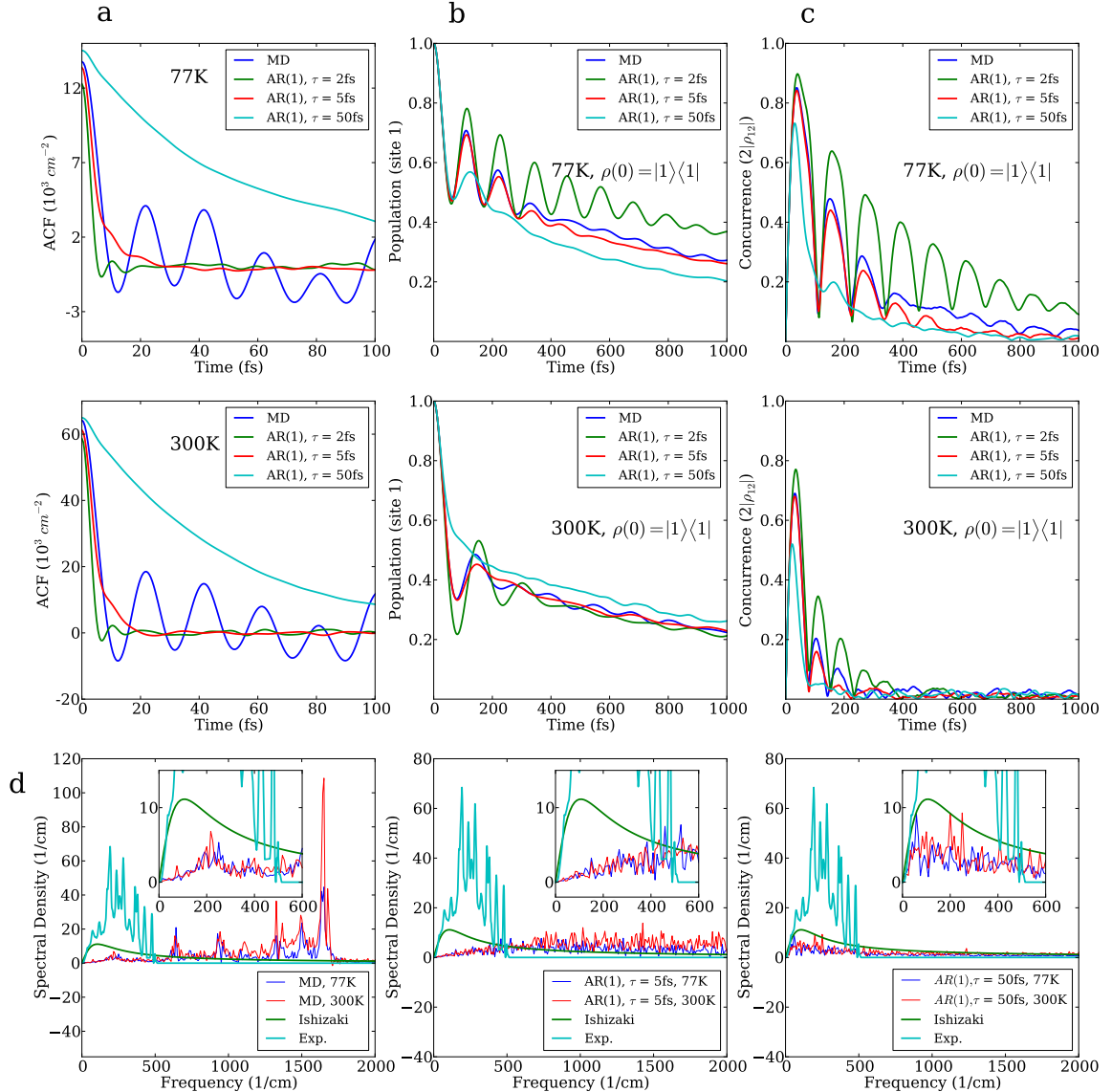


Figure 3.5: Panel **a**: Site 1 autocorrelation functions using MD and AR(1) processes generated with time constant equal to 2fs, 5fs, and 50fs at 77K and 300K. Panel **b**: Site 1 population dynamics of MD and AR(1) processes with the different time constants at 77K and 300K. Panel **c**: The change of pairwise coherence between site 1 and 2 of MD and AR(1) processes with the different time constants at 77K and 300K. Panel **d**: Spectral density of site 1 of the FMO complex from the MD simulation at 77K and 300K. They clearly show the characteristic vibrational modes of the FMO complex. High-frequency modes are overpopulated due to the ultraviolet catastrophe observed in classical mechanics. The Ohmic spectral density used by Ishizaki and Fleming in [29] was presented for comparison. The spectral densities of site 1 from AR(1) processes are also presented.

corresponding bath autocorrelation function $C(t)$ [103, 104],

$$J(\omega) = \frac{2}{\pi\hbar} \tanh(\beta\hbar\omega/2) \int_0^\infty C(t) \cos(\omega t) dt. \quad (3.11)$$

With the present data the spectral density exhibits characteristic phonon modes from the dynamics of the FMO complex, see Fig. 3.5 d) first panel. However, high-frequency modes tend to be overpopulated due to the limitation of using classical mechanics. Most of these modes are the local modes of the pigments, which can be seen from the pigment-only calculation in [65]. There are efforts to incorporate quantum effects into the classical MD simulation in the context of vibrational coherence [68–70]. We are investigating the possibilities of incorporating corrections based on a similar approach. Moreover, we also obtain a discrepancy of the spectral density in the low frequency region. On one hand, the origin could lie in the harmonic approximation of the bath modes leading to the tanh prefactor in Eq. (3.11) or in the force field used in this work. On the other, the form of the standard spectral density is from [17] which measures fluorescence line-narrowing on a much longer timescale, around ns, than considered in our simulations (around ps). Assuming correctness of our result, this implies that for the simulation of fast exciton dynamics in photosynthetic light-harvesting complexes a different spectral density than the widely used one has to be employed.

Site energy cross-correlations between chromophores due to the protein environment have been postulated to contribute to the long-lived coherence in photosynthetic systems [24]. Many studies have explored this issue, e.g. recently [14, 51, 60, 107–109]. We tested this argument by de-correlating the site energies. For each unitary evolution, the site energies of different molecules at the same time were taken from

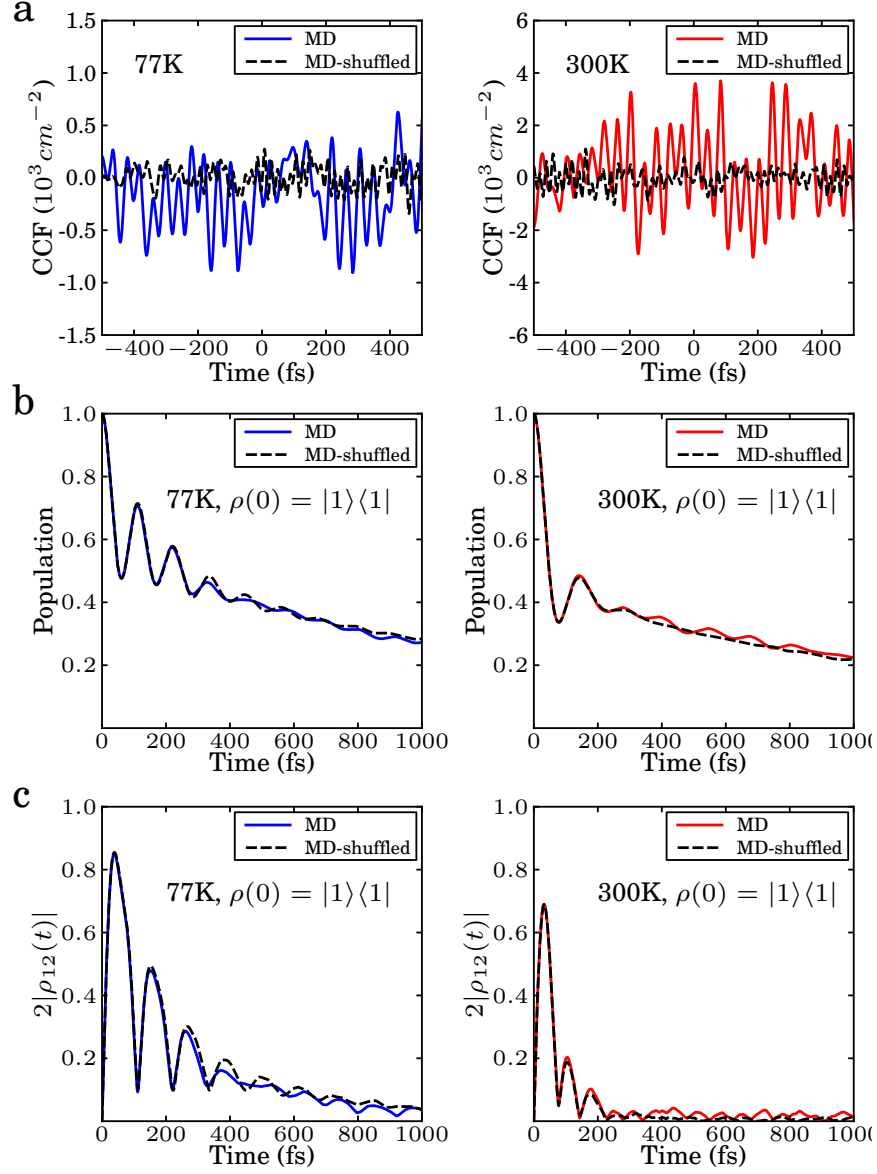


Figure 3.6: Panel **a**: Cross-correlation function of the original MD trajectory and a randomly shuffled trajectory between sites 1 and 2 at 77K and 300K. Panel **b**: Site 1 population dynamics of the original dynamics and the shuffled dynamics at 77K and 300K. **c**, The pairwise coherence between sites 1 and 2. Original and shuffled dynamics are virtually identical at both temperatures.

different parts of the MD trajectory. In this way, we could significantly reduce potential cross correlation between sites while maintaining the autocorrelation function of each site. As can be seen in Fig. 3.6, no noticeable difference between the original and shuffled dynamics is observed.

3.4 Conclusion

The theoretical and computational studies presented in this article show that the long-lived quantum coherence in the energy transfer process of the FMO complex of *Prosthecochloris aestuarii* can be simulated with the atomistic model of the protein-chromophore complex. Unlike traditional master equation approaches, we propagate in a quantum/classical framework both the system and the environment state to establish the connection between the atomistic details of the protein complex and the exciton transfer dynamics. Our method combines MD simulations and QM/MM with TDDFT/TDA to produce the time evolution of the excitonic reduced density matrix as an ensemble average of unitary trajectories.

The conventional assumption of unstructured and uncorrelated site energy fluctuations is not necessary for our method. No *ad hoc* parameters were introduced in our formalism. The temperature and decoherence time were extracted from the site energy fluctuation by the MD simulation of the protein complex. The simulated dynamics clearly shows the characteristic quantum wave-like population change and the long-lived quantum coherence during the energy transfer process in the biological environment. On this note it is worth mentioning that one has to be careful in the choice of force-field and in the method used to calculate site energies. In fact as

presented in Olbrich et al. [110] a completely different energy transfer dynamics was obtained by using the semiempirical ZINDO-S/CIS to determine site energies.

Moreover, we determined the correlations of the site energy fluctuations for each site and between sites through the direct simulation of the protein complex. The spectral density shows the influence of the characteristic vibrational frequencies of the FMO complex. This spectral density can be used as an input for quantum master equations or other many-body approaches to study the effect of the structured bath. The calculated linear absorption spectrum we obtained is comparable to the experimental result, which supports the validity of our method. The characteristic beating of exciton population and pairwise quantum coherence exhibit excellent agreement with the results obtained by the HEOM method. It is also worth noting the remarkable agreement of the dephasing timescales of the MD simulations, the HEOM approach, and experiment.

Recently, characterization of the bath in the LH2 [103, 104] and FMO [109] photosynthetic complexes were reported using MD simulation and quantum chemistry at room temperature. Those studies mostly focused on energy and spatial correlations across the sites, the linear absorption spectrum, and spectral density. The detailed study in [109] also suggests that spatial correlations are not relevant in the FMO dynamics.

This work opens the road to understanding whether biological systems employed quantum mechanics to enhance their functionality during evolution. We are planning to investigate the effects of various factors on the photosynthetic energy transfer process. These include: mutation of the protein residues, different chromophore

molecules, and temperature dependence. Further research in this direction could elucidate on the design principle of the biological photosynthesis process by nature, and could be beneficial for the discovery of more efficient photovoltaic materials and in biomimetics research.

Chapter 4

Path integral Monte Carlo with importance sampling for excitons interacting with arbitrary phonon bath environment

4.1 Introduction

Recent 2D non-linear spectroscopy experiments suggested the existence of long-lived quantum coherence during the electronic energy transfer process within the Fenna-Matthews-Olson complex of green sulfur bacteria, marine algae and plants even under physiological conditions [23, 25, 47, 111–113]. These results attracted a large amount of attention from theoretical physicists and chemists. The energy transfer process usually has been modeled as the dynamics of excitons coupled to a

phonon bath in thermal equilibrium within the single exciton manifold. This approximation leads to the famous spin-Boson Hamiltonian. The solution of this type of Hamiltonian has been studied extensively. For example, by assuming a certain relative magnitude between the reorganization energy and coupling terms, one can obtain quantum master equations valid in specific regimes[19, 67, 114]. Another approximation, the Haken-Strobl-Reineker model works in both the coherent and incoherent regimes, but incorrectly converges to the high temperature limit in the long time even at the low temperature [39, 40]. More recently, numerically exact approaches which interpolate both limits have been investigated and applied to many systems of interest. Two of the most popular methods are the hierarchical equation of motion [29, 87, 115] and the quasiadiabatic path integral method [116, 117]. These methods are being actively developed, improved, and applied to many systems of interests [58].

Although having been successful in many applications, many of the models described above have assumed the phonon bath to be a set of independent harmonic oscillators and encode all the complexity of the bath environment in the spectral density, which is essentially a frequency dependent distribution of exciton-phonon coupling. However, for studying the anharmonic effects of a very sophisticated bath environment, like the protein complexes of photosynthesis, being able to directly include the atomistic details of the bath structure into the exciton dynamics has a distinct advantage. In other words, approaches that can evaluate the influence functional first suggested by Feynman and Vernon [118] have more straightforward descriptions and are applicable to arbitrary systems. Evaluation of the exact influence functional for arbitrary environment requires the simulation of the full quantum dynamics, which is

still not practical with currently available computational resources. There have been several attempts to incorporate atomistic details of the large scale bath by combining the exciton dynamics and molecular dynamics simulations [32, 104, 119]. However, these theories are still in their early stages and the propagation scheme used does not satisfy some fundamental properties, like the detailed balance condition at finite temperature. In pursuit of more accurate theory, it is crucial to know the correct asymptotic behavior in the limit of infinite time. In this context, we decided to explore the numerically exact reduced density matrix in a finite temperature using path integral Monte Carlo [120–123] method. Recently, Moix *et al* applied path integral Monte Carlo for the equilibrium reduced density matrix of the FMO complex within the framework of open quantum systems [124].

4.2 Theory

4.2.1 Path Integral Formulation of the Reduced Thermal Density Matrix

We want to evaluate the reduced density matrix of an excitonic system coupled to phonons on arbitrary Born-Oppenheimer surfaces at a finite temperature. For photosynthetic energy transfer, we usually restrict the excitons to be within the single exciton manifold because at normal light intensity, in average, one photon is present at a given time in the complexes of interest. However, the formulation itself is not limited to the single exciton manifold. The Hamiltonian operator for such a system

can be written as

$$\begin{aligned} \hat{H} = & \underbrace{\sum_m \int d\mathbf{R} [V_m(\mathbf{R}) - V_g(\mathbf{R})] |m\rangle\langle m| \otimes |\mathbf{R}\rangle\langle \mathbf{R}| + \sum_{m \neq n} \int d\mathbf{R} J_{mn}(\mathbf{R}) |m\rangle\langle n| \otimes |\mathbf{R}\rangle\langle \mathbf{R}|}_{\hat{H}_{\text{exc}} = \hat{H}_S + \hat{H}_{SB}} \\ & + \underbrace{|\mathbf{1}\rangle\langle \mathbf{1}| \otimes \left[\hat{T} + \int d\mathbf{R} V_g(\mathbf{R}) |\mathbf{R}\rangle\langle \mathbf{R}| \right]}_{\hat{H}_B}. \end{aligned} \quad (4.1)$$

The Hamiltonian was written in terms of the diabatic basis $|m, \mathbf{R}\rangle \equiv |m\rangle \otimes |\mathbf{R}\rangle$, where m is the index for the exciton state and \mathbf{R} is the phonon coordinate. $V_g(\mathbf{R})$ is the potential energy surface (PES) of the phonons in the electronic ground state and $V_m(\mathbf{R})$ is the PES of the phonons in the m th exciton state. \hat{T} is the kinetic operator of the phonons defined as $\hat{T} = -\frac{\hbar^2}{2} \mathcal{M}^{-1} \nabla^2$, where \mathcal{M} is the mass tensor of the phonons. This expression is generally applicable to any molecular system with multiple potential energy surfaces. The reduced thermal density matrix ρ_S is defined as the partial trace of the full thermal density matrix with respect to the bath degrees of freedom:

$$\begin{aligned} \rho_S &= \frac{1}{Z(\beta)} \text{Tr}_B \exp(-\beta \hat{H}) \\ &= \frac{1}{Z(\beta)} \int d\mathbf{R}_0 \langle \mathbf{R}_0 | \exp(-\beta \hat{H}) | \mathbf{R}_0 \rangle, \end{aligned} \quad (4.2)$$

where $Z(\beta)$ is the partition function of the total system. We proceed by relying on the following identity:

$$\begin{aligned}
 \langle \mathbf{R}_0 | \exp(-\beta \hat{H}) | \mathbf{R}_0 \rangle &= \langle \mathbf{R}_0 | \left\{ \exp \left(-\frac{\beta \hat{H}}{M} \right) \right\}^M | \mathbf{R}_0 \rangle \\
 &= \int d\mathbf{R}_1 \int d\mathbf{R}_2 \cdots \int d\mathbf{R}_{M-1} \\
 &\quad \times \langle \mathbf{R}_0 | \exp \left(-\frac{\beta \hat{H}}{M} \right) | \mathbf{R}_{M-1} \rangle \langle \mathbf{R}_{M-1} | \exp \left(-\frac{\beta \hat{H}}{M} \right) | \mathbf{R}_{M-2} \rangle \\
 &\quad \times \cdots \times \langle \mathbf{R}_2 | \exp \left(-\frac{\beta \hat{H}}{M} \right) | \mathbf{R}_1 \rangle \langle \mathbf{R}_1 | \exp \left(-\frac{\beta \hat{H}}{M} \right) | \mathbf{R}_0 \rangle. \quad (4.3)
 \end{aligned}$$

For any positive integer M , the expression above is exact. When the Trotter decomposition is applied, an imaginary timestep $\tau \equiv \frac{\beta \hbar}{M}$ is usually defined for convenience. Then, the thermal density matrix can be interpreted as an imaginary time evolution. In the limit of an infinitesimal imaginary timestep, the Trotter decomposition converges to the exact result,

$$\begin{aligned}
 \langle \mathbf{R}_1 | \exp \left(-\frac{\beta \hat{H}}{M} \right) | \mathbf{R}_0 \rangle &= \langle \mathbf{R}_1 | \exp \left(-\tau \hat{H} / \hbar \right) | \mathbf{R}_0 \rangle \\
 &= \langle \mathbf{R}_1 | e^{-\tau \hat{H}_{\text{exc}} / 2\hbar} e^{-\tau \hat{H}_B / \hbar} e^{-\tau \hat{H}_{\text{exc}} / 2\hbar} | \mathbf{R}_0 \rangle + O(\tau^3) \\
 &= \int d\mathbf{R}_2 \int d\mathbf{R}_3 \langle \mathbf{R}_1 | e^{-\tau \hat{H}_{\text{exc}} / 2\hbar} | \mathbf{R}_3 \rangle \\
 &\quad \times \langle \mathbf{R}_3 | e^{-\tau \hat{H}_B / \hbar} | \mathbf{R}_2 \rangle \langle \mathbf{R}_2 | e^{-\tau \hat{H}_{\text{exc}} / 2\hbar} | \mathbf{R}_0 \rangle + O(\tau^3). \quad (4.4)
 \end{aligned}$$

Subsequently, we will recast the system part of \hat{H}_{exc} as a single matrix to simplify

the notation,

$$\begin{aligned}\hat{H}_{\text{exc}} &= \sum_{m,n} \int d\mathbf{R} E_{mn}(\mathbf{R}) |m\rangle \langle n| \otimes |\mathbf{R}\rangle \langle \mathbf{R}|, \\ E_{mm}(\mathbf{R}) &= \begin{cases} V_m(\mathbf{R}) - V_g(\mathbf{R}) & \text{for } m = n, \\ J_{mn}(\mathbf{R}) & \text{for } m \neq n. \end{cases}\end{aligned}\quad (4.5)$$

With the single exciton manifold assumption, E_{mm} corresponds to the optical gap of the m -th site. Now, the three terms in the integrand of the Eq. 4.4 can be written without Dirac notation,

$$\begin{aligned}\langle \mathbf{R}_1 | e^{-\tau \hat{H}_{\text{exc}}/2\hbar} | \mathbf{R}_3 \rangle &= \delta(\mathbf{R}_1 - \mathbf{R}_3) e^{-\tau E(\mathbf{R}_3)/2\hbar}, \\ \langle \mathbf{R}_3 | e^{-\tau \hat{H}_B/\hbar} | \mathbf{R}_2 \rangle &= (4\pi\tau|\lambda|)^{-1/2} e^{-\tau V_g(\mathbf{R}_3)/2\hbar} e^{-(\mathbf{R}_3 - \mathbf{R}_2)^T \lambda^{-1} (\mathbf{R}_3 - \mathbf{R}_2)/4\tau} e^{-\tau V_g(\mathbf{R}_2)/2\hbar} \\ &\quad + O(\tau^3), \\ \langle \mathbf{R}_2 | e^{-\tau \hat{H}_{\text{exc}}/2\hbar} | \mathbf{R}_0 \rangle &= \delta(\mathbf{R}_2 - \mathbf{R}_0) e^{-\tau E(\mathbf{R}_0)/2\hbar},\end{aligned}\quad (4.6)$$

where $\lambda \equiv \frac{\hbar \mathcal{M}^{-1}}{2}$. By the Eq. 4.4 and Eq. 4.6,

$$\begin{aligned}\langle \mathbf{R}_1 | \exp\left(-\frac{\beta \hat{H}}{M}\right) | \mathbf{R}_0 \rangle &= (4\pi\tau|\lambda|)^{-1/2} e^{-\tau V_g(\mathbf{R}_1)/2\hbar} e^{-(\mathbf{R}_1 - \mathbf{R}_0)^T \lambda^{-1} (\mathbf{R}_1 - \mathbf{R}_0)/4\tau} e^{-\tau V_g(\mathbf{R}_0)/2\hbar} \\ &\quad \times e^{-\tau E(\mathbf{R}_1)/2\hbar} e^{-\tau E(\mathbf{R}_0)/2\hbar} + O(\tau^3).\end{aligned}\quad (4.7)$$

Note that Eq. 4.7 is a matrix with the same dimension as the reduced density matrix

of the system. Substituting Eq. 4.7 to Eq. 4.2, we obtain

$$\begin{aligned}
\rho_S &= \frac{1}{Z(\beta)} \int d\mathbf{R}_0 \int d\mathbf{R}_1 \cdots \int d\mathbf{R}_{M-1} \\
&\times e^{-\tau E(\mathbf{R}_0)/2\hbar} e^{-\tau E(\mathbf{R}_{M-1})/\hbar} \dots e^{-\tau E(\mathbf{R}_1)/\hbar} e^{-\tau E(\mathbf{R}_0)/2\hbar} \\
&\times e^{-\tau V_g(\mathbf{R}_0)/\hbar} e^{-\tau V_g(\mathbf{R}_1)/\hbar} \dots e^{-\tau V_g(\mathbf{R}_{M-1})/\hbar} \\
&\times e^{-(\mathbf{R}_0 - \mathbf{R}_{M-1})^T \lambda^{-1} (\mathbf{R}_0 - \mathbf{R}_{M-1})/4\tau} e^{-(\mathbf{R}_{M-1} - \mathbf{R}_{M-2})^T \lambda^{-1} (\mathbf{R}_{M-1} - \mathbf{R}_{M-2})/4\tau} \\
&\times \dots \times e^{-(\mathbf{R}_1 - \mathbf{R}_0)^T \lambda^{-1} (\mathbf{R}_1 - \mathbf{R}_0)/4\tau} \\
&= \int d\mathbf{R}_0 \int d\mathbf{R}_1 \cdots \int d\mathbf{R}_{M-1} \\
&\times \underbrace{\frac{K}{Z(\beta)} e^{-\tau E(\mathbf{R}_0)/2\hbar} e^{-\tau E(\mathbf{R}_{M-1})/\hbar} \dots e^{-\tau E(\mathbf{R}_1)/\hbar} e^{-\tau E(\mathbf{R}_0)/2\hbar}}_{\rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})} \\
&\times \underbrace{\frac{1}{K} e^{-\beta V_{\text{PIMC}}(\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_{M-1})}}_{f_g(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})}, \tag{4.8}
\end{aligned}$$

where,

$$\begin{aligned}
V_{\text{PIMC}}(\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_{M-1}) &= \frac{1}{M} \sum_{i=0}^{M-1} V_g(\mathbf{R}_i) \\
&+ \sum_{i=0}^{M-1} \frac{M}{2\beta^2 \hbar^2} \{\mathbf{R}_i - \mathbf{R}_{\text{mod}(i+1, M)}\}^T \mathcal{M} \{\mathbf{R}_i - \mathbf{R}_{\text{mod}(i+1, M)}\}. \tag{4.9}
\end{aligned}$$

The expressions above show that the reduced thermal density matrix ρ_S can be evaluated as an expectation value of $\rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})$ where the joint probability density function of the M N -dimensional random variables $(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})$ is f_g . This type of multidimensional integral can be efficiently evaluated using Monte Carlo integration. Because $f_g(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})$ is invariant to cyclic permutation of the phonon coordinate, usually the averaged estimator ρ_{PIMC} over the cyclic permutation

is used in the actual Monte Carlo evaluation:

$$\rho_{\overline{\text{PIMC}}}(\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_{M-1}) = \frac{1}{M} \sum_{i=0}^{M-1} \rho_{\text{PIMC}}(\mathbf{R}_i, \mathbf{R}_{\text{mod}(i+1, M)}, \dots, \mathbf{R}_{\text{mod}(i+M-1, M)}). \quad (4.10)$$

4.2.2 Population-Normalized Estimator and Importance Sampling

In the previous approach described in Eq. 4.8, the phonon coordinates are sampled according to the electronic ground state PES. The estimator should converge to the target quantity in the long time limit, taking into account the discretization error. As long as $f_g(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})$ is positive definite everywhere in the phonon space, the sampling efficiency depends on the selection of the probability density. Obviously, the actual distribution of the phonon coordinate depends heavily on the excited state PES. Therefore, the Monte Carlo points coordinates sampled according to the reduced dynamics of the bath by taking the partial trace with respect to the *exciton degrees of freedom*, as explored in multiple surface path integral Monte Carlo approaches, are expected to give the better estimates. This choice of the probability density reweights the estimator in the following way:

$$\begin{aligned} f_I(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}) &= \text{Tr}_S [\rho_{\overline{\text{PIMC}}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})] f_g(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}), \\ \rho_I(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}) &= \frac{\rho_{\overline{\text{PIMC}}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})}{\text{Tr}_S [\rho_{\overline{\text{PIMC}}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})]}. \end{aligned} \quad (4.11)$$

In the expression above, we call $\rho_I(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})$ the population normalized estimator for the reduced density matrix because the sum of its populations is always constrained to be 1. The effective energy gap term of $-\frac{1}{\beta} \log \text{Tr} \rho_{\overline{\text{PIMC}}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})$

was added to the Eq. 4.9 to enable the phonons follow the excited state dynamics depending on the exciton state ρ_S . For the estimator of the reduced density matrix in Eq. 4.8, the normalization must be obtained by the estimates of its diagonal elements, leading to more uncertainties in the coherence. However, the population-normalized estimator preserves the correct normalization by construction, and does not suffer from any additional uncertainty.

Local gradient information can improve the efficiency and scaling of the sampling procedure by means of a gradient-based approach such as the Metropolis-adjusted Langevin algorithm (MALA). [125, 126] However, the exact closed form of the gradient of the effective energy gap term, $\log \text{Tr}_S \rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})$ can only be expressed as a function of a power series of matrices. Nevertheless, with the following approximation:

$$\sum_{k=0}^n A^k B A^{n-k} \approx \sum_{k=0}^n \frac{1}{2^n} \binom{n}{k} A^k B A^{n-k}, \quad (4.12)$$

an accurate approximation of the gradient can be obtained and employed in the sam-

pling procedure,

$$\begin{aligned}
\frac{\partial}{\partial R_{ij}} \log \text{Tr}_S [\rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})] &= \frac{\text{Tr}_S \left[\frac{\partial}{\partial R_{ij}} \rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}) \right]}{\text{Tr}_S [\rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})]} \\
&\approx \frac{\text{Tr}_S \left[-\frac{\tau}{2\hbar} \frac{\partial E(\mathbf{R}_i)}{\partial R_{ij}} \rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}) \right]}{\text{Tr}_S [\rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})]}, \\
\nabla_i \log f_g(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}) &= -\frac{\beta}{M} \nabla_i V_g(\mathbf{R}_i) \\
&\quad + \frac{M}{2\beta\hbar^2} \mathcal{M}(\mathbf{R}_{\text{mod}(i+1, M)} + \mathbf{R}_{\text{mod}(i-1, M)} - 2\mathbf{R}_i), \\
\mu_i(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}) &= \frac{\text{Tr}_S \left[-\frac{\tau}{2\hbar} \frac{\partial E(\mathbf{R}_i)}{\partial R_{ij}} \rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}) \right]}{\text{Tr}_S [\rho_{\text{PIMC}}(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})]} \\
&\quad + \nabla_i \log f_g(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}) \\
&\approx \nabla_i \log f_I(\mathbf{R}_0, \dots, \mathbf{R}_{M-1}). \tag{4.13}
\end{aligned}$$

Here, ∇_i is the gradient operator with respect to \mathbf{R}_i .

Note that if we choose an appropriate Metropolis criterion, no bias in the distribution is introduced even with the approximate gradient [127]. Firstly, a trial move \mathbf{R}'_i obtained by

$$\mathbf{R}'_i = \mathbf{R}_i + \mu_i(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})\Delta t + \xi_i\sqrt{\Delta t}, \tag{4.14}$$

where Δt is the timestep for the Monte Carlo step and ξ_i is a N -dimensional vector of independent standard Gaussian random variables. Then, \mathbf{R}'_i is probabilistically accepted according to the acceptance ratio,

$$\frac{f_I(\mathbf{R}'_0, \dots, \mathbf{R}'_{M-1})}{f_I(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})} \times \frac{\prod_{i=0}^{M-1} \exp \left[-\frac{|\mathbf{R}'_i - \{\mathbf{R}_i + \mu_i(\mathbf{R}_0, \dots, \mathbf{R}_{M-1})\}|^2}{2\Delta t} \right]}{\prod_{i=0}^{M-1} \exp \left[-\frac{|\mathbf{R}_i - \{\mathbf{R}'_i + \mu_i(\mathbf{R}'_0, \dots, \mathbf{R}'_{M-1})\}|^2}{2\Delta t} \right]}. \tag{4.15}$$

The Monte Carlo timestep Δt is only a tunable parameter for the Monte Carlo sampling procedure and not related to the physics of the simulated system.

Parameters	Value
k_{11}	4×10^{-5}
k_{22}	3.2×10^{-5}
x_{11}	7
x_{22}	10.5
ε_{11}	0
ε_{22}	2.2782×10^{-5}
c	5×10^{-5}
α	0.4
x_{12}	8.75
m	3.6743×10^3

Table 4.1: Summary of the parameters for the model system by Alexander *et al* [128]. All values are given in atomic units.

4.3 Application

4.3.1 Alexander’s 1D Test Model

Our formulation is equivalent to the multiple electronic state extension of matrix multiplication path integral (MMPI) method of Alexander [120, 128] when the population normalized estimator is chosen and only the vibrational degrees of freedom are considered. Therefore, the 1D model employed in Ref. [128] was calculated to test the validity of our method. The elements of the electronic Hamiltonian in this model are given by,

$$\begin{aligned}
 V_{11}(x) &= \frac{1}{2}k_{11}(x - x_{11})^2 + \varepsilon_{11}, \\
 V_{22}(x) &= \frac{1}{2}k_{22}(x - x_{22})^2 + \varepsilon_{22}, \\
 V_{12}(x) &= c \exp \left[-\alpha(x - x_{12})^2 \right],
 \end{aligned} \tag{4.16}$$

The total nuclear probability density evaluated as histograms from the Metropolis random walk and MALA simulations are compared to the grid-based result from

Alexander *et al.* [128] in Fig. 4.1. The distributions converged to the exact probability density after 2×10^7 steps with 8 beads at both temperatures of 8K and 30K.

4.3.2 Model of a Chromophore Heterodimer with Displaced Harmonic Oscillators

To test the proposed method, a system of two chromophores in a photosynthetic complex was modeled using displaced harmonic oscillator model. In this model, the ground and excited electronic states of the monomer are modeled as harmonic oscillators with different displacement, but the same harmonic constant [67]. The thermal reduced density matrix was calculated within the single exciton manifold. The Hamiltonian for this model is then given as follows:

$$\begin{aligned}
 V_g(x_1, x_2) &= \frac{1}{2}(k_1 x_1^2 + k_2 x_2^2), \\
 V_e(x_1, x_2) &= \begin{pmatrix} \frac{1}{2}k_1\{(x_1 - d_1)^2 - x_1^2\} + \varepsilon_1 & J \\ J & \frac{1}{2}k_2\{(x_2 - d_2)^2 - x_2^2\} + \varepsilon_2 \end{pmatrix}, \\
 \mathcal{M} &= \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}.
 \end{aligned} \tag{4.17}$$

Some of the parameters were set according to our molecular dynamics/quantum chemistry calculation of the FMO complex [32]. The parameter values are listed in table 4.2.

The model system was simulated at seven different temperatures ranging from 30K to 300K with a number of beads (discretization number) of 4, 8, 16, 32 and 64. The number of timesteps propagated in each simulation was 4×10^7 . The value of

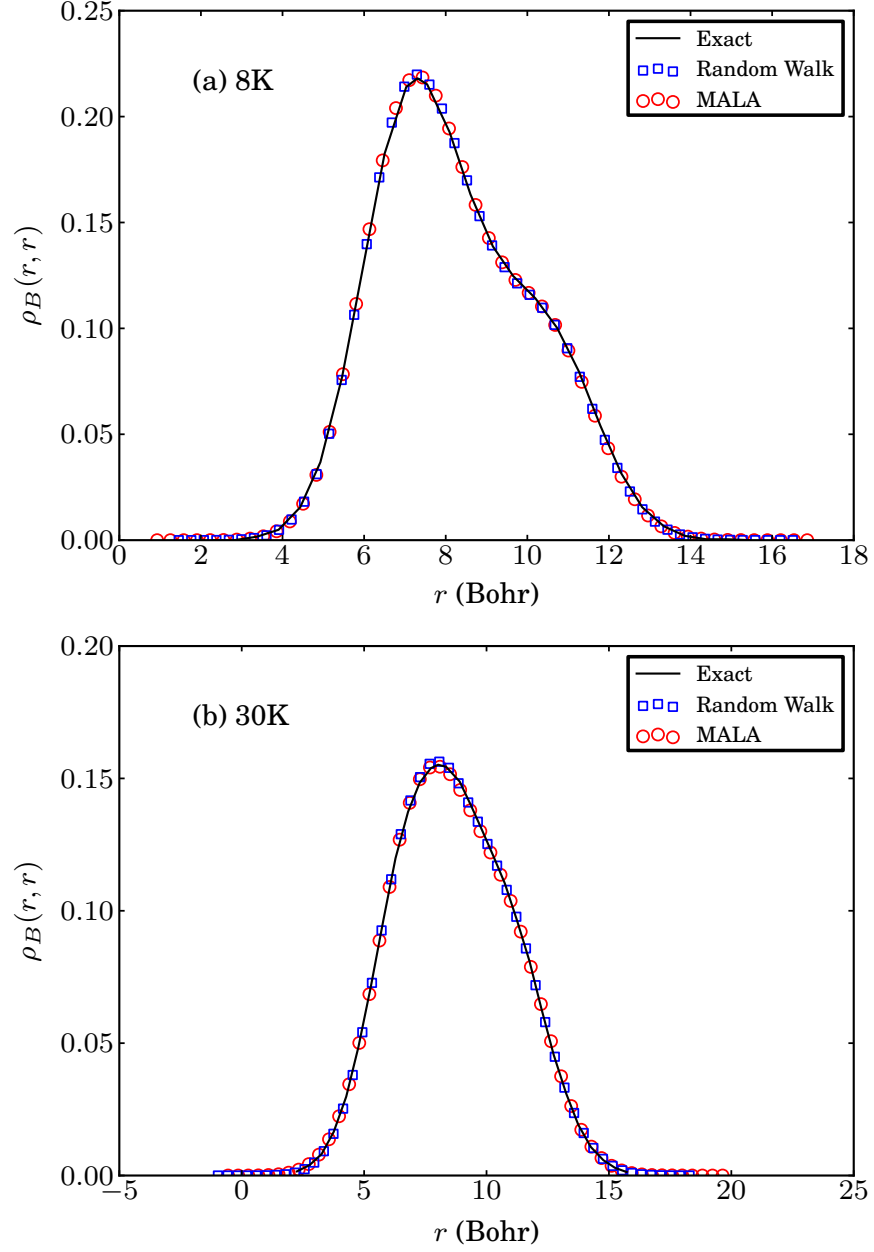


Figure 4.1: The estimated nuclear probability densities of Alexander's model [128] at (a) 8K and (b) 30K. For path integral Monte Carlo simulations, densities were obtained by histograms with 50 bins. The discretization number of 8 was enough to converge to the exact probability densities.

Parameter	Value
k_1	2.227817×10^{-3}
k_2	2.227817×10^{-3}
d_1	3.00000
d_2	2.00000
ε_1	8.064745×10^{-2}
ε_2	7.976238×10^{-2}
J	-4.738588×10^{-4}
m_1	3.418218×10^6
m_2	3.418218×10^6

Table 4.2: Summary of the parameters for the displaced harmonic oscillator model used in Sec. 4.3.2. All values are given in atomic units.

each timestep was tuned so that the acceptance ratio of the MALA run is close to 0.574, and 0.234 for the Metropolis random walk as maintaining these acceptance ratio is known to provide most efficient sampling [126]. We used non-overlapping batch means [129] with a batch size of 10^6 to estimate the standard error of the correlated samples. The batch size was adjusted so that the null hypothesis of uncorrelated batches was not rejected by using Ljung-Box test [130] at a significance level of 5%.

As shown in Fig. 4.2, the standard error of the simulation decreases modestly as the number of Monte Carlo steps increases. Fig. 4.3 shows the temperature dependence of the estimates of reduced density matrix elements as a function of various discretization numbers using MALA. Although the Metropolis random walk simulation gives a smaller confidence interval for the 4 bead case, MALA provides better estimates as the dimension of the sample space increases. The Metropolis random walk result is given in Fig. 4.4. While the population of the low energy site decreases as the temperature increases, the quantum coherence does not monotonically decrease. We believe that this phenomenon is an artifact of an insufficient discretiza-

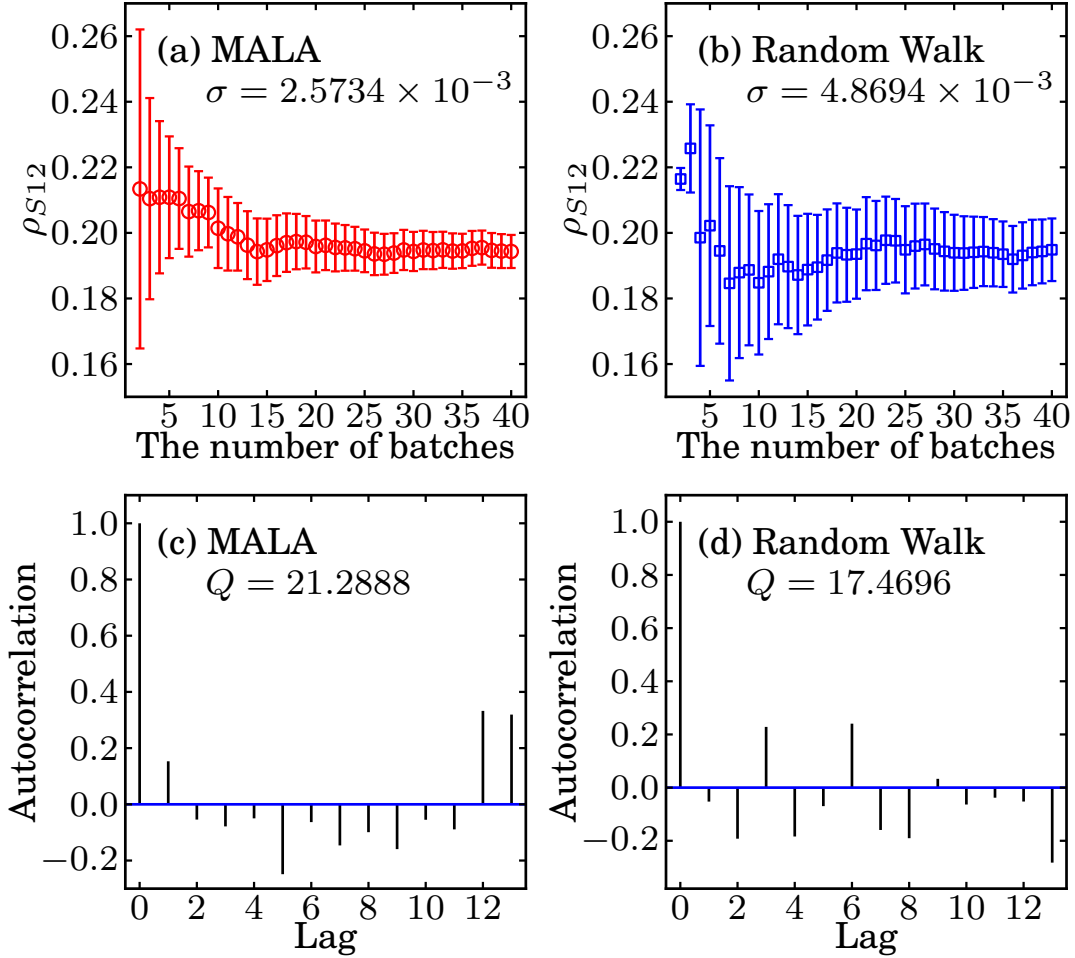


Figure 4.2: Estimates of (1,2) matrix elements of the thermal reduced density matrix evaluated using MALA and Metropolis random walk at 77K with 64 beads. MALA estimate has a smaller confidence interval thus a more accurate estimate than that of the Metropolis random walk. The error bar indicates the 95% confidence interval evaluated with the batch means. The 0.95 quantile of the χ^2 distribution with 13 degrees of freedom is 22.362 and both Ljung-Box statistics (Q) are smaller. Thus, the uncorrelation hypothesis is not rejected in both cases at the 5% significance level.

tion number at low temperatures. As can be seen in Fig. 4.3, 64 or more beads are needed for the coherence to converge at 77K, while 16 beads are enough at 300K with acceptable accuracy. This is a well known limitation of imaginary time path integral Monte Carlo simulations. Figure 4.5 shows the probability density function of the phonon coordinate at 77K and 300K. The population difference in the reduced density matrix is reflected to the difference in the probability mass of the two diabatic potential energy minimum at $(3, 0)$ and $(0, 2)$.

4.4 Conclusion

We explore a method for obtaining the thermal reduced density matrix of an exciton system coupled to an arbitrary phonon bath for path integral Monte Carlo simulation. Note that our scheme is closely related to the path integral Monte Carlo simulation for nonadiabatic systems for vibrational coherence [128, 131, 132]. Although the phonon state can be obtained as a byproduct, we mainly focused on the evaluation of the reduced density matrix of the excitonic system to explore the asymptotic behavior of the populations and coherences in this paper. In addition, we implemented an importance sampling scheme for better spatial scaling and sampling efficiency. Although the path integral Monte Carlo cannot evaluate the real time evolution of density matrices, the method gives the exact asymptotic values with all quantum effects from both the system and bath environments if a sufficient number of beads are used. We believe that in some of the cases where the bath has a nontrivial coupling to the system, or the non-Markovianity of the bath manifests very strongly, treating the environment around the system of interest as a set of harmonic oscillators

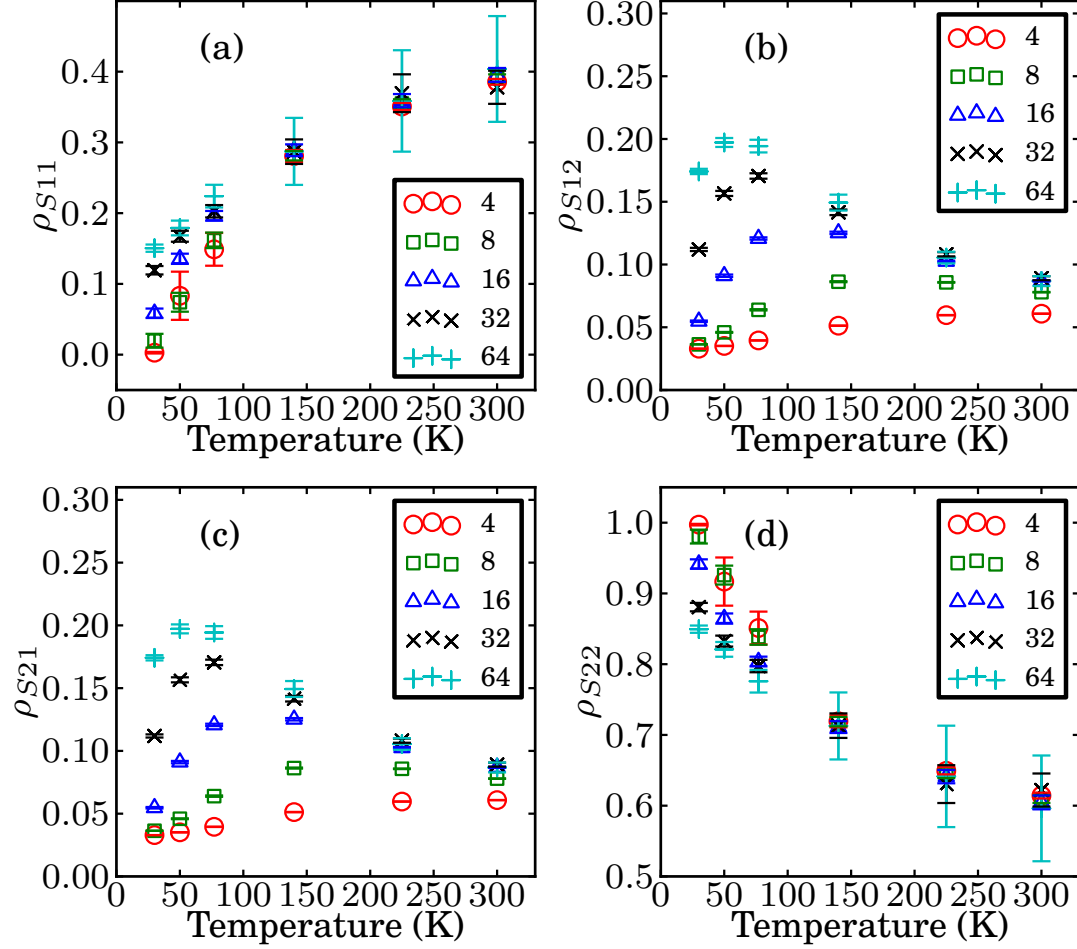


Figure 4.3: Estimates of matrix elements of the thermal reduced density matrix evaluated at 30K, 50K, 77K, 140K, 225K and 300K with different discretization numbers of 4, 8 and 16 using MALA. (a) is the (1,1) element, (b), (c) and (d) are (1,2), (2,1) and (2,2) elements, respectively. The error bar indicates the 95% confidence interval evaluated with the batch means.

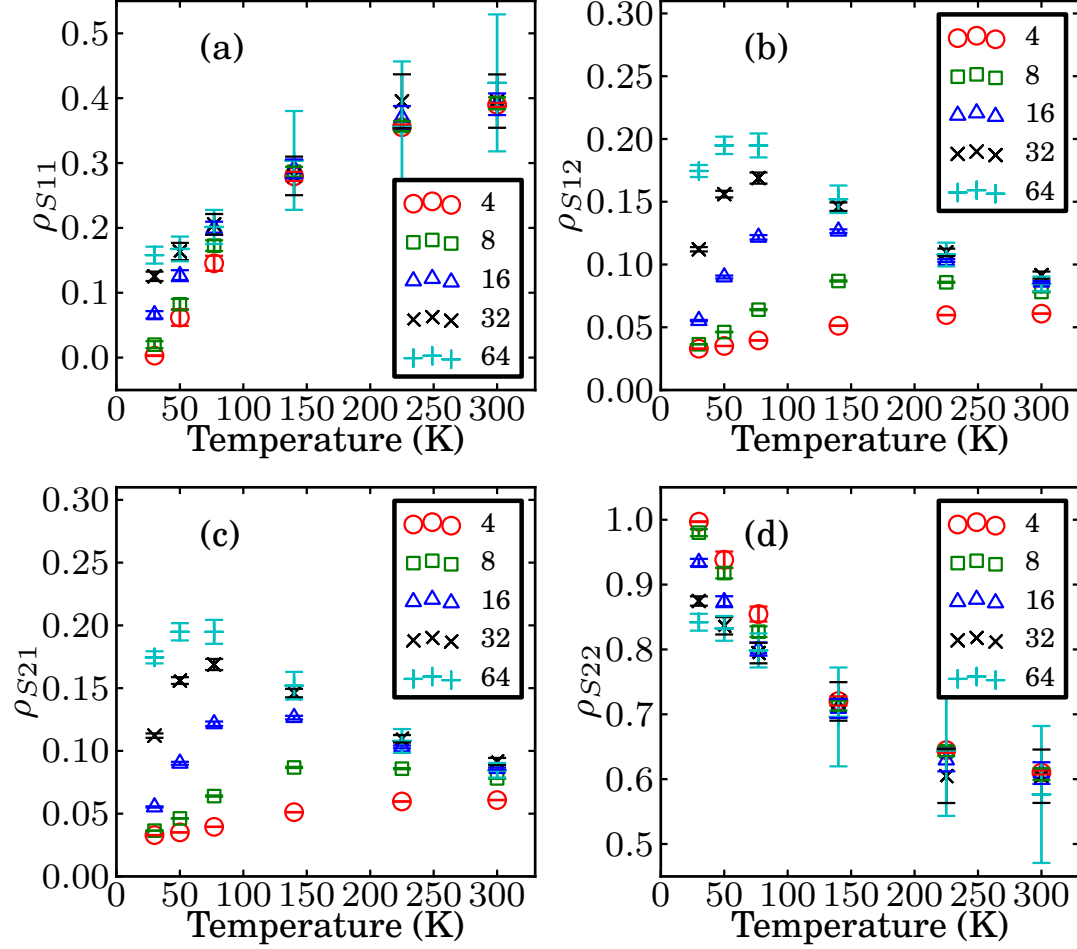


Figure 4.4: Estimates of matrix elements of the thermal reduced density matrix evaluated at 30K, 50K, 77K, 140K, 225K and 300K with different discretization numbers of 4, 8 and 16 using random walk Metropolis. (a) is the (1,1) element, (b), (c) and (d) are (1,2), (2,1) and (2,2) elements, respectively. The error bar indicates the 95% confidence interval evaluated with the batch means.

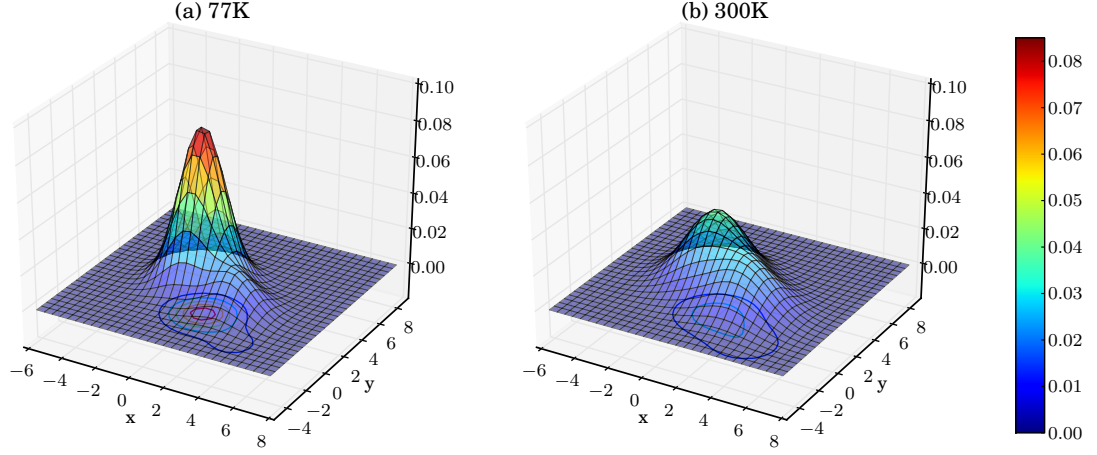


Figure 4.5: The phonon probability density function evaluated at (a) 77K and (b) 300K with 16 beads using MALA. At the lower temperature, the contribution of the exciton with lower energy at $(0, 2)$ becomes larger. Therefore, the population difference becomes more distinct, as can be seen in the temperature dependence of the exciton population in Fig. 4.3.

is not sufficient. If this is the case, the system should be studied in its entirety. We are trying to develop a real time propagation scheme to treat the system exactly, and the bath semiclassically. The method studied in this paper offers a foundation for it by providing the correct asymptotic behaviors.

Part II

Simulations of Molecular Systems and Applications

Chapter 5

First-principles semiclassical initial value representation molecular dynamics

5.1 Introduction

Algorithms for the simulation of molecular dynamics belong to the fundamental toolset of modern theoretical chemical physics. Classical simulation methods are able to study systems with up to millions of particles but are unable to describe quantum effects such as tunnelling and delocalization. Exact quantum mechanical methods are restricted to a few quantum particles, especially when pre-computed analytical potential energy surfaces (PES) are employed.

First-principles molecular dynamics (FPMD) algorithms have been introduced as an alternative to the pre-calculation of the PES. FPMD avoids any source of error

originated from the fitting of the PES. This is particularly true for many degrees of freedom, where the fitting procedure might not represent the many-dimensional surface accurately. In this family of methods, the potential and its derivatives are calculated *on-the-fly* as the dynamical simulation progresses and are directly obtained from electronic structure calculations. In the Born-Oppenheimer molecular dynamics (BOMD) approach, the electronic structure calculations for a given simulation step are converged based on previous step information. This approach can lead to systematic energy drifts and several methods have been proposed to avoid this effect [133]. Alternatively, extended Lagrangian molecular dynamics approaches (ELMD) [134–137] involve the propagation of nuclear and electronic degrees of freedom simultaneously. The electronic degrees of freedom are assigned to classical variables that are propagated using classical equations of motion and these can be expanded in terms of plane waves [134], Gaussian functions [136] or real-space grids [137]. Usually ELMD propagation is computationally more efficient, however questions have raised on whether the resultant energy surface remains close to the actual Born-Oppenheimer one and about disturbing dependencies on the fictitious electronic masses [136, 138].

While the evaluation of the potential *on-the-fly* can be easily integrated with classical simulations, the delocalized nature of quantum mechanical propagation has led to the development of many alternative approaches for the simulation of quantum dynamics. For example, the path-integral centroid molecular dynamics approach [139] includes quantum nuclear effects employing an extended Lagrangian. Alternatively, in the variational multi-configuration Gaussian wavepacket method (vMCG) [140] the quantum wavepackets are represented by fixed-width Gaussian functions for which the

potential is approximated to be locally harmonic. Other approaches introduce a mean field approximation and then update the dynamics in a time-dependent self-consistent fashion [141, 142].

Semiclassical molecular dynamics methods [143–152] are based on classical trajectories and therefore are amenable for carrying out *on-the-fly* calculation of the potential. The benefits of calculating the potential only when needed have been suggested by Heller and co-workers [152, 153]. In between formally exact quantum methods and classical dynamics, semi-classical methods include quantum effects approximately. Two representative semi-classical approaches are the coupled coherent states (CCS) technique [154] and the ab initio multiple spawning method (AIMS) algorithm [155]. In the CCS approach, several grids of coherent states are classically propagated and their trajectories can be derived from first principle dynamics. In AIMS, the nuclear wavefunction are spawned onto a multiple potential surface basis set. This set is made of adaptive time-dependent fixed-width Gaussian functions, which are generated by classical Newtonian dynamics.

5.2 First-Principles SC-IVR

In this work, we show how the semiclassical initial value representation (SC-IVR) method [144] can be coupled tightly and naturally, without any mayor change in formulation, with first principles electronic structure approaches to carry out classical molecular dynamics. We show how the method is able to reproduce approximately quantum effects such as the vibrational power spectra using a single, short classical trajectory using computational resources comparable to those employed in

first-principles molecular dynamics calculations. Calculations employing multiple trajectories can in principle be more accurate (and more computational intense as well), but here we focus on analyzing the predictive power of single trajectory runs. Finally, we describe how different approaches can be used in conjunction with this method for studying the symmetry of the vibrational states either by arranging the initial conditions of the classical trajectory or by employing the symmetry of the coherent state basis.

In the SC-IVR method, the propagator in F dimension is approximated by the phase space integral,

$$e^{-i\hat{H}t/\hbar} = \frac{1}{(2\pi\hbar)^F} \int d\mathbf{p}(0) \int d\mathbf{q}(0) C_t(\mathbf{p}(0), \mathbf{q}(0)) \times e^{iS_t(\mathbf{p}(0), \mathbf{q}(0))/\hbar} |\mathbf{p}(t), \mathbf{q}(t)\rangle \langle \mathbf{p}(0), \mathbf{q}(0)|, \quad (5.1)$$

where $(\mathbf{p}(t), \mathbf{q}(t))$ are the set of classically-evolved phase space coordinates, S_t is the classical action and C_t is a pre-exponential factor. In the Heller-Herman-Kluk-Kay version of the SC-IVR [151, 156], the prefactor involves mixed phase space derivatives,

$$C_t(\mathbf{p}(0), \mathbf{q}(0)) = \sqrt{\frac{1}{2} \left| \frac{\partial \mathbf{q}(t)}{\partial \mathbf{q}(0)} + \frac{\partial \mathbf{p}(t)}{\partial \mathbf{p}(0)} - i\hbar\gamma \frac{\partial \mathbf{q}(t)}{\partial \mathbf{p}(0)} + \frac{i}{\gamma\hbar} \frac{\partial \mathbf{p}(t)}{\partial \mathbf{q}(0)} \right|}, \quad (5.2)$$

as well as a set of reference states,

$$\langle \mathbf{q} | \mathbf{p}(t), \mathbf{q}(t) \rangle = \prod_i (\gamma_i/\pi)^{F/4} \exp[-\gamma_i \cdot (q_i - q_i(t))/2 + ip_i(t) \cdot (q_i - q_i(t))/\hbar], \quad (5.3)$$

of fixed width γ_i . For bound systems, the widths are usually chosen to match the widths of the harmonic oscillator approximation to the wave function at the global minimum and no significant dependency has been found under width variation [145]. By introducing a $2F \times 2F$ symplectic monodromy matrix $\mathbf{M}(t) \equiv$

$\partial((\mathbf{p}_t, \mathbf{q}_t) / \partial(\mathbf{p}_0, \mathbf{q}_0))$, one can calculate the pre-factor of Eq. 5.2 from blocks of $F \times F$ size and monitor the accuracy of the classical approximate propagation by the deviation of its determinant from unity. Wang *et al.* suggested calculating the determinant of the positive-definite matrix $\mathbf{M}^T \mathbf{M}$ instead [157] and we monitored the same quantity for this work. The spectral density is obtained as a Fourier transform of the surviving probability[151]. The SC-IVR expression of the probability of survival for a phase-space reference state $|\chi\rangle = |p_N, q_N\rangle$ is,

$$\begin{aligned} \langle \chi | e^{-i\hat{H}t/\hbar} | \chi \rangle &= \frac{1}{(2\pi\hbar)^F} \int d\mathbf{p}(0) \int d\mathbf{q}(0) C_t(\mathbf{p}(0), \mathbf{q}(0)) \\ &\times e^{iS_t(\mathbf{p}(0), \mathbf{q}(0))/\hbar} \langle \chi | \mathbf{p}(t), \mathbf{q}(t) \rangle \langle \mathbf{p}(0), \mathbf{q}(0) | \chi \rangle. \end{aligned} \quad (5.4)$$

The phase-space integral of Eq. 5.4 is usually computed using Monte Carlo methods. If the simulation time is long enough, the phase space average can be well approximated by a time average integral. This idea has been suggested and implemented by Kaledin and Miller [158] to obtain the time averaging (TA-) SC-IVR approximation [159] for the spectral density,

$$\begin{aligned} I(E) &= \frac{1}{(2\pi\hbar)^F} \int d\mathbf{p}(0) \int d\mathbf{q}(0) \frac{\text{Re}}{\pi\hbar T} \int_0^T dt_1 \int_{t_1}^T dt_2 C_{t_2}(\mathbf{p}(t_1), \mathbf{q}(t_1)) \\ &\times \langle \chi | \mathbf{p}(t_2), \mathbf{q}(t_2) \rangle e^{i(S_{t_2}(\mathbf{p}(0), \mathbf{q}(0)) + Et_2)/\hbar} \left[\langle \chi | \mathbf{p}(t_1), \mathbf{q}(t_1) \rangle e^{i(S_{t_1}(\mathbf{p}(0), \mathbf{q}(0)) + Et_1)/\hbar} \right]^*, \end{aligned} \quad (5.5)$$

where $(\mathbf{p}(t_1), \mathbf{q}(t_1))$ and $(\mathbf{p}(t_2), \mathbf{q}(t_2))$ are variables that evolve from the same initial conditions but to different times, and T is the total simulation time. The advantage of this approach is that the additional time integral can in principle replace the need for phase-space averaging in the large-time limit of a single trajectory. Calculations of the vibrational spectra of systems such as the water molecule have proved to be

very accurate using the TA-SC-IVR approach and its inexpensive single-trajectory variant showed significant improvements over the simple harmonic approximation for excited vibrational levels [158]. In order to make Eq. 5.5 less computationally demanding, one can employ the separable approximation [158], where the pre-factor of Eq. 5.5 is approximated as a phase, $C_{t_2}(\mathbf{p}(t_1), \mathbf{q}(t_1)) = \text{Exp}[i(\phi(t_2) - \phi(t_1))/\hbar]$, and $\phi(t) = \text{phase}[C_t(\mathbf{p}(0), \mathbf{q}(0))]$. Using this approximation, Eq. 5.5 becomes

$$I(E) = \frac{1}{(2\pi\hbar)^F} \frac{1}{2\pi\hbar T} \times \int d\mathbf{p}(0) \int d\mathbf{q}(0) \left| \int_0^T dt \langle \chi | \mathbf{p}(t), \mathbf{q}(t) \rangle e^{i(S_t(\mathbf{p}(0), \mathbf{q}(0)) + Et + \phi_t(\mathbf{p}(0), \mathbf{q}(0))/\hbar)} \right|^2 \quad (5.6)$$

leading to a simplification of the double-time integration to a single time integral. The resulting integral is positive definite, making more amenable for Monte Carlo integration. Our numerical tests show that the results of carrying out this approximation are essentially identical to the double time integral approach when using a single trajectory. In this paper results will be reported by use of this last approximation, since it is computationally cheaper and numerically more stable than Eq. 5.5.

For this work, we compute the potential energy surface at each nuclear configuration directly from the Kohn-Sham orbitals expanded on a non-orthogonal Gaussian basis. Gradients and Hessians at each nuclear configuration are obtained analytically from electronic orbitals. The evaluation of the potential represents most of the computational effort of our approach, which is roughly a few hours of computer time using standard desktop machines for a 1 cm^{-1} spectrum resolution. The nuclear equations

of motion are,

$$M_I \ddot{\mathbf{R}}_I = -\nabla_I \min_{\mathbf{C}} E_{DFT}[\mathbf{C}, \mathbf{R}_I], \quad (5.7)$$

where \mathbf{C} is the rectangular matrix of the lowest occupied orbitals and the classical propagation is performed according to the velocity-Verlet algorithm, as implemented in the Q-Chem package [160]. At each time step, the potential, nuclear gradient and Hessian are used to calculate the action, pre-factor and coherent state overlaps necessary for the TA-SC-IVR method (Eq. 5.5 and 5.6). A schematic representation of an implementation of the algorithm for a multithreaded machine is shown in Fig. (5.1). At each time step, results are accumulated for time-average integration. The results presented on this work were carried out on a single thread. For each classical trajectory, the procedure is repeated and the final integration gives the spectrum intensity $I(E)$ for a given parametric value of E . The same procedure is repeated for next $E + \Delta E$, where in our calculation $\Delta E = 1\text{cm}^{-1}$. As previously mentioned, the trajectory is monitored by calculating at each time step the deviation of the determinant of the monodromy matrix from unity. The difference in the determinants was always smaller than 10^{-6} during the course of the calculations. A time step of 10 a.u. has been always found to satisfy the strict monodromy matrix restrictions even for the lightest atoms.

The calculation of the full dimensional vibrational power spectrum of the CO_2 molecule is a challenging test for FP-SC-IVR method: A successful method should reproduce spectral features such as degenerate bending modes, strong intermodal couplings and Fermi resonances. To evaluate the FP-SC-IVR method, we compare vibrational spectrum of CO_2 molecule from FP-SC-IVR method to numerically-exact

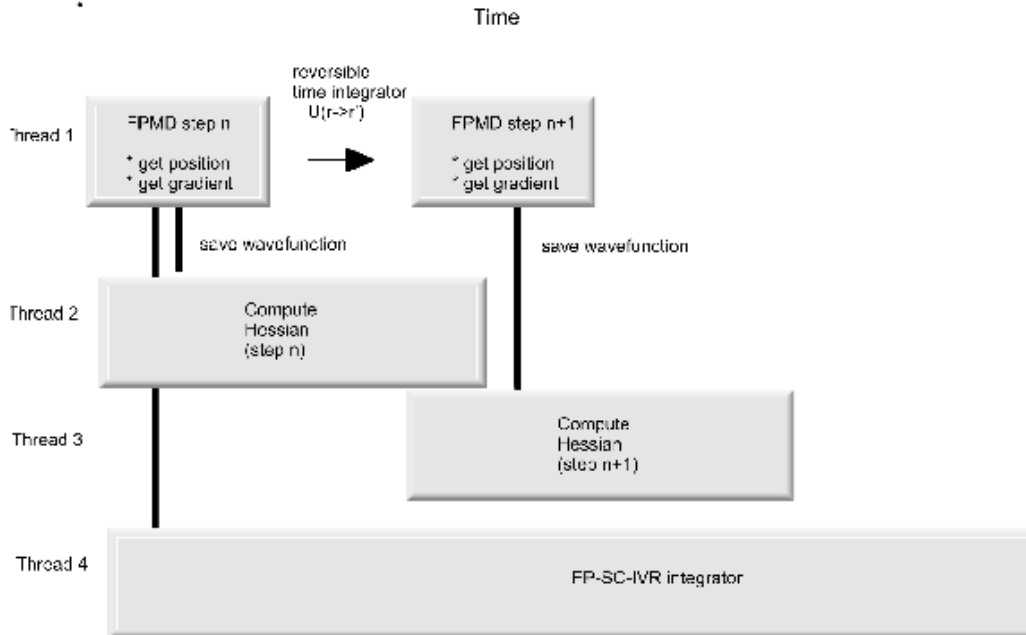


Figure 5.1: First-principles SC-IVR algorithm: At each time step electronic wavefunction are saved to calculate nuclear Hessian. Nuclear positions, gradients and Hessian are accumulated for the spectral time-average integral.

discrete variable representation (DVR) eigenvalue calculations on a potential fitted to a set of first-principles points obtained at the same level of theory. The next section describes the details of the potential fitting and DVR calculation. Following, we continue on the discussion of the FP-SC-IVR method.

5.3 Potential Fitting and Grid Calculations

The CO_2 molecule is a linear molecule with four vibrational normal modes: a symmetric stretching mode (ν_1), degenerate bending modes (ν_2 and $\bar{\nu}_2$), and an antisymmetric stretching mode (ν_3). A 3d potential energy grid in internal coordinates is calculated using the B3LYP density functional [161] with the cc-pVDZ

basis set [162]. The grid points are then fitted to a potential energy surface [163] represented by a fourth-order Morse-cosine expansion,

$$V(r_1, r_2, \theta) = \sum_{i,j,k=0}^4 K_{ijk} (1 - e^{-a_1(r_1-r_e)})^i \times (\cos\theta - \cos\theta_e)^j (1 - e^{-a_2(r_2-r_e)})^k, \quad (5.8)$$

where the parameter $r_e = 2.206119$ a.u. and $\theta_e = 180$ specify the equilibrium coordinates of the CO₂ molecule. The Morse parameters $a_1 = a_2 = 1.2489$ a.u. were determined so as to minimize the standard deviation of the differences of the fitted potential from the *ab initio* result using the Levenberg-Marquardt non-linear least square algorithm [164]. Instead, r_e was obtained by geometry optimization within the Q-Chem *ab initio* package [160].

The 35 K_{ijk} coefficients were subject to the non-linear least square fitting procedure to the DFT energies. Since these coefficients must be the same once r_1 and r_2 are swapped, 13 linear constraints of the type $K_{ijk} = K_{kji}$ were imposed during the fitting procedure. Additionally, to ensure that the equilibrium geometry was fitted to the predetermined equilibrium parametric distance, the coefficients K_{100} and K_{001} were constrained to be zero. Consequently, we employed a total number of 14 fitting constraints (K_{000} term is always constant). A total of 2500 *ab initio* grid points were chosen for the fitting process. These grid points range from 1.42 a.u. to 7.09 a.u. for r_1 and r_2 , and from 113.6 to 180 for the angle variable. The calculated expansion coefficients K_{ijk} are reported in Table 5.1.

As far as the numerically exact eigenvalues calculations is concerned, we used an exact DVR (Discrete Variable Representation) matrix diagonalization procedure. The CO₂ molecule was described for grid calculations in internal coordinates, while *on-the-fly* classical trajectories and the SC-IVR calculations described previously were

coeff.	attoJ	coeff.	attoJ
K_{001}	+0.000000	K_{100}	$= K_{001}$
K_{002}	+1.442886	K_{200}	$= K_{002}$
K_{003}	-0.032125	K_{300}	$= K_{003}$
K_{004}	+0.003630	K_{400}	$= K_{004}$
K_{010}	+0.726891	K_{111}	+0.392310
K_{011}	-0.443422	K_{110}	$= K_{011}$
K_{012}	-0.162970	K_{210}	$= K_{012}$
K_{013}	-0.101077	K_{310}	$= K_{013}$
K_{020}	+0.488451	K_{121}	+0.606572
K_{021}	-0.358126	K_{120}	$= K_{021}$
K_{022}	-0.210888	K_{220}	$= K_{022}$
K_{030}	+0.175981	K_{202}	+0.097300
K_{031}	-0.184503	K_{130}	$= K_{031}$
K_{112}	+0.103205	K_{211}	$= K_{112}$
K_{101}	+0.210532	K_{040}	+0.155374
K_{102}	+0.067998	K_{201}	$= K_{102}$
K_{103}	+0.068693	K_{301}	$= K_{103}$

Table 5.1: Expansion coefficients K_{ijk} for the CO₂ B3LYP/cc-pVDZ fitted potential energy surface in attoJoule units.

performed in Cartesian coordinates. No significant contamination between the rotational (set to zero kinetic energy) and vibrational motion was found within the simulation time. To this end, the deviation from simplecticity of the monodromy matrix in the vibrational sub-space were never more than 10^{-6} as previously mentioned.

The coordinates r_1 and r_2 are CO distances, and θ is the angle between the CO bonds. In these coordinates the kinetic part of the Hamiltonian for $J = 0$ is,

$$\begin{aligned}
 T = & \frac{p_1^2}{2\mu_{CO}} + \frac{p_2^2}{2\mu_{CO}} + \frac{j^2}{2\mu_{CO}r_1^2} + \frac{j^2}{2\mu_{CO}r_2^2} \\
 & + \frac{p_1p_2\cos\theta}{m_C} - \frac{p_1p_\theta}{m_Cr_2} - \frac{p_2p_\theta}{m_Cr_2} - \frac{\cos\theta j^2 + j^2\cos\theta}{2m_Cr_1r_2},
 \end{aligned} \tag{5.9}$$

where $p_k = -i\frac{\partial}{\partial r_k}$, $p_\theta = -i\frac{\partial}{\partial\theta}\sin\theta$, and $j^2 = -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}$. The carbon mass were taken to be $m_C = 12.0$ a.m.u., while the oxygen mass $m_O = 15.9949$ a.m.u. and the

reduced mass is as usual $1/\mu_{CO} = 1/m_C + 1/m_O$.

As previously mentioned, in order to calculate exact eigenvalues, a sine-DVR basis for the coordinates r_1 and r_2 and a Legendre-DVR basis for θ has been used [165]. For each degree of freedom 50 DVR functions were used and eigenvalues were converged to at least 10^{-3}cm^{-1} . The sine-DVR ranged from 1.51 a.u. to 3.78 a.u. and the magnetic quantum number m of the Legendre-DVR was zero.

Because of the restriction of total angular momentum $J = 0$, we couldn't observe all degenerate bending excitations. However, ZPE and several vibrational energy levels were obtained and compared with that ones coming from a single *on-the-fly* semiclassical trajectory.

5.4 First-Principles SC-IVR Calculations

The full power spectrum obtained using Eq. (5.5) after 3000 BOMD steps of 10 a.u. each is shown on the bottom of Fig. 5.2. For longer simulations, the monodromy matrix symplectic properties as well as the resolution of the spectrum started to deteriorate. The calculated vibrational zero-point energy (ZPE) value was 2518 cm^{-1} versus the exact value of 2514.27 cm^{-1} and both are in good agreement with the experimental value of 2508 cm^{-1} . In contrast, harmonic normal-mode analysis (whose frequencies are 656.62, 1363.46, 2423.47 wavenumbers) predicts a frequency of 2550.08 cm^{-1} . Thus, the TA-SC-IVR method successfully reproduces the ZPE anharmonic effects with the use of a single classical trajectory. Some representative frequencies of the power spectrum are presented in Table 5.2. The ZPE was shifted to zero for comparison with reported classical ELMD simulations on the same system

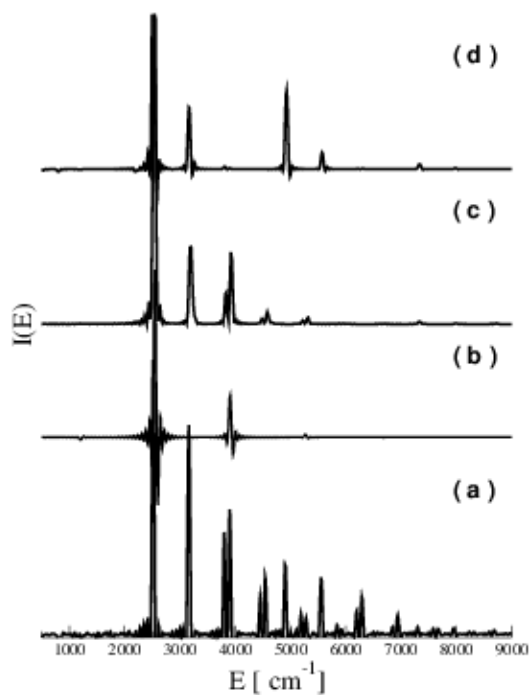


Figure 5.2: CO_2 Vibrational Power Spectrum: Initial kinetic energy on: (a) all modes; (b) symmetric mode; (c) one bending and symmetric modes; (d) bending and asymmetric modes.

that cannot reproduce the ZPE or higher vibrational states [166, 167] but only single modes frequencies. For these studies of Refs. 166 and 167, the vibrational data were obtained from the Fourier transform of correlation functions of classical trajectories in plane-wave DFT calculations. The ELMD approach predicts the following fundamental frequencies 648, 1368, 1428 and 2353 for Ref. 166 and 663, 1379, 1456 and 2355 for Ref. 167. These classical results are similar but limited to a normal mode analysis.

Table 5.2 compares our TA-SC-IVR results with the exact ones and to those obtained by Filho [168] with the same density functional and a basis set of comparable quality (6-31+G*) [169], using a perturbative approximation of the eigenvalue expansion. One can see how a different basis set results a significant deviation of vibrational levels spacing, once the comparison is performed in units of wavenumbers.

A major difficulty on the CO₂ power spectrum simulations is the calculation of the Fermi resonance splittings. These are the result of anharmonic couplings, and they represent a stringent test for a semi-classical method that relies on a single short trajectory. The Fermi resonances occur when an accidental degeneracy between two excited vibrational levels of the same symmetry exists and it results in a repulsion between the corresponding energy levels. The sources of these resonances are purely anharmonic and are only present in polyatomic potentials. For the CO₂ molecule, the unperturbed frequencies for the symmetric stretching are roughly equal to the first bending overtone ($\nu_1 \cong 2\nu_2$). For these modes, the wavefunctions are transformed as the irreducible representation of $D_{\infty h}$, *i.e.* $\nu_1(10^00)$ as Σ_g^+ , at the experimental frequency of 1388 cm⁻¹, and $\nu_2^2(02^00)$ as $\Sigma_g^+ + \Delta_g$, at an experimental frequency of

Exp. ^(a)	mode ^(b)	Harmonic ^(c)	FP-SCIVR-SA ^(d)	DVR	Ref. 168
667.4	0, 1 ¹ , 0	656.62	644		657.2
1285.4 [^]	0, 2 ⁰ , 0	1313.24	1288	1252.91	1283.4
1388.2 [^]	1, 0 ⁰ , 0	1363.46	1381	1372.29	1408.8
1932.5 [†]	0, 3 ¹ , 0	1969.86	1932		1930.2
2003.2	0, 3 ³ , 0	1969.86	2024		2004.9
2076.9 [†]	1, 1 ¹ , 0	2020.08	2106		2098.5
2349.1	0, 0 ⁰ , 1	2423.47	2388	2359.51	2411.5
2548.4 [‡]	0, 4 ⁰ , 0	2626.48	2515	2482.95	2553.3
2585.0 [*]	0, 4 ² , 0	2626.48	2578		2591.2
2671.7 [‡]	0, 4 ⁴ , 0	2626.48	2669	2640.15	2716.5
2760.7 [*]	1, 2 ² , 0	2676.70	2759		2796.3
2797.2 [‡]	2, 0 ⁰ , 0	2726.92	2793	2757.14	2845.2
4673.3	0, 0 ⁰ , 2	4846.94	4690 ⁺	4693.24	4797.8
6972.6	0, 0 ⁰ , 3	7270.41	6803 ⁺	6821.35	7152.9

Table 5.2: (a) Experimental frequencies in cm^{-1} from Ref. 173. (b) First number is the symmetric stretch quantum, second are the degenerate bendings, and third one is the asymmetric stretch. The exponent of the second number is the l_i degeneracy index. (c) Vibrational levels according to a normal modes harmonic model. (d) Using the Separable approximation of Eq. 5.6. Some of the calculated vibrational energy eigenvalues are tabulated. All data are in wavenumbers. Fermi Resonances group of frequencies are indicated by the same superscript symbols. Uncertain peaks are marked with (+). The first column represents the experimental vibrational frequencies associated with the modes listed on the second column. The third column shows the harmonic DFT results. In the fourth and fifth columns, we show our FP-SCIVR and exact numerical DVR calculations in the B3LYP/cc-PVDZ model chemistry used for the FP-SCIVR calculations. The fifth column shows perturbative DFT calculations carried out using a similar functional and basis set.

1285 cm^{-1} . Another Fermi doublet results from the addition of a quantum of bending mode to the previous Fermi doublet to yield the following states: $\nu_1\nu_2(11^10)$, at an experimental frequency of 2077 cm^{-1} and the $\nu_2^3(03^10)$ state, at an experimental frequency of 1932 cm^{-1} . Higher-energy Fermi resonances are indicated in Table 5.2 by using the same superscript symbols. The first Fermi terms are located at 1313 and 1363 in a harmonic approximation and corrected to 1288 and 1381 wavenumbers for FP-TA-SC-IVR. Thus, the original levels have been repelled by Fermi couplings. One mode is located at a higher frequency than the harmonic prediction, while the other is at a lower frequency. The latter effect could be explained also by simple anharmonicity, but the former is evidence of the ability of the single trajectory FP-TA-SC-IVR method even when the separable approximation is used to capture Fermi resonance effects partially. The same reasoning can explain the second Fermi doublet located at 1932 and 2106 for FP-TA-SC-IVR, while the harmonic estimate at 1970 and 2020 wavenumbers.

With the FP-TA-SC-IVR method, one can also identify the couplings between vibrational modes and the appearance of Fermi resonance splittings by carrying out simulations with different initial conditions. This can be achieved by selectively setting the initial velocity of some vibrational modes to zero. The anharmonic coupling between levels leads to a consistent reproduction of the ZPE peak in the spectrum for all simulations. However the excited vibrational peaks related to the modes with zero initial kinetic energy show a very small signal in the power spectrum. Vibrational energy redistribution processes can be studied as well, by carrying out simulations at different timescales. In Fig. 5.2, we show the resulting power spectra for different

initial conditions. If the initial state contains only purely symmetric motion, the lowest Fermi resonance peaks in Fig. 5.2(b) are absent as well as for a bending (without symmetric stretching) motion in Fig. 5.2(d). These results and the intensity of their peaks respect to that ones located at the same frequencies in Fig. 5.2(a) suggest that the Fermi resonance is indeed originated from the coupling between bending and the symmetric modes. One can reach the same conclusions by inspecting the lower Fermi doublet peaks intensity: by adding a bending mode (from Fig. 5.2(b) to Fig. 5.2(c)) and a second one (from Fig. 5.2(c) to Fig. 5.2(a)) the intensity of both peaks is gradually raised. This is called “intensity borrowing” and it arises from the strong mixing of the zero order states. These observations reinstate that “repulsion and mixing are the hallmarks of Fermi resonances” [170]. Also, for a distinct set of initial conditions, an additional peak at 5500 cm^{-1} related to the asymmetric stretch was observed. Using the proposed approach, one can carefully detect the characteristics of each peak even for complicated power spectra.

An attractive method for obtaining the symmetry properties of the eigenstates involves arranging the initial basis vectors [158, 171]. The basis for this method is the direct product of coherent states $|\chi\rangle = \prod_{k=1}^4 |p_i^{(k)}, q_i^{(k)}\rangle^{\epsilon_k}$. These states can be chosen to have an initial symmetry by employing linear combinations of the form $|p_i^{(k)}, q_i^{(k)}\rangle^{\epsilon_k} = \left(|p_i^{(k)}, q_i^{(k)}\rangle + \epsilon_k |-p, -q_i^{(k)}\rangle \right) / \sqrt{2}$. The k -th mode can be made symmetric ($\epsilon_k = 1$), antisymmetric ($\epsilon_k = -1$) or have no symmetry restrictions ($\epsilon_k = 0$). In order to assign the proper symmetry to each peak on Fig. 5.3, the reduced D_{2h} symmetry group was adopted. All irreducible representations were reproduced and peaks were grouped by symmetry as reported in Fig. 5.3. Note that (d) and (e)

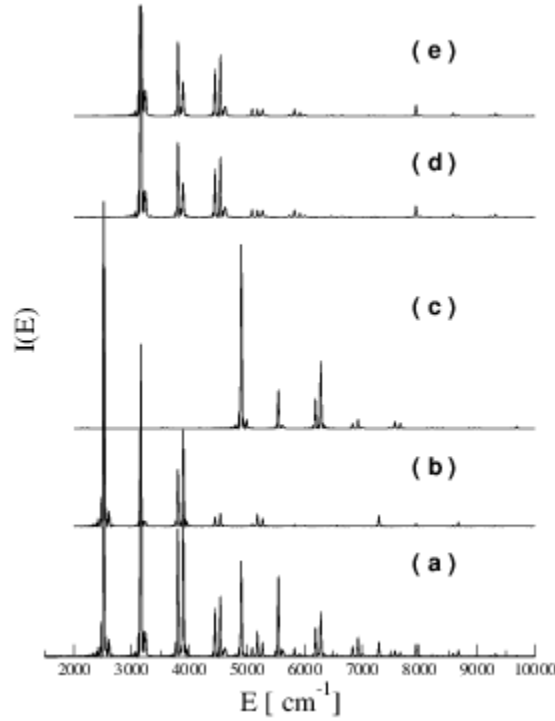


Figure 5.3: CO₂ Vibrational Power Spectrum (Separable approximation): Different basis set symmetries for ν_1 (symmetric stretching mode), ν_2 and $\bar{\nu}_2$ (bending modes) and ν_3 (asymmetric mode) and the corresponding D_{2h} irreducible representation; (a) all ϵ s are zero; (b) (B_{1u}): $\epsilon(v_1) = 0, \epsilon(v_2) = 1, \epsilon(\bar{\nu}_2) = 0, \epsilon(v_3) = -1$; (c) (A_g): $\epsilon(v_1) = 1, \epsilon(v_2) = 0, \epsilon(\bar{\nu}_2) = 0, \epsilon(v_3) = 1$; (d) (B_{2u}): $\epsilon(v_1) = 0, \epsilon(v_2) = -1, \epsilon(\bar{\nu}_2) = 0, \epsilon(v_3) = 1$, (e) (B_{3u}) $\epsilon(v_1) = 0, \epsilon(v_2) = 0, \epsilon(\bar{\nu}_2) = -1, \epsilon(v_3) = 1$. B_{2u} and B_{3u} representations are degenerated in the $D_{\infty h}$ subspace as shown.

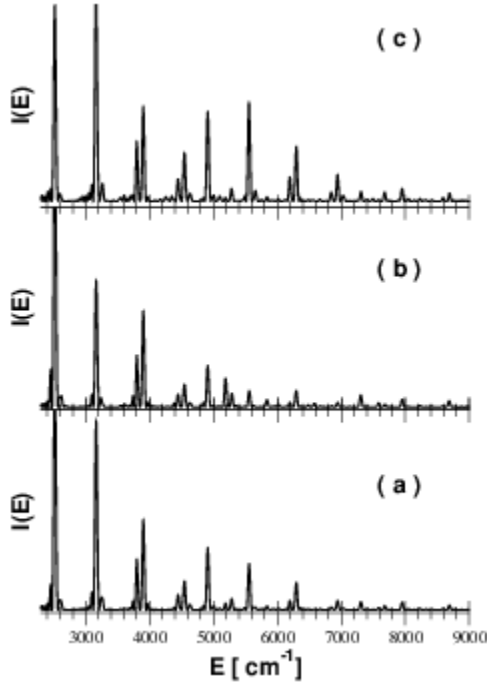


Figure 5.4: Gaussian width variations and related power spectra: a) $\gamma_i = \omega_i$; b) $\gamma_i = 2\omega_i$; c) $\gamma_i = \omega_i/2$, where ω_i are the i -*esime* normal mode frequency. The FP-SCIVR power spectra are fairly insensitive to variations in the value of the coherent state width.

plots are identical since they only differ trivially by swapping coefficients between the degenerate bending modes in the original $D_{\infty h}$ symmetry group.

Finally we have investigated the stability of the propagator versus variations of the coherent states gaussian width parameters γ_i . Previous calculations [156] have shown that there is no significant dependency on energy and norm conservation for the semiclassical propagator if suitable values of γ_i are chosen. For power spectra calculation we have chosen to look at vibrational levels variations under different values of coherent states width. Since a single trajectory was used in the FP-TA-SC-IVR approach, no Monte Carlo integration is performed in phase space coordinates and the changes of γ_i are confined to the coherent states overlap and to the prefactor in

Eq. 5.2. As reported in Fig. 5.4 and checked on a finer scale, no significant variation was observed beyond 1 cm^{-1} . These findings are in agreements with previous calculations on the same propagator [156]. Interestingly, a different distribution in peaks intensity were found in each panel. Since the peaks magnitude is proportional to the overlap between the reference state and the actual eigenfunction, the anharmonic choice ($\gamma_i = \omega_i/2$) is a more suitable solution as clearly showed on panel (c) of Fig. 5.4.

5.5 Conclusions

In conclusion, we have shown that SC-IVR can be implemented easily and efficiently using first principles molecular dynamics. With the modest computational cost of a single classical trajectory, the vibrational density of states of the CO_2 molecule was calculated. On Fig. 5.5 we report a graphical comparison between the harmonic and the FP-TA-SC-IVR approximations, versus the exact vibrational value for the Fermi resonance multiplets. One can notice how the single trajectory FP-TA-SC-IVR goes far beyond the harmonic approximation by removing the harmonic degeneracy and including part of anharmonicity. Fermi splittings are well mimicked not only for the first doublet, but also for the higher ones. The numerically exact DVR vibrational energy levels constrained by $J = 0$ are represented on the last column. The FP-TA-SC-IVR values are similar to the DVR results, when comparison is possible. However, a closer look at Table 5.2 shows how these single trajectory FP-TA-SC-IVR calculations can include only part of the anharmonicity and that their precision gets worse for higher vibrational levels. In particular, the spacing of the higher-energy states is

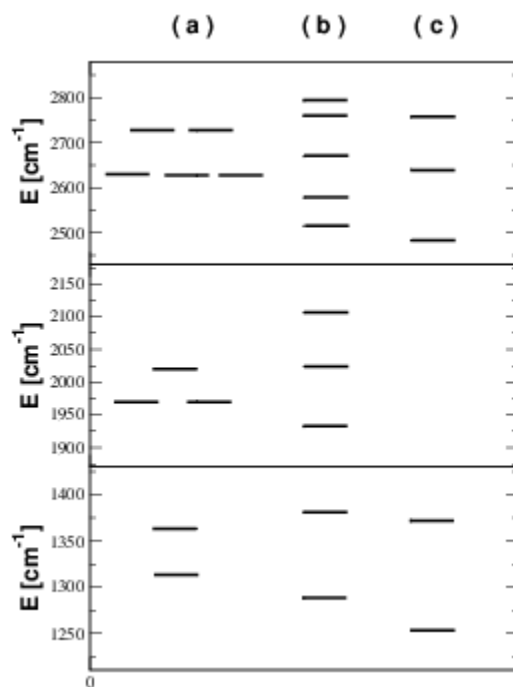


Figure 5.5: Fermi Resonance states vibrational energy level: (a) in harmonic approximation; (b) single FP-SC-IVR trajectory calculation; (c) exact grid calculation on splined potential.

harmonic-like and this is the major limitation of using a single classical trajectory.

These and previous calculations on model potentials [158] has shown how the single trajectory TA-SC-IVR gives reasonable results and performs better for higher frequencies modes. The computational cost of the method is essentially the same as classical propagation, and therefore, if broadly implemented in electronic structure codes, it can provide a description of quantum effects at a comparable computational cost to that of classical approaches. Possible applications of this method or related ones are the study of excited electronic states and Franck-Condon transitions, such as vibrational absorption spectra [174]. Although this single trajectory approach may be a practical tool for the simulation of more complex systems, the use of more trajectories is probably required to remove any harmonic “ghost states”. We are currently exploring the use of a small number of a set of systematically determined trajectories for further improvement of the results. If the number of required trajectories grows as a low polynomial of the system size, semi-classical methods could be competitive with currently-employed numerical approximations to obtain anharmonic vibrational effects. Finally, we expect that the representation of the potential energy in terms of normal coordinates will become less suitable when large amplitude motions or non adiabatic effects come into play.

Chapter 6

Simplified Sum-Over-States Approach for Predicting Resonance Raman Spectra

6.1 Introduction

Resonance enhancement of Raman scattering, which occurs whenever the frequency of the incident radiation approaches molecular excitation frequencies, was reported some 20 years after the initial experimental observation of the Raman effect [175, 176]. The large degree of enhancement spanning several orders of magnitude is useful for detection of the inherently inefficient spontaneous Raman scattering. Moreover, the shapes of Raman spectra change considerably at resonance with molecular excitations and provide information on structures and properties of electronic excited states. Resonance Raman spectroscopy is a sensitive spectroscopic

technique for strongly absorbing chemical constituents such as nucleic acid bases, aromatic aminoacids, and heme chromophores [177–179].

Another important manifestation of resonance enhancement emerges in surface-enhanced Raman scattering (SERS) [180–182]. The surface enhancement of Raman scattering is observed in molecules adsorbed on rough or nanostructured noble-metal surfaces and comprises an electromagnetic and a chemical contribution [182, 183]. The chemical contribution to the SERS intensities, while generally smaller in magnitude than the electromagnetic enhancement, is sensitive to the electronic structure of the adsorbate. The chemical effect leads to characteristic changes in the relative intensities of Raman bands and alters the overall shape of the Raman spectra compared to neat substance. Chemical effects are satisfactorily described by cluster models and can be attributed to resonance enhancement due to interface states [184, 185]. The combination of surface enhancement with intramolecular excitations gives rise to surface-enhanced resonance Raman scattering (SERRS) which provides an extraordinary sensitivity, even to the level of single-molecule detection [186–188].

While theoretical descriptions of resonance Raman scattering has been developed early on by Shorygin and co-workers [176, 189] and by Albrecht [190, 191], calculations of resonance Raman scattering from medium-size and large molecules are not often routinely performed. Raman scattering is a second-order process and its cross sections are given by the Kramers–Heisenberg–Dirac (KHD) dispersion relation [192, 193]. The classical expression for Raman cross sections involving derivatives of electronic polarizabilities with respect to vibrational normal modes can be obtained via closure of the sum over intermediate *vibronic* states in the KHD expression [194, 195].

The description of Shorygin and co-workers [189, 196] represents the polarizability derivatives as a sum over *electronic* states and introduces parameters for the resulting derivatives of excitation energies and oscillator strengths of the lowest excited state with respect to vibrational normal modes.

On the other hand, Albrecht’s approach is rooted in the vibronic coupling theory [197, 198] and introduces the Herzberg–Teller expansion into the sum over *vibronic* states of the KHD dispersion relation. Each vibronic state contributes four different terms denoted A, B, C, and D by Albrecht. The A term is due to vibrational wavefunction overlap of the initial and the intermediate state and of the intermediate and the final state. The B and C terms arise from the dependence of transition dipole moments on vibrational coordinates and are analogous to the intensity borrowing terms of vibronic coupling theory [197, 198]. The B term is derived from the coupling between the intermediate electronic excited state to other excited states, while the C term is due to the coupling between the ground electronic state to excited states and is customarily assumed to be small. The D term is of higher order in the coupling between electronic states and is often neglected. Albrecht’s treatment involves a full sum over all vibronic states of the molecule and is thus rarely computationally tractable for larger systems. Nevertheless, it constituted a major breakthrough in the understanding of Raman scattering in that it provided a unified picture for both non-resonant and resonant Raman spectra. Sums over vibronic states can be evaluated in the displaced harmonic oscillator approximation [199].

A different approach to resonance Raman scattering was proposed by Heller and co-workers [200]. It amounts to a transformation of the KHD dispersion relation into

the time domain, which represents the resonance Raman process as a propagation of vibrational wavepackets (multiplied with transition dipole moments) on the excited-state potential energy surface. Often, the short-time approximation to propagation dynamics is introduced [201], which has proven remarkably useful in interpreting resonance Raman spectra [201–203].

Finally, resonance Raman cross section can be expressed in a fashion analogous to the non-resonant case by introducing finite lifetimes for the intermediate states, or in other words, by computing Raman cross sections from derivatives of electronic polarizabilities evaluated at complex frequencies $\tilde{\omega} = \omega + i\gamma$ [204, 205]. Here ω is the excitation frequency and γ corresponds to an averaged lifetime of excited states, which is usually treated as an empirical parameter.

The purpose of the present work is to provide a simple and computationally tractable approximation for resonance Raman cross sections. To this end, we reduce the summation over vibronic states of the KHD dispersion relation to a summation of *electronic* states, similar to the parametric method of Shorygin and co-workers, and apply the double harmonic approximation, which is commonly used in calculations of vibrational spectra. This approximation requires only excitation energies, transition dipole moments, and their respective geometric derivatives to be computed for the electronic excited states included in the sum-over-states expression. In contrast to Shorygin’s work, all parameters in the sum-over-states expression are provided from *ab initio* calculations, while the summation runs over all electronic excitations in a given energy range. Analytical gradient techniques make computation of geometric derivatives particularly efficient in the framework of time-dependent density

functional theory (TDDFT) [206]. In addition, the sum-over-states approach may be used to identify major contributions to resonance Raman intensities. We apply the present approach to assign and interpret resonance Raman scattering in nucleic acid bases.

6.2 Theory

The polarizability theory of Raman scattering due to Placzek relates the Raman scattering cross section to frequency-dependent electronic polarizabilities at the frequency of the incident radiation [191, 195],

$$\alpha^{mn}(\omega) = \sum_k \left[\frac{\mu_{0k}^m \mu_{0k}^n}{\Omega_k - \omega} + \frac{\mu_{0k}^n \mu_{0k}^m}{\Omega_k + \omega} \right]. \quad (6.1)$$

m and n are Cartesian directions. We use atomic units throughout. The summation is over all electronic excited states $k > 0$ with excitation energies Ω_k and transition dipole moments μ_{0k}^m . The polarizability theory of Raman scattering is based on the separability of the electronic and nuclear wavefunctions (Born–Oppenheimer approximation) and the assumption that the incident radiation is sufficiently far from resonance such that energy differences between vibronic levels of the KHD expression may be approximated by electronic excitation energies Ω_k . In the double harmonic approximation, the Raman scattering cross sections are proportional to derivatives of $\alpha^{mn}(\omega)$ with respect to vibrational normal modes [191, 195]. Straightforward differentiation of the sum-over-states expansion for $\alpha(\omega)$ with respect to the vibrational normal mode Q yields the following expression for the Raman scattering cross section

of the vibration Q ,

$$\left(\frac{\partial \sigma}{\partial \Omega}\right)_Q = \frac{(\omega - \omega_Q)^4}{2\omega_Q c^4} |\langle \boldsymbol{\sigma}_Q(\omega) \rangle|^2, \quad (6.2)$$

where the components of the Raman scattering tensor $\sigma_Q^{mn}(\omega)$ are given by

$$\begin{aligned} \sigma_Q^{mn}(\omega) = & \sum_k \left[-\mu_{0k}^m \mu_{0k}^n \left[\frac{(\Omega_k - \omega)^2 - \gamma_k^2}{((\Omega_k - \omega)^2 + \gamma_k^2)^2} + \frac{i 2(\Omega_k - \omega)\gamma_k}{((\Omega_k - \omega)^2 + \gamma_k^2)^2} \right] \frac{\partial \Omega_k}{\partial Q} \right. \\ & \left. + \left[\mu_{0k}^m \frac{\partial \mu_{0k}^n}{\partial Q} + \frac{\partial \mu_{0k}^m}{\partial Q} \mu_{0k}^n \right] \left[\frac{\Omega_k - \omega}{(\Omega_k - \omega)^2 + \gamma_k^2} + \frac{i \gamma_k}{(\Omega_k - \omega)^2 + \gamma_k^2} \right] \right]. \end{aligned} \quad (6.3)$$

Here, ω_Q is the vibrational frequency, c is the speed of light. Angle brackets denote the appropriate orientational average over components of the Raman scattering tensor $\boldsymbol{\sigma}_Q(\omega)$. The excited states $k > 0$ have linewidths γ_k associated with them, which are chosen as empirical parameters independent of k in most studies. We will follow this practice here. The analogous expression for $\sigma_Q^{mn}(\omega)$ with uniform linewidths $\gamma_k = \gamma$ for all excited states may be obtained by differentiation of the polarizability evaluated at the complex frequency $\tilde{\omega} = \omega + i\gamma$ [204, 205]. In contrast, in the present approach different linewidths γ_k may be chosen for individual excited states to reflect differences in their lifetimes. Ultimately, the excited-state linewidths may be rigorously derived from an open-system formulation, e. g., in the framework of TDDFT [207–209].

In practice, the sum over electronic excited states has to be truncated. The number of excited states contributing significantly to the Raman cross sections in Eq. 6.3 will be small in the vicinity of a resonance ($|\omega - \Omega_k| \approx \gamma_k$) but might increase significantly in the non-resonant case. While truncation of the sum-over-states is a potential source of error not present in the finite-lifetime approach [204, 205], we find that convergence is sufficiently fast even in the non-resonant regime for nucleic acid bases considered here.

Differentiation of the frequency-dependent electronic polarizabilities (Eq. 6.1) with respect to the vibrational normal mode Q gives rise to two kinds of terms for each excited state. The first term in Eq. 6.3 is proportional to the Cartesian derivative (gradient) of the excitation energy $\frac{\partial \Omega_k}{\partial Q}$. It may be compared to the A term in Albrecht’s approach, which arises from the energy differences between vibronic states in the energy denominator [190, 191]. By analogy, we will refer to these contributions as the A terms in the following. Only totally symmetric vibrational modes Q yield non-zero energy derivatives $\frac{\partial \Omega_k}{\partial Q}$, therefore A terms are only present for totally symmetric vibrations. The second term in Eq. 6.3 results from the dependence of transition moments μ_{0k}^m on the vibrational normal modes. In the language of Herzberg–Teller coupling [190, 197, 198], this dependence results from the interaction of the ground state or the electronic excited state k with other electronic states induced by nuclear displacements along the vibrational mode Q . The corresponding contributions are denoted B and C terms, respectively, in Albrecht’s approach. The terms in Eq. 6.3 that are proportional to derivatives of transition dipole moments $\frac{\partial \mu_{0k}^m}{\partial Q}$ have the same origin and hence will be referred to as B terms. B terms are non-zero for vibrational modes that transform like components of the polarizability tensor; the selection rules for the B term are equivalent to those for non-resonant Raman scattering [190, 191].

The frequency dependence of Raman spectra is defined by the molecular electronic excitation spectrum. In the strictly resonant case ($\omega = \Omega_k$) the excited electronic state k dominates the sum in Eq. 6.3. In this limit, the shape of the resonance Raman spectrum reflects the structure of the potential energy surface of the excited state k . Since the A term is quadratic in the resonance denominator $((\Omega_k - \omega)^2 + \gamma_k^2)^{-1}$, while

the B term is linear in it, the A term contribution can be expected to be predominant at resonance. In the opposite limiting case the excitation frequency is far from any electronic excitations (non-resonant Raman scattering), and a considerable number of electronic excited states contributes to the sum-over-states expression (Eq. 6.3.) The B term contributions become dominant in Raman cross sections, while the A terms are scaled down by their large energy denominators. Smooth interpolation between both limiting cases (non-resonant and strictly resonant) requires that both A and B terms be treated on equal footing.

Analytical derivative techniques allow to compute excitation energy gradients and non-resonant polarizability derivatives in an efficient fashion using TDDFT [206, 210]. In this work, derivatives of transition dipole moments are computed by numerical differentiation. However, an analytical implementation is possible starting from a Lagrangian formulation [211], similar to that for gradients of excitation energies [212, 213] and frequency-dependent polarizabilities [210].

6.3 Resonance Raman Spectra of nucleic acid bases

In the following, we explore the characteristic changes in resonance Raman spectra of guanosine for excitations in the range between 200–266 nm, which contains a number of electronic excitations. In addition, we consider Raman excitation profiles of ring-breathing modes of nucleosides. Raman excitation profiles describe the dependence of Raman cross sections on the energy of the incident radiation. Finally, we determine the relative contributions of the A and B terms to Raman cross sections of guanosine both at resonance and in the non-resonant case.

All calculations have been performed using the PBE0 functional [214] and triple-zeta valence basis sets with two sets of polarization functions (TZVPP) [215]. The PBE0 functional has been chosen because it has proven quite accurate both for polarizabilities [216, 217] and Raman intensities [210, 218]. However, vibrational frequencies [219] and electronic excitation energies [220] are often overestimated with PBE0. 20 excited electronic states were included in the sum-over-states expressions. Linewidth parameters were assumed to be 0.1 eV for all electronic states. All calculations were performed using the program package TURBOMOLE [221].

In Fig. 6.1(a)–(c), we compare experimental and computed resonance Raman spectra of guanosine at excitation wavelengths of 266 nm, 218 nm, and 200 nm. In addition, we show experimental and computed non-resonant Raman spectra of guanosine at 514.5 nm in Fig. 6.1(d). The experimental spectra are from Refs. 222 and 223. The considered range of excitation energies includes the two overlapping electronic absorption bands of guanosine observed experimentally at 4.4–4.6 eV and 4.8–5.1 eV [224–226]. Deconvolution of the UV absorption spectrum of guanosine in water yields 4.56 eV and 5.04 eV for the positions of the absorption maxima [224]. PBE0 predicts the two lowest electronic excited states of guanosine at 4.97 eV and 5.39 eV to be strongly allowed. At still higher excitation energies, a second pair of strongly allowed electronic absorption bands is observed experimentally [224, 226], with maxima at 6.17 eV and 6.67 eV, respectively. The computed excitation energies for these transitions are 6.79 eV and 6.99 eV. We refer to supplementary information for a full overview of computed and experimental excitation energies of guanosine. The overestimation of excitation energies observed here is quite typical for the PBE0

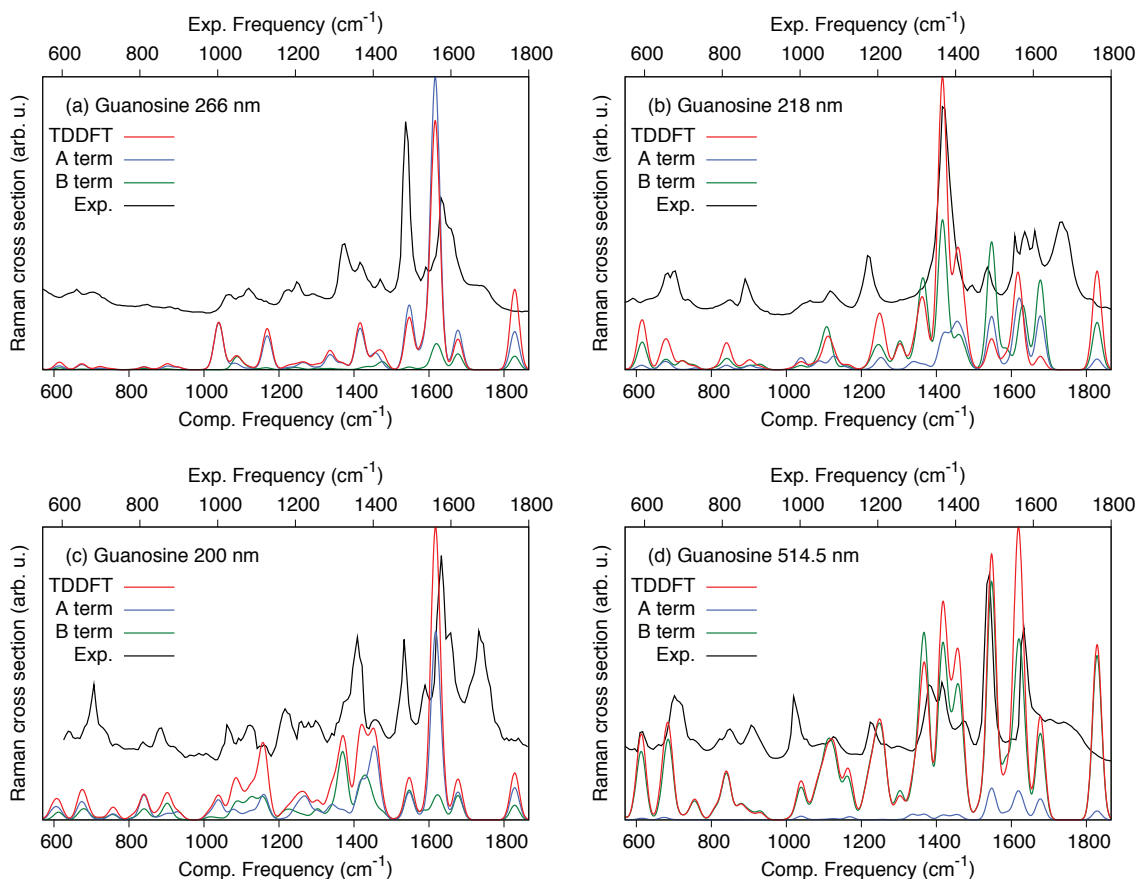


Figure 6.1: Experimental and computed Raman spectra of guanosine at 266, 218, 200, and 514.5 nm excitations. Experimental spectra of guanosine-5'-monophosphate (GMP) are from Refs. 222 and 223. Note that different frequency scales are applied to experimental and computed Raman spectra. See text for computational details.

functional [220] and is in part due to the lack of solvation effects in the calculations. Since the shape of resonance Raman spectra is sensitive to the relative position of the frequency of the incident light in the electronic excitation spectrum, we correct for the systematic error in excitation energies with PBE0. To this end, we first introduce a linear regression between the computed and experimental excitation energies based on the four strongly allowed electronic transitions of guanosine. The slope of the linear regression is 1.02, the offset is 0.35 eV. In addition, frequency scales in experimental and computed Raman spectra are adjusted in Fig. 6.1(a)–(d) to reflect the systematic overestimation of vibrational frequencies with PBE0 functional [219]. This corresponds to an effective scaling factor of 0.96.

The experimental resonance Raman spectrum at 266 nm excitation (Fig. 6.1(a)) is characterized by a strong 1492 cm^{-1} Raman peak and a slightly less intense 1581 cm^{-1} band. The former vibrational band was attributed to an imidazole ring vibration while the latter was assigned to a pyrimidine ring stretch mode [227]. A complete assignment of intensive Raman bands of guanosine is given in the supplementary information. To facilitate comparison between experimental and theoretical results, we compute the resonance Raman spectra at an excitation frequency shifted according to the linear regression results, see above. The experimental results obtained using 266 nm (4.66 eV) excitation are thus compared to computed Raman spectra at the 245 nm (5.07 eV) excitation. The computed Raman spectrum at 245 nm is dominated by contributions from the S_1 excited state. The strongest vibrational band is found at 1631 cm^{-1} and stems from the $\nu(\text{N}_7\text{--C}_8)$ bond stretch. The pyrimidine ring stretch is observed as a weaker band at 1547 cm^{-1} . Comparison with the resonance Raman

spectrum computed for the S_2 electronic excitation shows the opposite pattern, with a strong band at 1547 cm^{-1} and a somewhat less intense one at 1631 cm^{-1} . The predicted spectrum at resonance with the S_2 state seems to be in a better overall agreement with the experimental resonance Raman spectrum at 266 nm than the computed spectrum at resonance with the S_1 state, see supplementary information for more details. This findings underscore the importance of an accurate determination of the relative position of the frequency of the incident radiation relative to the electronic excitation spectrum of the molecule. The linear regression between experimental and computed excitations used here is perhaps the simplest possible correction scheme, while more rigorous approaches would contributions from the A terms, as is expected for an excitation close to resonance.

The experimental resonance Raman spectrum for the 218 nm excitation is characterized by a strong vibrational band at 1367 cm^{-1} assigned to an in-plane purine ring vibration. The $\nu(\text{C}_6=\text{O})$ Raman band is observed at 1685 cm^{-1} . The corresponding computed spectrum is obtained for the 203 nm (6.11 eV) excitation. The intermediate $\pi \rightarrow \pi^*$ excited state S_{10} of guanosine of low intensity (computed excitation energy 6.25 eV) has the largest contribution to the computed resonance Raman spectrum. It might be associated with the electronic transition observed at 215 nm (5.77 eV) in circular dichroism (CD) spectra of guanosine [228]. Due to the low oscillator strength of the S_{10} transition (0.05), the resonance Raman intensity is derived from both the A and the B terms. The strongest vibrational band in the computed resonance Raman spectrum at 203 nm excitation purine ring stretch mode predicted at 1416 cm^{-1} .

The experimental resonance Raman spectrum at 200 nm shows a strong pyrimidine

ring stretching band at 1578 cm^{-1} vibrational band as well as three Raman peaks of nearly equal intensity at 1679 cm^{-1} , 1489 cm^{-1} , and 1364 cm^{-1} , which are assigned to the $\nu(\text{C}_6=\text{O})$ stretch, a pyrimidine ring stretch, and an imidazole ring stretch, respectively. The low-frequency part of the resonance Raman spectrum is dominated by the ring breathing mode. The computed resonance Raman spectrum at 187 nm (6.63 eV) is close in energy to the strongly allowed $\pi \rightarrow \pi^*$ state (S_{13}) at 6.79 eV . The pyrimidine ring stretch vibration at 1631 cm^{-1} is predicted as the strongest vibrational band. The intensities of the $\nu(\text{C}_6=\text{O})$ vibration at 1829 cm^{-1} , the imidazole ring vibration at 1547 cm^{-1} , and the ring deformation mode at 1416 cm^{-1} , which correspond to the three intense Raman bands observed experimentally, are underestimated relative to the strongest Raman peak. Since the excitation at 200 nm is close to strict resonance, the A terms are dominant in the resonance Raman spectrum.

The non-resonant Raman spectrum of guanosine at 514.5 nm is shown in Fig. 6.1(d). Assignments of the non-resonant Raman spectra of guanine and its derivatives have been published previously [229–231]. As expected for Raman spectra far from resonance, the B terms are dominant, while the A terms are comparatively small. The non-resonant case is characterized by a significant number of excited electronic states, each contributing only a small amount to Raman cross sections. Under these circumstances, the closure of the sum over states is applicable, and the resulting Raman cross sections are represented as a ground state response property [191, 195]. The sum-over-states results for guanosine Raman spectra at 514.5 nm including 20 excited electronic states is in very good agreement with the conventional result obtained from

derivatives of frequency-dependent electronic polarizabilities, see supplementary information.

The changes observed in the experimental resonance Raman spectra can be well described within the sum-over-states formalism. A comprehensive assignment of Raman peaks can be achieved. The relative changes in resonance Raman spectra depend on the relative position of the frequency of the incident radiation within the electronic excitation spectrum. Thus a balanced description of a large number of electronic excitations is required, which represents a considerable challenge for the existing DFT methodology. Generally, the sum-over-states approach reproduces the characteristic changes in the overall shape of resonance Raman spectra reasonably well. This suggests that the main source of error in these calculations is due to electronic excitation energies, while the local properties of excited states, such as energy gradients and derivatives of transition dipole moments, are better reproduced. Similar results have been found for relaxed structures of excited states [206, 212, 213]. However, we note that all comparisons include relative Raman cross sections only. Accurate determination of absolute Raman cross sections is a challenging task both for experiments and computation and is not considered here.

Raman excitation profiles (REPs) describe the dependence of Raman scattering cross sections on excitation frequency. In Fig. 6.2 we show the REPs for the ring breathing modes of adenosine, guanosine, cytidine, and uridine. These low-frequency totally symmetric vibrational modes correspond to an in-phase expansion or contraction of the entire heteroaromatic ring system. Experimental spectra are from Ref. 232. For consistency, the correction for systematic errors in excitation energies derived for

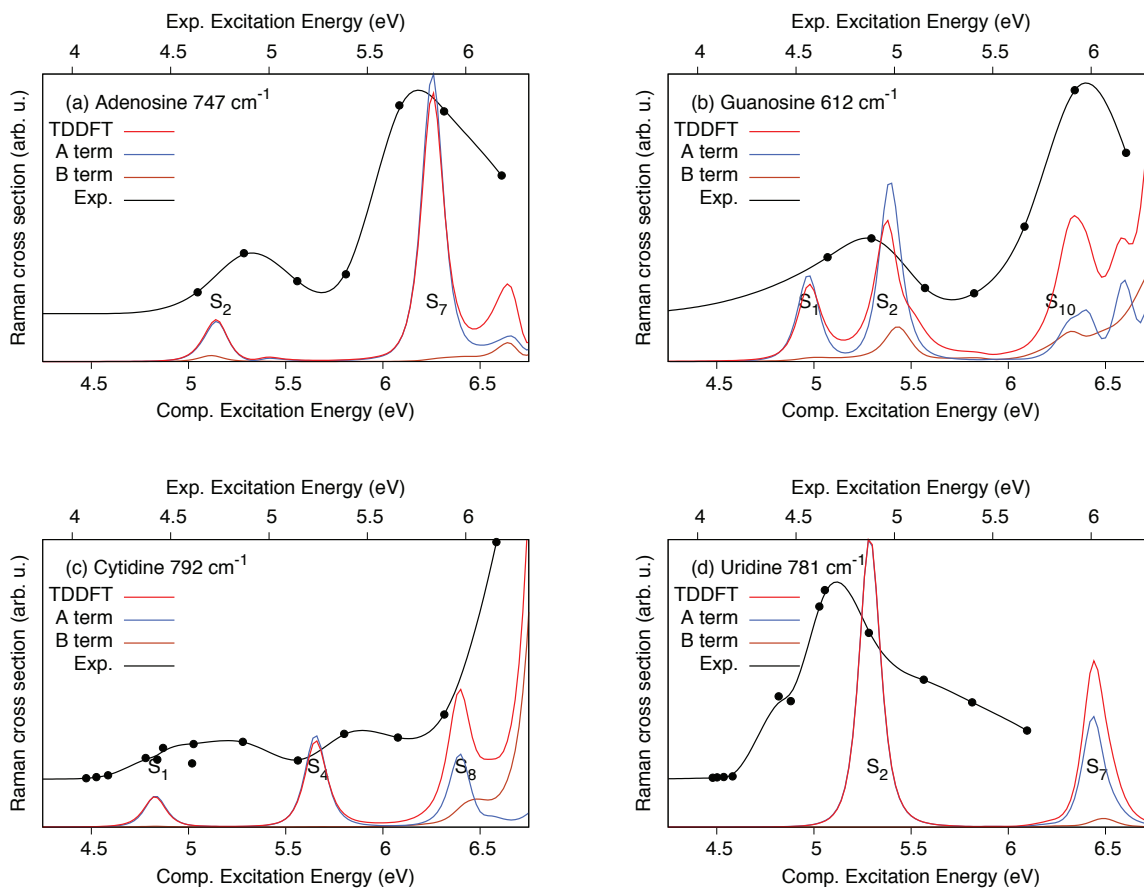


Figure 6.2: Experimental and computed Raman excitation profiles of ring breathing modes of nucleosides. Experimental data for nucleoside 5'-monophosphates are from Ref. 232. Solid lines in experimental data are obtained by interpolation and serve solely to guide the eye. Note that different energy scales are applied to experimental and computed Raman excitation profiles. See text for computational details.

guanosine (see above for details) is used for all nucleosides.

The ring breathing mode of adenosine (Fig. 6.2(a)) is observed at 729 cm^{-1} in experimental spectra, while the computed vibrational frequency is 747 cm^{-1} . The experimental REP shows two maxima at the positions of the electronic absorption bands of adenosine. They are assigned to the strongly allowed $\pi \rightarrow \pi^*$ excited states S_2 and S_7 , respectively. The Raman cross section of the ring breathing mode is larger at resonance with the higher-energy absorption band, in line with experimental data. Since the ring breathing vibration is totally symmetric, its intensity is almost entirely due to A term contributions.

The ring breathing mode of guanosine (Fig. 6.2(b)) is observed at 670 cm^{-1} ; the computed vibrational frequency is 612 cm^{-1} . Two broad maxima are observed in the experimental REP at the positions of the two electronic absorption bands. The first REP maximum at 4.5–5.0 eV covers the two closely lying dipole-allowed states S_1 and S_2 , while the second REP maximum peaked at ca. 6 eV includes the weakly allowed S_{10} state as well as the strongly absorbing S_{13} and S_{17} states. The larger Raman cross section at the second maximum is reproduced by theoretical results. The significant contribution from B terms, which grows with increasing excitation energy, suggests that the ring breathing modes is strongly coupled to non-totally symmetric vibrations.

Experimental and computed REPs of the ring breathing mode of cytidine are shown in Fig. 6.2(c). Experimental vibrational frequency is 782 cm^{-1} , the computed frequency is 792 cm^{-1} . Two moderately strong maxima are present in the experimental REP, followed by a significant increase at the high-energy edge of the REP. The two maxima are attributed to the two $\pi \rightarrow \pi^*$ transitions of cytidine (S_1 and

S_4). The increase at above 6.2 eV is due to the S_8 and higher-lying excited states. While the Raman cross sections are almost exclusively due to A terms at the first maxima, the contribution of B terms increases at higher excitation energies. The ring breathing mode of uridine shows a single broad peak at about 4.7 eV (Fig. 6.2(d)), which is assigned to the strongly allowed S_2 excited state. Experimental vibrational frequency of the ring breathing mode is 783 cm^{-1} , computed value is 781 cm^{-1} .

The two limiting cases of the sum-over-states expression are the strictly resonant situation, in which one resonant electronic state dominates the sum, with the A terms outweighing the corresponding B terms. In this case, the sum reduces to the short-time approximation [233]. The other limiting case, far from resonance, is usually well described by the polarizability theory of Placzek [195], in which polarizability derivatives are often even approximated by their static limits. As was pointed out by Albrecht, B terms are dominant in the non-resonant case [190], while A terms are all but negligible. The polarizability approximation is usually adequate for the range of excitation frequencies below the lowest electronic excitation. In the intermediate regime, e. g., above the first electronic excitation, both A and B terms from different electronic excited states contribute to Raman cross section, and a smooth interpolation, such as the one offered by the present approach, becomes necessary.

The presented sum-over-states approach ignores the details of vibronic structure and includes the contributions from a given electronic excited state in an aggregate manner only. Thus, it is likely to be problematic for molecules with well-resolved vibronic transitions such as small gas-phase species. However, vibronic structure is typically “washed out” in most medium-size and large molecules or in the presence

of a solvent so that the averaged description appears appropriate in these cases. The contributions of different electronic excited states to the Raman scattering tensor $\sigma_Q(\omega)$ are additive, cf. Eq. 6.2. Therefore, the quality of the description of the strictly resonant case may be improved upon by treating the contribution of the resonant electronic state in a more accurate way such as explicit time propagation [200, 201] or summation over vibronic states [199].

The computed resonance Raman cross sections include excitation energies, transition moments and their geometric derivatives. As a consequence, they offer a sensitive test of the TDDFT methodology. Our results indicate that largest source of error for relative resonance Raman cross sections are excitation energies, and the results are found to improve if the excitation energies are corrected for errors intrinsic to the method. Corrections using experimental excitation energies might be used for this purpose if available. Alternatively, corrections for excitation energies might be obtained from more accurate theoretical methods such as coupled-cluster response approaches.

6.4 Conclusion

In this work, we presented a simple approximation to resonance Raman cross section based on the sum-over-states expression for frequency-dependent electronic polarizabilities. Each electronic excited state contributes two types of terms to the Raman cross section, which we term A and B terms, in analogy to Albrecht’s treatment. The A terms are dominant in the strictly resonant case, while the B terms determine the Raman cross sections in the non-resonant limit. By using both terms,

the present method can treat the resonant and non-resonant cases on equal footing. Resonance Raman spectra and Raman excitation profiles of nucleosides can be predicted with reasonable accuracy using the sum-over-states approach. The major source of error seem to be electronic excitation energies, which are can be off by up to 0.5 eV with TDDFT. Improved description of resonance Raman spectra and Raman excitation profiles is expected from a combination of the present sum-over-states formulation with more accurate approaches for the few strictly resonant electronic states as well as an first-principles framework for computing electronic state linewidths from the open-system formulation of TDDFT [209].

Chapter 7

Summary and Future Directions

In this dissertation, quantum effects in biological systems were investigated by simulations at the molecular level. To study the effect of the protein environment on long-lived quantum coherences observed during the energy transfer process in the FMO complex, the protein complex embedding the chromophores were simulated in atomistic detail using molecular dynamics and TD-DFT. SC-IVR and DFT calculations were combined to give an accurate quantum dynamics. A formula for the resonance Raman spectra was developed using TD-DFT and analytic gradients and applied to the nucleic acid bases.

Our exciton propagation method with molecular dynamics and TD-DFT is a phenomenological model like the Haken-Strobl-Reineker method. The effect of the bath is included as classical stochastic terms in the system Hamiltonian, and the reduced density matrix is evaluated by averaging over the realization of quantum trajectories. Although very convenient in carrying out the propagation using stochastic simulation, this type of stochastic Liouville equation cannot reproduce the correct asymptotic

behavior at a finite temperature. Several propagation methods based on trajectories generated by stochastic generators [59, 99, 234–237] or real time path integral Monte Carlo [238–241] are known to produce exact dynamics. We expect that these stochastic methods are more appropriate than master equation based approaches for atomistic simulations. However, most of these stochastic approaches are obtained by mathematical unraveling of the master equation or influence functional rather than contemplating the physical system. Therefore, further investigations are needed to combine these methods with atomistic simulations.

Simulation of the protein environment is currently limited to classical mechanics due to the large degrees of freedom. As explored in Chapter 5, very accurate quantum dynamics can be obtained on top of classical mechanics using SC-IVR. Another possibility is to use mixed quantum-classical dynamics [242–245]. This formalism was developed to treat the dynamics of a quantum subsystem interacting with a classical bath by propagating the classical bath and the quantum reduced density matrix in phase space using the Wigner representation. However, both of these methods are not scalable enough to directly simulate the entire photosynthetic system. Other promising approaches would be to introduce the quantum correction factors to the classical bath correlation. These factors have been studied in the context of vibrational energy relaxation [69, 70, 246] and are expected to be straightforwardly applicable to the energy transfer dynamics.

Bibliography

- [1] P. A. M. Dirac, *P. R. Soc. Lond. A-Conta.* **123**, pp. 714 (1929).
- [2] J. A. McCammon, B. R. Gelin, and M. Karplus, *Nature* **267**, 585 (1977).
- [3] M. Karplus and J. A. McCammon, *Nat. Struct. Biol.* **9**, 646 (2002).
- [4] W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman, *Annu. Rev. Bioph. Biom.* **30**, 211 (2001).
- [5] T. Hansson, C. Oostenbrink, and W. van Gunsteren, *Curr. Opin. Struc. Biol.* **12**, 190 (2002).
- [6] M. Karplus and J. Kuriyan, *Proc. Natl Acad. Sci. USA* **102**, 6679 (2005).
- [7] E. Villa, A. Balaeff, L. Mahadevan, and K. Schulten, *Multiscale Model. Sim.* **2**, 527 (2004).
- [8] R. Phillips, M. Dittrich, and K. Schulten, *Ann. Rev. Mater. Res.* **32**, 219 (2002).
- [9] G. Ayton, S. G. Bardenhagen, P. McMurtry, D. Sulsky, and G. A. Voth, *J. Chem. Phys.* **114**, 6913 (2001).

- [10] B. Matthews, R. Fenna, M. Bolognesi, M. Schmid, and J. Olson, *J. Mol. Biol.* **131**, 259 (1979).
- [11] Y.-F. Li, W. Zhou, R. E. Blankenship, and J. P. Allen, *J. Mol. Biol.* **271**, 456 (1997).
- [12] J. Olson, *Photosynth. Res.* **80**, 181 (2004).
- [13] S. I. E. Vulto et al., *J. Phys. Chem. B* **102**, 10630 (1998).
- [14] J. Adolphs and T. Renger, *Biophys. J.* **91**, 2778 (2006).
- [15] A. Freiberg, S. Lin, K. Timpmann, and R. E. Blankenship, *J. Phys. Chem. B* **101**, 7211 (1997).
- [16] S. Savikhin, D. R. Buck, and W. S. Struve, *J. Phys. Chem. B* **102**, 5556 (1998).
- [17] M. Wendling et al., *J. Phys. Chem. B* **104**, 5825 (2000).
- [18] D. Tronrud, J. Wen, L. Gay, and R. Blankenship, *Photosynth. Res.* **100**, 79 (2009).
- [19] T. Förster, *Ann. Phys.* **437**, 55 (1948).
- [20] M. Wendling et al., *Photosynth. Res.* **71**, 99 (2002).
- [21] M. Cho, H. M. Vaswani, T. Brixner, J. Stenger, and G. R. Fleming, *J. Phys. Chem. B* **109**, 10542 (2005).
- [22] F. Müh et al., *Proc. Natl Acad. Sci. USA* **104**, 16862 (2007).
- [23] G. S. Engel et al., *Nature* **446**, 782 (2007).

- [24] H. Lee, Y.-C. Cheng, and G. R. Fleming, *Science* **316**, 1462 (2007).
- [25] G. Panitchayangkoon et al., *Proc. Natl Acad. Sci. USA* **107**, 12766 (2010).
- [26] P. Rebentrost, M. Mohseni, and A. Aspuru-Guzik, *J. Phys. Chem. B* **113**, 9942 (2009).
- [27] P. Rebentrost, M. Mohseni, I. Kassal, S. Lloyd, and A. Aspuru-Guzik, *New J. Phys.* **11**, 033003 (2009).
- [28] A. Ishizaki and G. R. Fleming, *J. Chem. Phys.* **130**, 234110 (2009).
- [29] A. Ishizaki and G. R. Fleming, *Proc. Natl Acad. Sci. USA* **106**, 17255 (2009).
- [30] G. Tao and W. H. Miller, *J. Phys. Chem. Lett.* **1**, 891 (2010).
- [31] P. Nalbach, A. Ishizaki, G. R. Fleming, and M. Thorwart, *New J. Phys.* **13**, 063040 (2011).
- [32] S. Shim, P. Rebentrost, S. Valleau, and A. Aspuru-Guzik, *Biophys. J.* **102**, 649 (2012).
- [33] T. C. Berkelbach, T. E. Markland, and D. R. Reichman, *J. Chem. Phys.* **136**, 084104 (2012).
- [34] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*, Oxford University Press, New York, 2002.
- [35] A. Croy and U. Saalman, *Phys. Rev. B* **80**, 073102 (2009).
- [36] J. Xu, R.-X. Xu, M. Luo, and Y. Yan, *Chem. Phys.* **370**, 109 (2010).

- [37] J. Hu, R.-X. Xu, and Y. Yan, *J. Chem. Phys.* **133**, 101106 (2010).
- [38] R. Kubo, *J. Phys. Soc. Jpn* **17**, 1100 (1962).
- [39] H. Haken and G. Strobl, *Z. Phys.* **262**, 135 (1973).
- [40] H. Haken and P. Reineker, *Z. Phys.* **249**, 253 (1972).
- [41] P. Reineker and R. Kuhne, *Z. Phys. B Con. Mat.* **22**, 193 (1975).
- [42] M. A. Palenberg, R. J. Silbey, C. Warns, and P. Reineker, *J. Chem. Phys.* **114**, 4386 (2001).
- [43] X. Chen and R. J. Silbey, *J. Chem. Phys.* **132**, 204503 (2010).
- [44] C. Aslangul and P. Kottis, *Phys. Rev. B* **10**, 4364 (1974).
- [45] R. E. Blankenship, *Molecular Mechanisms of Photosynthesis*, Wiley-Blackwell, 1st edition, 2002.
- [46] Y.-C. Cheng and G. R. Fleming, *Ann. Rev. Phys. Chem.* **60**, 241 (2009).
- [47] E. Collini et al., *Nature* **463**, 644 (2010).
- [48] M. B. Plenio and S. F. Huelga, *New J. Phys.* **10**, 113019 (2008).
- [49] M. Mohseni, P. Rebentrost, S. Lloyd, and A. Aspuru-Guzik, *J. Chem. Phys.* **129**, 174106 (2008).
- [50] S. Jang, Y.-C. Cheng, D. R. Reichman, and J. D. Eaves, *J. Chem. Phys.* **129**, 101104 (2008).

- [51] J. Wu, F. Liu, Y. Shen, J. Cao, and R. J. Silbey, *New J. Phys.* **12**, 105012 (2010).
- [52] M. Sarovar, A. Ishizaki, G. R. Fleming, and K. B. Whaley, *Nat. Phys.* **6**, 462 (2010).
- [53] F. Caruso, A. W. Chin, A. Datta, S. F. Huelga, and M. B. Plenio, *Phys. Rev. A* **81**, 062346 (2010).
- [54] F. Fassioli and A. Olaya-Castro, *New J. Phys.* **12**, 085006 (2010).
- [55] J. Cai, G. G. Guerreschi, and H. J. Briegel, *Phys. Rev. Lett.* **104**, 1 (2010).
- [56] A. Olaya-Castro, C. Lee, F. Olsen, and N. Johnson, *Phys. Rev. B* **78**, 7 (2008).
- [57] Q. Shi, L. Chen, G. Nan, R.-X. Xu, and Y. Yan, *J. Chem. Phys.* **130**, 084105 (2009).
- [58] J. Zhu, S. Kais, P. Rebentrost, and A. Aspuru-Guzik, *J. Phys. Chem. B* **115**, 1531 (2011).
- [59] J. Piilo, S. Maniscalco, K. Härkönen, and K.-A. Suominen, *Phys. Rev. Lett.* **100**, 180402 (2008).
- [60] P. Rebentrost, M. Mohseni, and A. Aspuru-Guzik, *J. Phys. Chem. B* **113**, 9942 (2009).
- [61] M. Tamoi et al., *J. Biol. Chem.* **285**, 15399 (2010).
- [62] P. Jahns, M. Graf, Y. Muneke, and T. Shikanai, *FEBS lett.* **519**, 99 (2002).

- [63] P. Pesaresi, D. Sandon, E. Giuffra, and R. Bassi, *FEBS Lett.* **402**, 151 (1997).
- [64] W. D. Cornell et al., *J. Am. Chem. Soc.* **117**, 5179 (1995).
- [65] M. Ceccarelli, P. Procacci, and M. Marchi, *J. Comp. Chem.* **24**, 129 (2003).
- [66] Y. Shao et al., *Phys. Chem. Chem. Phys.* **8**, 3172 (2006).
- [67] V. May and O. Kühn, *Charge and Energy Transfer Dynamics in Molecular Systems*, Wiley-VCH Verlag, Weinheim, 2004.
- [68] S. A. Egorov, K. F. Everitt, and J. L. Skinner, *J. Phys. Chem. A* **103**, 9494 (1999).
- [69] J. L. Skinner and K. Park, *J. Phys. Chem. B* **105**, 6716 (2001).
- [70] G. Stock, *Phys. Rev. Lett.* **102**, 118301 (2009).
- [71] J. C. Tully, *J. Chem. Phys.* **93**, 1061 (1990).
- [72] M. Ben-Nun, J. Quenneville, and T. J. Martínez, *J. Phys. Chem. A* **104**, 5161 (2000).
- [73] Y. Wu and M. F. Herman, *J. Chem. Phys.* **123**, 144106 (2005).
- [74] Y. C. Cheng and G. R. Fleming, *J. Phys. Chem. A* **112**, 4254 (2008).
- [75] S. Mukamel, *Principles of Nonlinear Optical Spectroscopy*, Oxford University Press, 1995.
- [76] I. L. Chuang and M. A. Nielsen, *J. Mod. Opt.* **44**, 2455 (1997).

- [77] J. F. Poyatos, J. I. Cirac, and P. Zoller, *Phys. Rev. Lett.* **78**, 390 (1997).
- [78] J. Yuen-Zhou, M. Mohseni, and A. Aspuru-Guzik, arXiv:1006.4866v3 (2010).
- [79] J. Yuen-Zhou and A. Aspuru-Guzik, arXiv:1101.2716v1 (2011).
- [80] E. C. G. Sudarshan, P. M. Mathews, and J. Rau, *Phys. Rev.* **121**, 920 (1961).
- [81] M. Cho, *Two Dimensional Optical Spectroscopy*, CRC Press, 2009.
- [82] M. Goldman, *Quantum Description of High-Resolution NMR in Liquids*, Oxford University Press, 1991.
- [83] N. S. Ginsberg, Y.-C. Cheng, and G. R. Fleming, *Acc. Chem. Res.* **42**, 1352 (2009).
- [84] D. M. Jonas, *Ann. Rev. Phys. Chem.* **54**, 425 (2003).
- [85] M. Cho, *Chem. Rev.* **108**, 1331 (2008).
- [86] B. Palmieri, D. Abramavicius, and S. Mukamel, *J. Chem. Phys.* **130**, 204512 (2009).
- [87] A. Ishizaki and G. R. Fleming, *J. Chem. Phys.* **130**, 234111 (2009).
- [88] J. Strumpfer and K. Schulten, *J. Chem. Phys.* **131**, 225101 (2009).
- [89] A. M. Virshup et al., *J. Phys. Chem. B* **113**, 3280 (2009).
- [90] F. Caruso, A. W. Chin, A. Datta, S. F. Huelga, and M. B. Plenio, *J. Chem. Phys.* **131**, 105106 (2009).

- [91] A. G. Dijkstra and Y. Tanimura, *New J. Phys.* **12**, 055005 (2010).
- [92] P. Rebentrost and A. Aspuru-Guzik, *J. Chem. Phys.* **134**, 101103 (2011).
- [93] M. Schmidt am Busch, F. Muh, M. El-Amine Madjet, and T. Renger, *J. Phys. Chem. Lett.* **2**, 93 (2011).
- [94] J. Gilmore and R. H. McKenzie, *J. Phys. Chem. A* **112**, 2162 (2008).
- [95] G. Ritschel, J. Roden, W. T. Strunz, A. Aspuru-Guzik, and A. Eisfeld, *J. Phys. Chem. Lett.* **2**, 2912 (2011).
- [96] J. Moix, J. Wu, P. Huo, D. Coker, and J. Cao, arXiv:1109.3416v1 (2011).
- [97] B. P. Krueger, G. D. Scholes, and G. R. Fleming, *J. Phys. Chem. B* **102**, 5378 (1998).
- [98] C.-P. Hsu, Z.-Q. You, and H.-C. Chen, *J. Phys. Chem. C* **112**, 1204 (2008).
- [99] J. Dalibard, Y. Castin, and K. Mølmer, *Phys. Rev. Lett.* **68**, 580 (1992).
- [100] Z. Vokcov and J. V. Burda, *J. Phys. Chem. A* **111**, 5864 (2007).
- [101] A. J. Leggett et al., *Rev. Mod. Phys.* **59**, 1 (1987).
- [102] R. M. Peralstein, *Theoretical Interpretation of Antenna Spectra*, CRC Press, New York, 1991.
- [103] A. Damjanović, I. Kosztin, U. Kleinekathöfer, and K. Schulten, *Phys. Rev. E* **65**, 031919 (2002).
- [104] C. Olbrich and U. Kleinekathofer, *J. Phys. Chem. B* **114**, 12427 (2010).

- [105] P. Huo and D. F. Coker, *J. Chem. Phys.* **133**, 184108 (2010).
- [106] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Application*, Cambridge University Press, 1993.
- [107] A. Nazir, *Phys. Rev. Lett.* **103**, 146404 (2009).
- [108] D. Abramavicius and S. Mukamel, *J. Chem. Phys.* **134**, 174504 (2011).
- [109] C. Olbrich, J. Strumpfer, K. Schulten, and U. Kleinekathofer, *J. Phys. Chem. B* **115**, 758 (2011).
- [110] C. Olbrich et al., *J. Phys. Chem. B* **115**, 8609 (2011).
- [111] G. S. Schlau-Cohen et al., *Nat. Chem.* **advance online publication**. (2012).
- [112] C. Y. Wong et al., *Nat. Chem.* **advance online publication**. (2012).
- [113] A. Kolli, E. J. O'Reilly, G. D. Scholes, and A. Olaya-Castro, arXiv:1203.5056v1 (2012).
- [114] A. G. Redfield, *IBM J. Res. Dev.* **1**, 19 (1957).
- [115] A. Ishizaki and Y. Tanimura, *J. Phys. Soc. Jpn* **74**, 3131 (2005).
- [116] M. Topaler and N. Makri, *J. Chem. Phys.* **97**, 9001 (1992).
- [117] N. Makri, *Chem. Phys. Lett.* **193**, 435 (1992).
- [118] R. Feynman and F. Vernon, *Ann. Phys.* **24**, 118 (1963).
- [119] C. Olbrich, J. Strümpfer, K. Schulten, and U. Kleinekathöfer, *J. Phys. Chem. B* **115**, 758 (2011).

- [120] D. Thirumalai, E. J. Bruskin, and B. J. Berne, *J. Chem. Phys.* **79**, 5063 (1983).
- [121] K. Allinger, B. Carmeli, and D. Chandler, *J. Chem. Phys.* **84**, 1724 (1986).
- [122] E. C. Behrman, G. a. Jongeward, and P. G. Wolynes, *J. Chem. Phys.* **83**, 668 (1985).
- [123] J. Cao and B. J. Berne, *J. Chem. Phys.* **99**, 2902 (1993).
- [124] J. M. Moix, Y. Zhao, and J. Cao, *Phys. Rev. B* **85**, 115412 (2012).
- [125] C. P. Robert and G. Casella, *Monte Carlo statistical methods.*, Springer Verlag, New York, 2004.
- [126] N. S. Pillai, A. M. Stuart, and A. H. Thiery, arXiv:1103.0542v2 (2011).
- [127] A. Aspuru-Guzik and W. A. Lester Jr., Quantum monte carlo methods for the solution of the schrödinger equation for molecular systems, in *Special Volume, Computational Chemistry*, edited by C. L. Bris, volume 10 of *Handbook of Numerical Analysis*, pages 485 – 535, Elsevier, 2003.
- [128] M. H. Alexander, *Chem. Phys. Lett.* **347**, 436 (2001).
- [129] J. M. Flegal and G. L. Jones, *Ann. Stat.* **38**, 1034 (2010).
- [130] G. M. Ljung and G. E. P. Box, *Biometrika* **65**, 297 (1978).
- [131] C. D. Schwieters and G. a. Voth, *J. Chem. Phys.* **111**, 2869 (1999).
- [132] J. R. Schmidt and J. C. Tully, *J. Chem. Phys.* **127**, 094103 (2007).
- [133] J. M. Herbert and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **7**, 3269 (2005).

- [134] R. Car and M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985).
- [135] H. B. Schlegel, J. M. Millam, S. S. Iyengar, G. A. Voth, A. D. Daniels, G. E. Scuseria, and M. J. Frisch, *J. Chem. Phys.* **114**, 9758 (2001).
- [136] J. M. Herbert and M. Head-Gordon, *J. Chem. Phys.* **121**, 11542 (2004).
- [137] Y. Liu, D. Yarne, M. E. Tuckerman, *Phys. Rev. B* **68**, 125110 (2003).
- [138] P. Tangney, *J. Chem. Phys.* **124**, 044111 (2006)
- [139] M. Pavese, D. R. Berard, and G. A. Voth, *Chem. Phys. Lett.* **300**, 93 (1999)
- [140] G. A. Worth, M. A. Robb, and I. Burghardt, *Faraday Discuss.* **127**, 307 (2004).
- [141] S. Iyengar and J. Jakowski, *J. Chem. Phys.* **122**, 114105 (2005).
- [142] O. Knospe and P. Jungwirth, *Chem. Phys. Lett.* **317**, 529 (2000).
- [143] W. H. Miller, *Adv. Chem. Phys.* **25**, 69 (1974); W. H. Miller, *Faraday Discuss.* **110**, 1 (1998)
- [144] W. H. Miller, *J. Chem. Phys.* **53**, 3578 (1970); *ibid.* **53**, 1949 (1970); W. H. Miller, *J. Phys. Chem. A* **105**, 2942 (2001); M. Thoss and H. Wang, *Annu. Rev. Phys. Chem.* **55**, 299 (2004); K. G. Kay, *Annu. Rev. Phys. Chem.* **56**, 255 (2005).
- [145] H. Wang, X. Sun, and W. H. Miller, *J. Chem. Phys.* **108**, 9726 (1998); X. Sun and W. H. Miller, *J. Chem. Phys.* **110**, 6635 (1999); M. Thoss, H. Wang, and W. H. Miller, *J. Chem. Phys.* **114**, 9220 (2001); T. Yamamoto, H. Wang, and

- W. H. Miller, *J. Chem. Phys.* **116**, 7335 (2002); T. Yamamoto W. H. Miller, *J. Chem. Phys.* **118**, 2135 (2003).
- [146] J. Ankerhold, M. Saltzer, and E. Pollak, *J. Chem. Phys.* **116**, 5925 (2002); S. Zhang and E. Pollak, *Phys. Rev. Lett.* **91**, 190201 (2003); S. S. Zhang and E. Pollak, *J. Chem. Phys.* **121**, 3384 (2004).
- [147] A. R. Walton, D. E. Manolopoulos, *Mol. Phys.* **87**, 961 (1996); A. R. Walton and D. E. Manolopoulos, *Chem. Phys. Lett.* **244**, 448 (1995); M. L. Brewer, J. S. Hulme, and D. E. Manolopoulos, *J. Chem. Phys.* **106**, 4832 (1997).
- [148] S. Bonella, D. Montemayor, and D. F. Coker, *Proc. Natl. Am. Soc.* **102**, 6715 (2005); S. Bonella and D. F. Coker, *J. Chem. Phys.* **118**, 4370 (2003).
- [149] Y. Wu , M. Herman, V. S. Batista, *J. Chem. Phys.* **122**, 114114 (2005); Y. Wu and V. S. Batista, *J. Chem. Phys.* **118**, 6720 (2003).
- [150] F. Grossmann, *Comment. At. Mol. Phys.* **34**, 243 (1999).
- [151] E. J. Heller, *J. Chem. Phys.* **62**, 1544 (1975); E. J. Heller, *J. Chem. Phys.* **75**, 2923 (1981).
- [152] E. J. Heller, *Acc. Chem. Res.* **14**, 368 (1981); E. J. Heller, *Acc. Chem. Res.* **39**, 127 (2006).
- [153] T. Van Voorhis and E. J. Heller, *J. Chem. Phys.* **119**, 12153 (2003).
- [154] D. V. Shalashilin and M. S. Child, *Chem. Phys.* **304**, 103 (2004); D. V. Shalashilin and M. S. Child, *J. Chem Phys.* **115**, 5367 (2001).

- [155] M. Ben-Nun and T. J. Martinez, *Adv. Chem. Phys.* **121**, 439 (2002).
- [156] M. F. Herman and E. Kluk, *Chem. Phys.* **91**, 27 (1984); K. G. Kay, *J. Chem. Phys.* **100**, 4377 (1994); K. G. Kay, *J. Chem. Phys.* **100**, 4432 (1994).
- [157] H. Wang, D. E. Manolopoulos, and W. H. Miller, *J. Chem. Phys.* **115**, 6317 (2001).
- [158] A. L. Kaledin and W. H. Miller, *J. Chem. Phys.* **118**, 7174 (2003); M. Ceotto, *PhD Dissertation*, University of California, Berkeley (2005); A. L. Kaledin and W. H. Miller, *J. Chem. Phys.* **119**, 3078 (2003).
- [159] Y. Elran and K. G. Kay, *J. Chem. Phys.* **110**, 3653 (1999); *ibid.* **110**, 8912 (1999).
- [160] Y. Shao, *et al. Phys. Chem. Chem. Phys.* **8**, 3172 (2006).
- [161] A.D. Becke, *J. Chem. Phys.* **98**, 5648 (1993); P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.* **98**, 11623 (1994).
- [162] T. Dunning Jr. *J. Chem. Phys.* **90**, 1007 (1989).
- [163] J. Zuniga, M. Alacid, A. Bastida, F. J. Carvajal, and A. Requena, *J. Mol. Spectr.* **195**, 137 (1999).
- [164] K. Levenberg, *Quart. Appl. Math.* **2**, 164 (1944); D. Marquardt, *Siam J. Appl. Math.* **11**, 431 (1965); M. I. A. Lourakis, Levenberg-Marquardt nonlinear least squares algorithms in C/C++, 2004.
Available from <http://www.ics.forth.gr/~lourakis/levmar/>.

- [165] M. H. Beck and H.-D. Meyer, *J. Chem. Phys.* **114**, 2036 (2001); G. A. Worth, M. H. Beck, A. Jäckle, and H.-D. Meyer, The MCTDH Package, Version 8.3, University of Heidelberg, Heidelberg, Germany, 2002.
Available from <http://www.pci.uni-heidelberg.de/tc/usr/mctdh/>.
- [166] F. Gygi, *Phys. Rev. B* **51**, 11190 (1995).
- [167] J. R. Chelikowsky, X. Jing, K. Wu, and Y. Saad, *Phys. Rev. B* **53**, 12071 (1994).
- [168] H. P. M. Filho, *Spectr. Acta Part A* **58**, 2621 (2002).
- [169] W. J. Hehre, R. Ditchfield, and J. A. Pople, *J. Chem. Phys.* **56**, 2257 (1972).
- [170] E. J. Heller, E. B. Stechel, and M. J. Davis, *J. Chem. Phys.* **73**, 4720 (1980).
- [171] X. Sun and W. H. Miller, *J. Chem. Phys.*, 1998, **108**, 8870.
- [172] A. M. N. Niklasson, C. J. Tymczak, and M. Challacombe, *Phys. Rev. Lett.* **97**, 123001 (2006); A. M. N. Niklasson, C. J. Tymczak, and M. Challacombe, *J. Chem. Phys.* **126**, 144103 (2007).
- [173] L. R. Brown and C. B. Farmer, *Appl. Opt.* **26**, 5154 (1987).
- [174] J. Tatchen and E. Pollak, *J. Chem. Phys.* **130**, 041103 (2009).
- [175] P. P. Shorygin, *Zh. Fiz. Khim.* **21**, 1125 (1947).
- [176] P. P. Shorygin and L. L. Krushinskij, *J. Raman Spectrosc.* **28**, 383 (1997).

- [177] F. S. Parker, *Applications of Infrared, Raman, and Resonance Raman Spectroscopy in Biochemistry*, Springer, New York, 1983.
- [178] T. G. Spiro, editor, *Biological Applications of Raman Spectroscopy*, volume 1–3, Wiley, New York, 1987.
- [179] J. M. Benevides, S. A. Overman, and G. J. Thomas, *J. Raman Spectrosc.* **36**, 279 (2005).
- [180] M. Moskovits, *Rev. Mod. Phys.* **57**, 783 (1985).
- [181] K. Kneipp, M. Moskovits, and H. Kneipp, editors, *Surface-Enhanced Raman Scattering: Physics and Applications*, Springer, Berlin, 2006.
- [182] L. Jensen, C. M. Aikens, and G. C. Schatz, *Chem. Soc. Rev.* **37**, 1061 (2008).
- [183] G. Schatz, M. Young, and R. Van Duyne, Electromagnetic mechanism of SERS, in *Surface-Enhanced Raman Scattering*, edited by K. Kneipp, M. Moskovits, and H. Kneipp, volume 103, pages 19–45, Springer, Berlin, 2006.
- [184] S. K. Saikin, R. Olivares-Amaya, D. Rappoport, M. Stopa, and A. Aspuru-Guzik, *Phys. Chem. Chem. Phys.* **11**, 9401 (2009).
- [185] S. K. Saikin, Y. Chu, D. Rappoport, K. B. Crozier, and A. Aspuru-Guzik, *J. Phys. Chem. Lett.*, 2740 (2010).
- [186] S. Nie and S. R. Emory, *Science* **275**, 1102 (1997).
- [187] K. Kneipp et al., *Phys. Rev. Lett.* **78**, 1667 (1997).
- [188] J. A. Dieringer et al., *J. Am. Chem. Soc.* **131**, 849 (2009).

- [189] P. Schorygin, L. Kuzina, and L. Ositjanskaja, *Microchim. Acta* **43**, 630 (1955).
- [190] A. C. Albrecht, *J. Chem. Phys.* **34**, 1476 (1961).
- [191] D. A. Long, *The Raman Effect*, Wiley, Chichester, 2nd edition, 2002.
- [192] H. A. Kramers and W. Heisenberg, *Z. Phys.* **31**, 681 (1925).
- [193] P. A. M. Dirac, *Proc. Roy. Soc. A* **114**, 710 (1927).
- [194] J. H. van Vleck, *Proc. Nat. Acad. Sci.* **15**, 754 (1929).
- [195] G. Placzek, Rayleigh-streuung und raman-effekt, in *Handbuch der Radiologie*, edited by E. Marx, volume VI/2, pages 209–374, Akademische Verlagsgesellschaft, Leipzig, 1934.
- [196] J. Behringer and J. Brandmüller, *Z. Elektrochem.* **60**, 643 (1956).
- [197] A. C. Albrecht, *J. Chem. Phys.* **33**, 156 (1960).
- [198] G. Fischer, *Vibronic Coupling: the Interaction between the Electronic and Nuclear Motions*, Academic Press, London, 1984.
- [199] K. A. Kane and L. Jensen, *J. Phys. Chem. C* **114**, 5540 (2010).
- [200] S.-Y. Lee and E. J. Heller, *J. Chem. Phys.* **71**, 4777 (1979).
- [201] D. J. Tannor, *J. Chem. Phys.* **77**, 202 (1982).
- [202] J. Neugebauer and B. A. Hess, *J. Chem. Phys.* **120**, 11564 (2004).
- [203] F. Neese, T. Petrenko, D. Ganyushin, and G. Olbrich, *Coord. Chem. Rev.* **251**, 288 (2007).

- [204] L. Jensen, J. Autschbach, and G. C. Schatz, *J. Chem. Phys.* **122**, 224115 (2005).
- [205] L. Jensen, L. L. Zhao, J. Autschbach, and G. C. Schatz, *J. Chem. Phys.* **123**, 174110 (2005).
- [206] F. Furche and D. Rappoport, Density functional methods for excited states: Equilibrium structure and electronic spectra, in *Computational Photochemistry*, edited by M. Olivucci, Theoretical and Computational Chemistry, pages 93–128, Elsevier, Amsterdam, 2005.
- [207] J. Yuen-Zhou, D. G. Tempel, C. A. Rodríguez-Rosario, and A. Aspuru-Guzik, *Phys. Rev. Lett.* **104**, 043001 (2010).
- [208] J. Yuen-Zhou, C. A. Rodríguez-Rosario, and A. Aspuru-Guzik, *Phys. Chem. Chem. Phys.* **11**, 4509 (2009).
- [209] D. G. Tempel, M. A. Watson, R. Olivares-Amaya, and A. Aspuru-Guzik, *J. Chem. Phys.* **134**, 074116 (2011).
- [210] D. Rappoport and F. Furche, *J. Chem. Phys.* **126**, 201104 (2007).
- [211] S. Coriani et al., *J. Chem. Theory Comput.* **6**, 1028 (2010).
- [212] F. Furche and R. Ahlrichs, *J. Chem. Phys.* **117**, 7433 (2002).
- [213] D. Rappoport and F. Furche, Excited states and photochemistry, in *Time-Dependent Density Functional Theory*, edited by M. A. L. Marques et al., chapter 23, pages 337–354, Springer, Berlin, 2006.

- [214] J. P. Perdew, M. Ernzerhof, and K. Burke, *J. Chem. Phys.* **105**, 9982 (1996).
- [215] A. Schäfer, C. Huber, and R. Ahlrichs, *J. Chem. Phys.* **100**, 5829 (1994).
- [216] C. Adamo, M. Cossi, G. Scalmani, and V. Barone, *Chem. Phys. Lett.* **307**, 265 (1999).
- [217] C. Van Caillie and R. D. Amos, *Chem. Phys. Lett.* **328**, 446 (2000).
- [218] C. Van Caillie and R. D. Amos, *Phys. Chem. Chem. Phys.* **2**, 2123 (2000).
- [219] J. P. Merrick, D. Moran, and L. Radom, *J. Phys. Chem. A* **111**, 11683 (2007).
- [220] C. Adamo, G. E. Scuseria, and V. Barone, *J. Chem. Phys.* **111**, 2889 (1999).
- [221] TURBOMOLE V6.2 2010, University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, 2007-.
Available from <http://www.turbomole.com/>.
- [222] S. P. A. Fodor, R. P. Rava, T. R. Hays, and T. G. Spiro, *J. Am. Chem. Soc.* **107**, 1520 (1985).
- [223] Y. Nishimura, M. Tsuboi, W. L. Kubasek, K. Bajdor, and W. L. Peticolas, *J. Raman Spectrosc.* **18**, 221 (1987).
- [224] L. B. Clark, *J. Am. Chem. Soc.* **116**, 5265 (1994).
- [225] M. K. Shukla and J. Leszczynski, *J. Comput. Chem.* **25**, 768 (2004).
- [226] M. Shukla and J. Leszczynski, *J. Biomol. Struct. Dyn.* **25**, 93 (2007).

- [227] A. Toyama, N. Hanada, J. Ono, E. Yoshimitsu, and H. Takeuchi, *J. Raman Spectros.* **30**, 623 (1999).
- [228] W. Voelter, R. Records, E. Bunnenberg, and C. Djerassi, *J. Am. Chem. Soc.* **90**, 6163 (1968).
- [229] M. Mathlouthi, A. M. Seuvre, and J. L. Koenig, *Carbohydr. Res.* **146**, 15 (1986).
- [230] J. Florián, *J. Phys. Chem.* **97**, 10649 (1993).
- [231] B. Giese and D. McNaughton, *Phys. Chem. Chem. Phys.* **4**, 5161 (2002).
- [232] M. Tsuboi, Y. Nishimura, and A. Y. Hirakawa, Resonance raman spectroscopy and normal modes of nucleic acid bases, in *Biological Applications of Raman Spectroscopy*, edited by T. G. Spiro, volume 2, pages 109–179, Wiley, New York, 1987.
- [233] D. J. Tannor, E. J. Heller, and R. Sundberg, *J. Phys. Chem.* **86**, 1822 (1982).
- [234] H.-P. Breuer, B. Kappler, and F. Petruccione, *Phys. Rev. A* **59**, 1633 (1999).
- [235] J. T. Stockburger and H. Grabert, *Chem. Phys.* **268**, 249 (2001).
- [236] J. Stockburger and H. Grabert, *Phys. Rev. Lett.* **88**, 2 (2002).
- [237] J. T. Stockburger, *Chem. Phys.* **296**, 159 (2004).
- [238] L. Mühlbacher, J. Ankerhold, and C. Escher, *J. Chem. Phys.* **121**, 12696 (2004).
- [239] L. Mühlbacher and J. Ankerhold, *J. Chem. Phys.* **122**, 184715 (2005).
- [240] L. Mühlbacher and E. Rabani, *Phys. Rev. Lett.* **100**, 1 (2008).

- [241] O. Mülken, L. Mühlbacher, T. Schmid, and A. Blumen, *Phys. Rev. E* **81**, 1 (2010).
- [242] R. Kapral and G. Ciccotti, *J. Chem. Phys.* **110**, 8919 (1999).
- [243] R. Kapral, *J. Phys. Chem. A* **105**, 2885 (2001).
- [244] M. Toutounji and R. Kapral, *Chem. Phys.* **268**, 79 (2001).
- [245] M. Toutounji, *J. Chem. Phys.* **123**, 244102 (2005).
- [246] J. L. Skinner and D. Hsu, *J. Phys. Chem.* **90**, 4931 (1986).