ПРЕДГОВОР

Тази книга съдържа лекциите по Числени методи I част, които чета на студентите от Факултета по математика и информатика при Софийски Университет "Св. Климент Охридски". В това издание са направени някои незначителни промени, поправени са печатни грешки и са добавени задачи за домашна работа.

Благодарен съм на колегите си проф. дмн Стефка Димова, доц. дмн Гено Николов, доц. Никола Найденов и ас. Иван Христов, които ми обърнаха внимание върху някои неточности и печатни грешки в предишното издание. Благодаря на гл.ас. Веселин Гушев за помощта при предпечатната подготовка.

13 април 2008 Борислав Боянов

Глава 1

ПРИБЛИЖАВАНЕ НА ФУНКЦИИ

Една основна задача в числените методи е задачата за приближаване на "сложни" функции с други, "по-прости". Под сложни функции ще разбираме функции, които предизвикват някои проблеми при компютърни пресмятания. Това могат да бъдат сложни аналитични изрази, чието числено пресмятане отнема много компютърно време или се съпътства с грешки от закръгляне. Към "сложните" функции ще причисляваме и функциите, които са зададени таблично, т.е. чрез таблица от аргументи $x_1, \ldots, x_n$ и съответстващи им стойности $y_1, \ldots, y_n$. Обикновено в такъв вид се представят функциите в задачи, идващи от практиката. В резултат на експерименти и измервания се получава приближено стойността y_k на функцията $f(x)$ при $x = x_k$. Понякога функцията зададена по този начин трябва да бъде диференцирана, интегрирана или подлагана на друг вид операции. Ясно е, че поради непълната информация за $f(x)$, получаването на относително точен числен резултат е свързано с определени проблеми. Затова и табличните функции ще третираме като сложни функции.

Под "прости функции" в този курс ще разбираме преди всичко алгебричните полиноми.

Агебричен полином от степен n е всеки израз от вида

$$p(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n,$$

където $a_0, \ldots, a_n$ са реални числа. Известен е много прост метод (наречен *правило на Хорнер*) за пресмятане стойността на полинома в точката x. Това става като се извършат n-те умножения и n-те събирания, в реда показан по-долу

$$p(x) = (\ldots(((a_0.x + a_1).x + a_2).x + a_3)\ldots + a_{n-1}).x + a_n.$$

Полиномите се диференцират и интегрират лесно. Те имат много интересни свойства, които са добре изучени. Ето защо полиномите се приемат като прости, "хубави" функции.

В този курс с π_n ще означаваме класа от всички алгебрични полиноми от степен по-малка или равна на n.

Друг клас от прости функции са тригонометричните полиноми. Напомняме, че всеки израз от вида

$$t_n(x) = a_0 + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx)$$

се нарича *тригонометричен полином* от ред n. Вижда се, че $t_n(x) = t_n(x + 2\pi)$ за всяко x. С други думи, тригонометричният полином $t_n(x)$ е 2π-периодична функция. Можем да си мислим, че $t_n(x)$ е определен върху единичната окръжност (окръжността с център нула и радиус 1). Тогава много от свойствата му се възприемат по-лесно. Подобно на алгебричните полиноми и тригонометричните се диференцират и интегрират точно. Те се използват най-вече за приближаване на сложни функции, които описват периодични явления. По-нататък ще разгледаме и други класове от прости функции.

Остана да изясним какво разбираме под израза "да приближим" една функция с друга. Естествено, това зависи от критерия за близост, който ще възприемем. В числените методи се използват най-разнообразни критерии за близост, но те по същество се разделят на две групи: *интерполационни* и *метрични* критерии. При интерполационните критерии се избират краен брой характеристики $L_1(f), \ldots, L_n(f)$ на функцията f (най-често това са линейни функционали, да речем $L_k(f) := f(x_k), k = 1, \ldots, n$) и две функции се сравняват по тези характеристики. Казваме, че f е близка до g, ако $L_k(f)$ съвпада с $L_k(g)$ за $k = 1, \ldots, n$.

При метричните критерии се използва понятието *разстояние* (*метрика*). Казваме, че в пространството $\mathcal{F}$ от функции е въведено разстояние, ако на всеки два елемента f и g от $\mathcal{F}$ се съпоставя число $\rho(f, g)$ и това съответствие удовлетворява следните условия:

$$
\begin{array}{ll}
1) & \rho(f,g) \geq 0 \quad (=) \iff f = g; \\
2) & \rho(f,g) = \rho(g,f), \quad \forall f,g \in \mathcal{F}; \\
3) & \rho(f,g) \leq \rho(f,h) + \rho(h,g), \quad \forall f,g,h \in \mathcal{F}.
\end{array}
$$

Тогава, естествено, функциите f и g от $\mathcal{F}$ са близки, ако разстоянието $\rho(f, g)$ е "малко".

1. ИНТЕРПОЛАЦИОННА ФОРМУЛА НА ЛАГРАНЖ

Ще разгледаме следната интерполационна задача.

Нека $x_0, \ldots, x_n$ са различни точки и $y_0, \ldots, y_n$ са дадени реални числа. Да се построи алгебричен полином $P(x)$ от степен $\le n$, който удовлетворява условията

$$(1) \qquad P(x_k) = y_k, \quad k = 0, \ldots, n.$$

С други думи, при дадени $n+1$ точки $\{(x_k, y_k)\}_{k=0}^{n}$ в равнината, да се построи полином P от степен n, чиято графика минава през дадените точки (x_k, y_k), $k = 0, \ldots, n$.

Да отбележим най-напред, че ако изобщо съществува решение на интерполационната задача (1), то трябва да е единствено. Наистина, да допуснем, че има два полинома P и Q от степен n, които удовлетворяват (1). Тогава тяхната разлика

$$R(x) := P(x) - Q(x)$$

ще бъде също полином от степен $\le n$ и освен това

$$R(x_k) = P(x_k) - Q(x_k) = y_k - y_k = 0$$

за $k = 0, \ldots, n$. И така, R е полином от степен n, който се анулира в $n+1$ точки. Тогава, по основната теорема на алгебрата, $R(x)$ е тъждествено равен на 0. Следователно $P \equiv Q$.

Съществуването и единствеността на решението на (1) се виждат и по следния начин. Да запишем полинома $P(x)$ в общия му вид

$$P(x) = a_0 x^n + \cdots + a_n.$$

Тогава условието (1) добива вида

$$\begin{array}{rcl} a_0 x_0^n + a_1 x_0^{n-1} + \cdots + a_{n-1} x_0 + a_n & = & y_0 \\ a_0 x_1^n + a_1 x_1^{n-1} + \cdots + a_{n-1} x_1 + a_n & = & y_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \cdots & \cdots \\ a_0 x_n^n + a_1 x_n^{n-1} + \cdots + a_{n-1} x_n + a_n & = & y_n. \end{array}$$

Това е една система от $n+1$ линейни уравнения по отношение на неизвестните $a_0, \ldots, a_n$. Детерминантата

$$V(x_0, \ldots, x_n) := \det \begin{bmatrix} x_0^n & \ldots & x_0 & 1 \\ x_1^n & \ldots & x_1 & 1 \\ \vdots & \ldots & \vdots & \vdots \\ x_n^n & \ldots & x_n & 1 \end{bmatrix}.$$

на тази линейна система е детерминанта на Вандермонд. А ние знаем от линейната алгебра, че детерминантата на Вандермонд съответстваща на точките $x_0, \dots, x_n$ е различна от нула, ако $x_i \neq x_j$ при $i \neq j$. Тъй като по условие точките $x_0, \dots, x_n$ в (1) са различни, $V(x_0, \dots, x_n) \neq 0$ и следователно системата, а оттук и задачата (1), има единствено решение.

Извеждане на формулата. Особено важен за нас е въпросът за построяване на полинома P, който решава интерполационната задача.

Решението на (1) е било дадено в явен вид за първи път от Нютон. Ние тук ще дадем най-напред формулата за построяване на P, изведена от Лагранж, а по-късно ще представим и решението на Нютон.

Тъй като единствеността на решението на P е очевидна, то Лагранж пристъпва направо към построяването на това решение по следния остроумен начин. При фиксирано k да намерим полинома $l_{nk}(x)$ от π_n, който удовлетворява условията

$$\begin{aligned} l_{nk}(x_i) &= 0 \quad \text{за} \quad i = 0, \dots, n, \quad i \neq k, \\ l_{nk}(x_k) &= 1. \end{aligned}$$

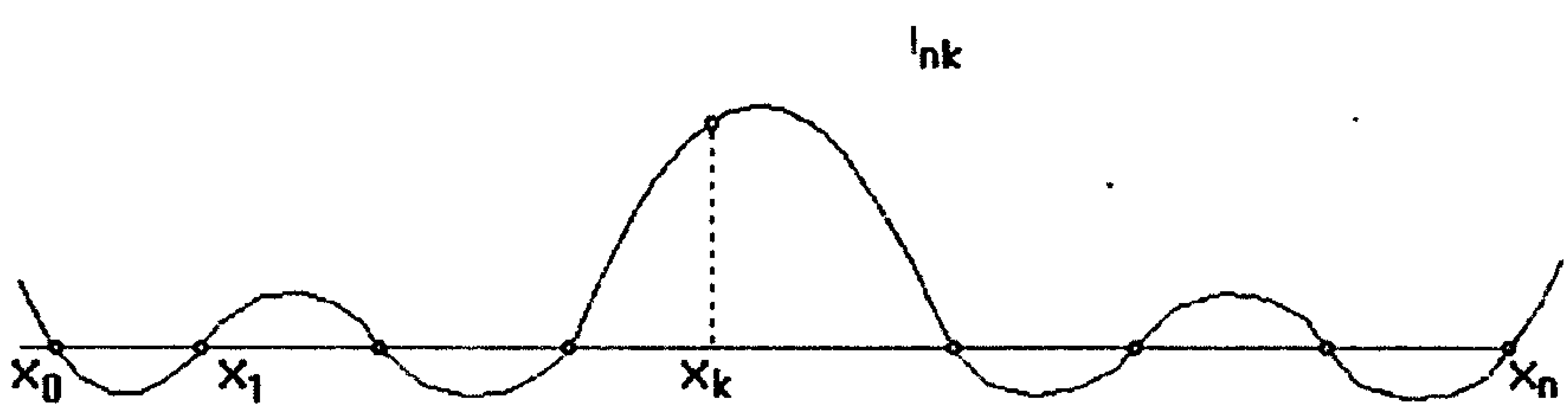

Рис. 1

Първото условие означава, че точките $x_0, \dots, x_{k-1}, x_{k+1}, \dots, x_n$ са нули на l_{nk} (виж Рис. 1). Но те са точно n на брой и l_{nk} е полином от степен n. Следователно това са всичките нули на l_{nk}. Тогава l_{nk} може да се запише така

$$l_{nk}(x) = A(x - x_0) \dots (x - x_{k-1})(x - x_{k+1}) \dots (x - x_n),$$

където A е някакво число. Това число ще определим от последното условие $l_{nk}(x_k) = 1$. Имаме

$$1 = l_{nk}(x_k) = A(x_k - x_0) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n).$$

Следователно

$$A = \frac{1}{\prod\limits_{i=0, i\neq k}^{n} (x_k - x_i)}$$

и окончателно

(2) $$l_{nk}(x) = \frac{(x-x_0)\dots(x-x_{k-1})(x-x_{k+1})\dots(x-x_n)}{(x_k-x_0)\dots(x_k-x_{k-1})(x_k-x_{k+1})\dots(x_k-x_n)}.$$

Полиномите $\{l_{nk}\}_{k=0}^{n}$ се наричат *базисни полиноми* на Лагранж. С тяхна помощ може лесно да се построи решението P на интерполационната задача (1). Ще покажем, че решението $P(x)$ на (1) се дава с формулата

(3) $$P(x) = \sum_{k=0}^{n} y_k \, l_{nk}(x).$$

Наистина, по построение

$$l_{nk}(x_i) = \delta_{ki} := \begin{cases} 1 & \text{при} \quad k = i \\ 0 & \text{при} \quad k \neq i. \end{cases}$$

Тогава

$$P(x_i) = \sum_{k=0}^{n} y_k \, l_{nk}(x_i) = y_i \, l_{ni}(x_i) = y_i.1 = y_i$$

за всяко $i = 0, 1, \dots, n$. И така, полиномът (3) е от π_n (защото $l_{nk} \in \pi_n$ за всяко k) и удовлетворява интерполационните условия (1). Следователно $P(x)$, даден в (3), е решение на интерполационната задача (1).

Най-често $\{y_k\}_{k=0}^{n}$ са стойности на някаква функция $f(x)$ в точките $x_0, \dots, x_n$, т.е.

$$y_k = f(x_k), \quad k = 0, \dots, n.$$

В такъв случай решението на интерполационната задача

$$P(x_k) = f(x_k), \quad k = 0, \dots, n,$$

се бележи с $L_n(f; x)$ и се нарича *интерполационен полином по Лагранж* за функцията f с *възли* $x_0, \dots, x_n$. Казваме още, че $L_n(f; x)$ *интерполира* $f(x)$ в точките $x_0, \dots, x_n$.

И така, доказахме следната теорема.

Теорема 1. *Нека $x_0 < \dots < x_n$ и $f(x)$ е определена в тези точки. Тогава съществува единствен полином от π_n, който интерполира f в $x_0, \dots, x_n$. Този полином се представя по формулата :*

(4) $$L_n(f; x) = \sum_{k=0}^{n} f(x_k) \prod_{i=0, i\neq k}^{n} \frac{x - x_i}{x_k - x_i}.$$

Твърдението следва веднага от (3) като вземем предвид, че съгласно (2).

$$l_{nk}(x) = \prod_{i=0, i\neq k}^{n} \frac{x - x_i}{x_k - x_i}.$$

Формулата (4) се нарича *интерполационна формула на Лагранж* .

Понякога ще използваме един по-кратък запис за l_{nk}. Той следва от връзката

$$\omega'(x_k) = (x_k - x_0)\dots(x_k - x_{k-1})(x_k - x_{k+1})\dots(x_k - x_n),$$

където

$$\omega(x) := (x - x_0)\dots(x - x_n).$$

Тази връзка се проверява директно като се диференцира $\omega(x)$ и се постави $x = x_k$. Тогава очевидно

$$l_{nk}(x) = \frac{\omega(x)}{(x - x_k)\,\omega'(x_k)}$$

и следователно

$$L_n(f;x) = \sum_{k=0}^{n} f(x_k)\frac{\omega(x)}{(x - x_k)\,\omega'(x_k)}.$$

Оценка на грешката. Обикновено интерполационният полином $L_n(f;x)$ се използва за приближение на по-сложна функция $f(x)$. Тогава възниква въпросът: Какво можем да кажем за грешката при това приближение, т.е. какво можем да кажем за разликата

$$R_n(f;x) := f(x) - L_n(f;x)$$

в някакво отнапред избрано x?

Да обърнем внимание, че полиномът $L_n(f;x)$ беше построен само въз основа на точките $\{(x_k, f(x_k))\}_{k=0}^n$. Но през тези същите точки минават графиките на безброй други непрекъснати функции $g(x)$ и очевидно за тях ще имаме $L_n(g;x) \equiv L_n(f;x)$. При това, за всяко дадено число $C > 0$, можем да построим непрекъсната функция g от разглеждания клас - такава, че $g(x) - L_n(f;x) \geq C$. Следователно грешката може да бъде произволно голяма, ако нищо не знаем за функцията освен това, че е непрекъсната. Затова в следващата теорема се налага едно допълнително условие за гладкост на f.

Теорема 1. *Нека $[a,b]$ е даден краен интервал и $x_0,\dots,x_n$ са различни точки в него. Нека функцията $f(x)$ има непрекъсната $(n+1)$-ва производна в $[a,b]$. Тогава за всяко $x \in [a,b]$ съществува точка $\xi \in [a,b]$ такава, че*

$$f(x) - L_n(f;x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)\dots(x-x_n).$$

(По-точно, $\xi \in (\min\{x,x_0,\dots,x_n\},\ \max\{x,x_0,\dots,x_n\})$.)

Доказателство. Да образуваме помощната функция

$$F(t) = f(t) - L_n(f;t) - C(t-x_0)\dots(t-x_n),$$

където C е параметър. Веднага се вижда, че $F(t)$ се анулира в точките $x_0,\dots,x_n$ при всеки избор на C. Наистина

$$F(x_k) = f(x_k) - L_n(f;x_k) - C.0 = f(x_k) - f(x_k) = 0.$$

Сега да изберем C така, че $F(t)$ да се анулира и в точката $t = x$. От равенството

$$f(x) - L_n(f;x) - C(x-x_0)\dots(x-x_n) = 0$$

определяме

$$C = \frac{R_n(f;x)}{(x-x_0)\dots(x-x_n)}. \tag{5}$$

И така, при този избор на C функцията $F(t)$ има $n+2$ нули. Това са точките $x,x_0,\dots,x_n$. По теоремата на Рол $F'(t)$ ще има поне $n+1$ нули, които лежат в интервала $(\min\{x,x_0,\dots,x_n\},\ \max\{x,x_0,\dots,x_n\})$, $F''(t)$ ще има поне n нули, и т.н., $F^{(n+1)}(t)$ ще има поне една нула, която лежи в $(\min\{x,x_0,\dots,x_n\},\ \max\{x,x_0,\dots,x_n\})$. Да я означим с ξ. Имаме $F^{(n+1)}(\xi) = 0$. От друга страна,

$$\begin{aligned} F^{(n+1)}(\xi) &= f^{(n+1)}(\xi) - L_n^{(n+1)}(f;\xi) - C(n+1)! \\ &= f^{(n+1)}(\xi) - C(n+1)! \end{aligned}$$

Следователно

$$C = \frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

Като сравним това равенство с (5) получаваме

$$R_n(f;x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)\dots(x-x_n).$$

Теоремата е доказана.

2. ПОЛИНОМИ НА ЧЕБИШОВ

От доказаната в предишната лекция теорема следва, че за всяко $x \in [a, b]$,

$$|f(x) - L_n(f; x)| \leq \frac{M_{n+1}}{(n+1)!} \max_{x \in [a,b]} |(x - x_0) \ldots (x - x_n)|,$$

където M_{n+1} е горна граница на $|f^{(n+1)}(x)|$ в $[a, b]$. Оттук се вижда, че оценката на грешката при приближаване с интерполационния полином на Лагранж зависи съществено от избора на интерполационните възли $x_0, \ldots, x_n$, тъй като величината

$$\max_{x \in [a,b]} |(x - x_0) \ldots (x - x_n)|$$

зависи от тях. Така възниква следната екстремална задача:

Да се намерят тези точки $\{x_k^*\}_{k=0}^n$, $a \leq x_0^* < \cdots < x_n^* \leq b$, *при които*

$$\max_{a \leq x \leq b} |(x - x_0^*) \ldots (x - x_n^*)| = \inf_{a \leq x_0 < \cdots < x_n \leq b} \max_{a \leq x \leq b} |(x - x_0) \ldots (x - x_n)|.$$

С други думи, трябва да се намери полином от вида $(x - x_0) \cdots (x - x_n)$, който се отклонява минимално от нулата в $[a, b]$. Решението на тази задача се дава чрез така наречените *полиноми на Чебишов от първи род.*

Полиномът на Чебишов от първи род от n-та степен се бележи обикновено с $T_n(x)$ и се определя в интервала $[-1, 1]$ чрез равенството

$$T_n(x) = \cos(n \arccos x), \quad x \in [-1, 1] . \tag{1}$$

Ще покажем най-напред, че изразът (1) е наистина полином от степен n. Непосредствено от определението следва, че

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= \cos(\arccos x) = x. \end{aligned}$$

Освен това, съгласно формулата за събиране на косинуси,

$$\begin{aligned} T_{n+1}(x) + T_{n-1}(x) &= \cos((n+1) \arccos x) + \cos((n-1) \arccos x) \\ &= 2 \cos(\arccos x) . \cos(n \arccos x) \\ &= 2x T_n(x) \end{aligned}$$

при всяко $n \geq 1$. Оттук получаваме рекурентната връзка

$$T_{n+1}(x) = 2x T_n(x) - T_{n-1}(x). \tag{2}$$

С нейна помощ можем да построим в явен вид следващите няколко полинома на Чебишов. Получаваме

$$\begin{aligned} T_2(x) &= 2xT_1(x) - T_0(x) = 2x.x - 1 = 2x^2 - 1, \\ T_3(x) &= 2xT_2(x) - T_1(x) = 4x^3 - 3x. \end{aligned}$$

Аналогично

$$\begin{aligned} T_4(x) &= 8x^4 - 8x^2 + 1\ , \\ T_5(x) &= 16x^5 - 20x^3 + 5x\ , \\ T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1\ , \\ T_7(x) &= 64x^7 - 112x^5 + 56x^7 - 7x\ . \end{aligned}$$

От рекурентната връзка се вижда, че коефициентът пред x^n в $T_n(x)$ се получава от коефициента пред x^{n-1} в $T_{n-1}(x)$ чрез умножение с 2. Тъй като $T_1(x) = 2^0x$, то $T_n(x)$ ще бъде от вида

$$T_n(x) = 2^{n-1}x^n + \cdots$$

И така, показахме, че $T_n(x)$ е алгебричен полином от степен n с коефициент 2^{n-1} пред x^n. Сега да отбележим и други интересни свойства на $T_n(x)$. От определението (1) веднага следва, че

(3) $$|T_n(x)| \le 1 \quad \text{за всяко } x \in [-1, 1]\ .$$

При това равенство се достига само за тези точки x от $[-1, 1]$, за които

$$|\cos(n \arccos x)| = 1,$$

т.е. при

$$n \arccos x = k\pi, \quad \text{където } k \text{ е цяло число.}$$

От това уравнение определяме екстремалните точки η_k на $T_n(x)$ в $[-1, 1]$ (т.е. точките η_k, за които $|T_n(\eta_k)| = 1$). Получаваме $\eta_k = \cos\frac{k\pi}{n}$. Когато k приема всички цели стойности, η_k описва циклично само $n + 1$ различни точки. Следователно всички екстремални точки на T_n в $[-1, 1]$ се дават с формулата

$$\eta_k = \cos\frac{k\pi}{n}, \quad k = 0, \ldots, n.$$

Директно се проверява, че

(4) $$T_n(\eta_k) = (-1)^k, \quad k = 0, \ldots, n.$$

Полиномите на Чебишов имат твърде интересно поведение в интервала $[-1,1]$ (вж. Рис. 2). Графиката на $T_n(x)$ лежи изцяло в квадрата $[-1,1] \times [-1,1]$, като се допира алтернативно в точките η_k до правите $y = 1$ и $y = -1$. Казваме, че $T_n(x)$ осъществява *алтернанс* в точките $\{\eta_k\}_{k=0}^n$.

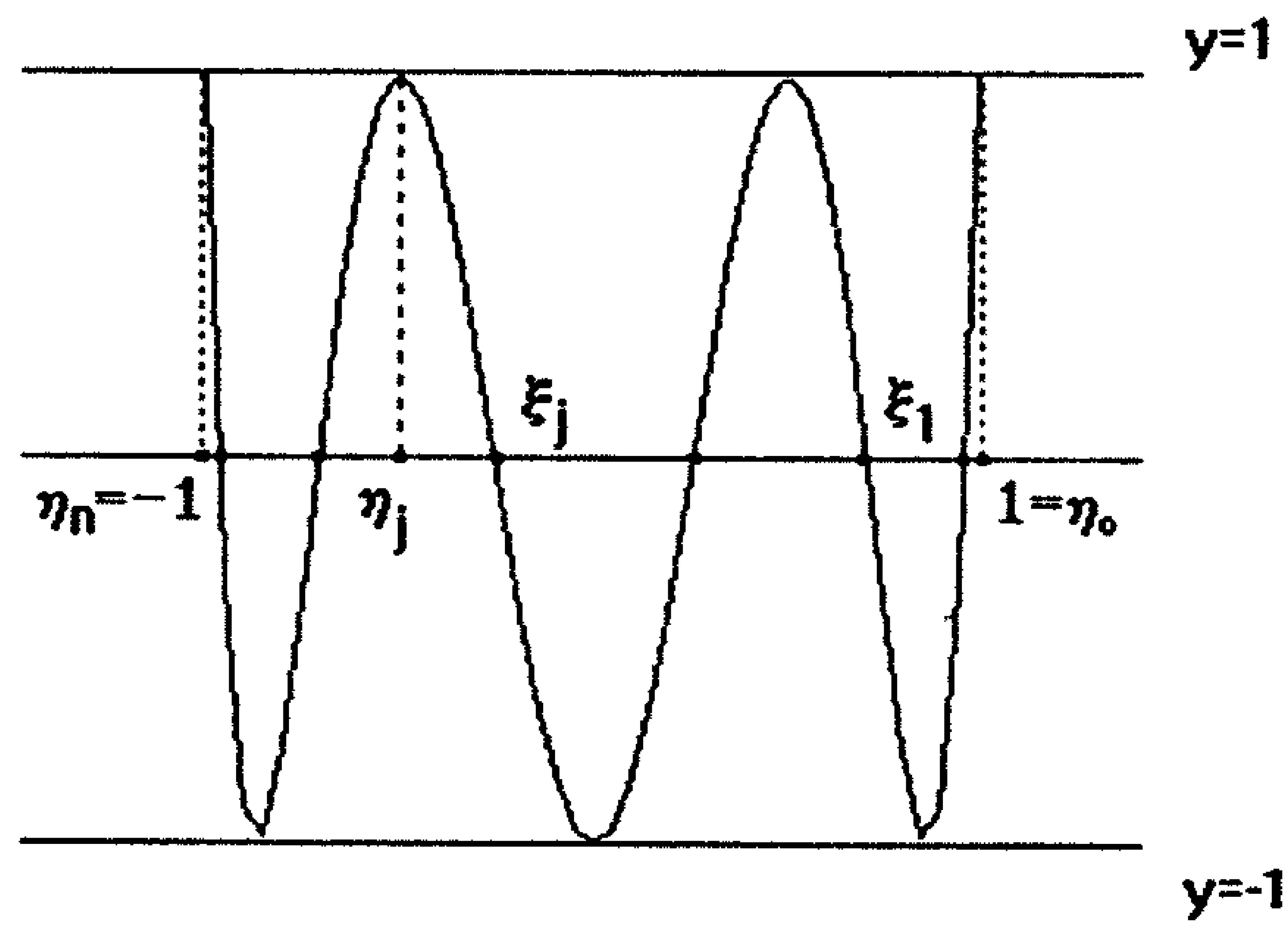

Рис. 2

От (4) следва, че $T_n(x)$ има точно n различни реални нули в $[-1,1]$. Те могат да се намерят веднага от израза (1). Очевидно $T_n(x) = 0$ при $n \arccos x = (2k-1)\frac{\pi}{2}$, $k = 1, 2, \ldots$ Оттук определяме нулите $\{\xi_k\}_{k=1}^n$ на $T_n(x)$:

$$\xi_k = \cos\frac{(2k-1)\pi}{2n}, \quad k = 1, \ldots, n.$$

Сега ще докажем следното екстремално свойство на полиномите на Чебишов.

Теорема 1. *Нека $P(x)$ е произволен алгебричен полином от степен n с коефициент 2^{n-1} пред x^n. Тогава*

(5) $$\max_{x\in[-1,1]} |T_n(x)| \le \max_{x\in[-1,1]} |P(x)|.$$

Равенство имаме само при $P(x) \equiv T_n(x)$.

Доказателство. От (3) знаем, че $\max\limits_{x\in[-1,1]} |T_n(x)| = 1$. Да допуснем, че има полином $P(x) = 2^{n-1}x^n + \cdots$, за който $|P(x)| \le 1$ при всяко $x \in [-1,1]$.

Тогава полиномът

$$Q(x) := T_n(x) - P(x)$$

ще бъде най-много от степен $n-1$ (защото коефициентите пред x^n в $T_n(x)$ и $P(x)$ са еднакви и се съкращават при изваждането). Освен това

$$Q(\eta_k) = (-1)^k - P(\eta_k), \quad k = 0, \dots, n.$$

Тъй като $|P(\eta_k)| \le 1$, то знакът на $Q(\eta_k)$ е равен на знака на $(-1)^k$ или $Q(\eta_k) = 0$. И така, ако $Q(\eta_k) \ne 0$ и $Q(\eta_{k-1}) \ne 0$, то $Q(\eta_k).Q(\eta_{k-1}) < 0$ и следователно Q има поне една нула в (η_k, η_{k-1}). Ако $Q(\eta_k) = 0$ за някое k, то очевидно $P(\eta_k) = (-1)^k = T_n(\eta_k)$ и тъй като $|T_n(x)| \le 1$ и $|P(x)| \le 1$, то графиката на P и графиката на T_n се допират до правата $y = (-1)^k$ в точката η_k. Тогава $P'(\eta_k) = T_n'(\eta_k) = 0$ и следователно η_k е нула с кратност 2 за Q - едната можем да свържем с интервала (η_k, η_{k-1}), а другата с интервала (η_{k+1}, η_k). По този начин на всеки интервал (η_i, η_{i-1}), $i = 1, \dots, n$ ще съответства поне по една нула на Q. От тези разсъждения се вижда, че $Q(x)$ има поне n нули в $[-1, 1]$ (броейки кратностите). Но $Q \in \pi_{n-1}$. Следователно $Q(x) \equiv 0$, т.е. $P(x) \equiv T_n(x)$.

Теоремата е доказана.

Следствие 2. *За всеки полином P от n-та степен с коефициент 1 пред x^n е в сила неравенството*

$$\frac{1}{2^{n-1}} = \max_{x \in [-1,1]} \frac{1}{2^{n-1}} |T_n(x)| \le \max_{x \in [-1,1]} |P(x)| \, .$$

Това твърдение следва от (5), като разделим двете страни на 2^{n-1}.

Следствие 3. *За всяка система от точки $\{x_k\}_0^n$ имаме*

$$\begin{aligned} \frac{1}{2^n} &= \max_{x \in [-1,1]} |(x - x_0^*) \dots (x - x_n^*)| \\ &\le \max_{x \in [-1,1]} |(x - x_0) \dots (x - x_n)|, \end{aligned}$$

където $\{x_k^*\}_0^n$ са нулите на полинома на Чебишов $T_{n+1}(x)$, т.е.

$$x_k^* = \cos \frac{(2k+1)\pi}{2(n+1)} \, , \quad k = 0, \dots, n.$$

И така, нулите на полинома на Чебишов $T_{n+1}(x)$ са най-добрите възли за интерполиране в интервала $[-1, 1]$, защото при тях се получава най-добра оценка на грешката $R_n(f)$.

Да запишем тази грешка като приложим Следствие 3 и оценката, дадена в началото на тази лекция. Получаваме

$$|R_n(f)| \le \frac{M_{n+1}}{(n+1)!}\,\frac{1}{2^n}\,.$$

Тази оценка се отнася за грешката при интерполиране в $[-1,1]$. Да видим сега как изглежда тя при произволен интервал $[a,b]$.

Линейната смяна $x = \frac{2}{b-a}t - \frac{a+b}{b-a}$ и нейната обратна $t = \frac{b-a}{2}x + \frac{a+b}{2}$ трансформират интервалите $[a,b]$ и $[-1,1]$ един в друг. Нека $\{t_k\}_{k=0}^n$ са произволни точки от интервала $[a,b]$. Да означим

$$x_k = \frac{2}{b-a}t_k - \frac{a+b}{b-a}, \quad k = 0,\dots,n.$$

Очевидно $x_k \in [-1,1]$ за $k = 0,\dots,n$. Тъй като

$$\begin{aligned}
|(t-t_0)\dots(t-t_n)| &= \left|\prod_{k=0}^{n}\left[\left(\frac{b-a}{2}x + \frac{a+b}{2}\right) - \left(\frac{b-a}{2}x_k + \frac{a+b}{2}\right)\right]\right| \\
&= \left(\frac{b-a}{2}\right)^{n+1} |(x-x_0)\dots(x-x_n)|,
\end{aligned}$$

то въз основа на Следствие 3 получаваме

$$\begin{aligned}
\max_{t\in[a,b]} |(t-t_0)\dots(t-t_n)| &\ge \left(\frac{b-a}{2}\right)^{n+1} \cdot \max_{x\in[-1,1]} |(x-x_0^*)\dots(x-x_n^*)| \\
&= \left(\frac{b-a}{2}\right)^{n+1} \frac{1}{2^n}\,.
\end{aligned}$$

Следователно, ако за интерполационни възли в $[a,b]$ изберем точките $t_k^* = \frac{b-a}{2}x_k^* + \frac{a+b}{2}$, където $\{x_k^*\}_0^n$ са нулите на полинома на Чебишов от първи род $T_{n+1}(x)$, то за грешката при интерполиране получаваме оценката

$$|R_n(f)| \le \frac{(b-a)^{n+1}}{2^{2n+1}}\,\frac{M_{n+1}}{(n+1)!}.$$

3. РАЗДЕЛЕНИ РАЗЛИКИ. ИНТЕРПОЛАЦИОННА ФОРМУЛА НА НЮТОН

Вече споменахме, че задачата за построяване на алгебричен полином p от π_n, който интерполира дадена функция f в $n+1$ точки $x_0, \ldots, x_n$, е била решена най-напред от Нютон. Сега ще представим неговото решение. За целта ще въведем едно ново понятие - *разделена разлика.*

Определение. Нека $x_0, \ldots, x_n$ са дадени различни точки (т.е. $x_i \neq x_j$ при $i \neq j$). *Разделената разлика на функцията* f в точките $x_0, \ldots, x_n$ се бележи с $f[x_0, \ldots, x_n]$ и се определя индуктивно със следната рекурентна връзка

$$(1) \qquad f[x_0, \ldots, x_n] = \frac{f[x_1, \ldots, x_n] - f[x_0, \ldots, x_{n-1}]}{x_n - x_0}, \quad n = 1, 2, \ldots,$$

като приемаме, че $f[x_i] := f(x_i)$ за всяка точка x_i.

Съществува тясна връзка между интерполационния полином на Лагранж с възли $x_0, \ldots, x_n$ и разделената разлика $f[x_0, \ldots, x_n]$. Тя се разкрива в следната теорема.

Теорема 1. *Разделената разлика $f[x_0, \ldots, x_n]$ съвпада с коефициента пред x^n в интерполационния полином на Лагранж $L_n(f; x)$ за функцията f с възли в същите точки $x_0, \ldots, x_n$.*

Доказателство. Доказателството се извършва по индукция относно броя на точките. При две точки x_0, x_1 имаме

$$\begin{aligned}
L_1(f; x) &= f(x_0)\frac{x - x_1}{x_0 - x_1} + f(x_1)\frac{x - x_0}{x_1 - x_0} \\
&= \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) + f(x_0) \\
&= f[x_0, x_1](x - x_0) + f(x_0)
\end{aligned}$$

и следователно коефициентът пред x в $L_1(f; x)$ е точно равен на разделената разлика $f[x_0, x_1]$. Да допуснем сега, че теоремата е вярна за произволни n точки. Ще я докажем за $n+1$ точки. И така, нека $x_0, \ldots, x_n$ са произволни $n+1$ различни точки. Да въведем полиномите $p(x)$ и $q(x)$ от π_{n-1} по следния начин:

$$\begin{array}{llll}
p(x) & \text{интерполира} \ f & \text{в} & x_1, \ldots, x_n. \\
q(x) & \text{интерполира} \ f & \text{в} & x_0, \ldots, x_{n-1}.
\end{array}$$

Да разгледаме полинома

$$r(x) := \frac{(x - x_0)p(x) - (x - x_n)q(x)}{x_n - x_0}.$$

Тъй като p и q са от π_{n-1}, то r е алгебричен полином от степен $\leq n$. Освен това, при $i \in \{1, \ldots, n-1\}$,

$$r(x_i) := \frac{(x_i - x_0)f(x_i) - (x_i - x_n)f(x_i)}{x_n - x_0} = f(x_i).$$

При $i = 0$ и $i = n$ имаме

$$\begin{aligned} r(x_0) &= -\frac{(x_0 - x_n)}{x_n - x_0} q(x_0) = f(x_0), \\ r(x_n) &= \frac{(x_n - x_0)}{x_n - x_0} p(x_n) = f(x_n). \end{aligned}$$

И така, $r \in \pi_n$ и $r(x)$ интерполира $f(x)$ в точките $x_0, \ldots, x_n$. От единствеността на интерполационния полином на Лагранж следва, че

$$r(x) \equiv L_n(f; x).$$

Следователно коефициентът пред x^n в $L_n(f; x)$ е равен на коефициента пред x^n в $r(x)$. Да означим с α и β коефициентите пред x^{n-1} в $p(x)$ и $q(x)$, съответно. Тогава от формулата за $r(x)$ се вижда, че коефициентът D пред x^n в $r(x)$ е равен на

$$\frac{\alpha - \beta}{x_n - x_0}.$$

Но съгласно индукционното предположение

$$\alpha = f[x_1, \ldots, x_n], \quad \beta = f[x_0, \ldots, x_{n-1}].$$

Следователно

$$D = \frac{f[x_1, \ldots, x_n] - f[x_0, \ldots, x_{n-1}]}{x_n - x_0} = f[x_0, \ldots, x_n].$$

Последното равенство следва от рекурентната връзка (1). Индукцията е завършена. Теоремата е доказана.

От Теорема 1 следват редица интересни свойства на разделената разлика. Ще отбележим някои от тях.

От записа

$$f[x_0, x_1] = f(x_0)\,\frac{1}{x_0 - x_1} + f(x_1)\,\frac{1}{x_1 - x_0}\ ,$$

се вижда, че разделената разлика $f[x_0, x_1]$ се представя като линейна комбинация на стойностите на функцията f в x_0 и x_1. Тогава от рекурентната връзка (1) следва, че всяка разделена разлика $f[x_0, \ldots, x_n]$ се представя като линейна комбинация на стойностите на функцията f в $x_0, \ldots, x_n$. Сега ще намерим коефициентите в това представяне. За целта ще използваме Теорема 1.

По формулата на Лагранж,

$$\begin{aligned} L_n(f;x) &= \sum_{k=0}^{n} f(x_k) \prod_{i=0, i\neq k}^{n} \frac{x-x_i}{x_k-x_i} \\ &= \sum_{k=0}^{n} \frac{f(x_k)}{\prod\limits_{i=0, i\neq k}^{n} (x_k-x_i)} \frac{\omega(x)}{x-x_k}, \end{aligned}$$

където $\omega(x) := (x-x_0)\dots(x-x_n)$. От последното равенство виждаме, че *коефициентът пред* x^n *в* $L_n(f;x)$ *е равен на*

$$\sum_{k=0}^{n} \frac{f(x_k)}{\prod\limits_{i=0, i\neq k}^{n} (x_k-x_i)}.$$

Следователно, съгласно Теорема 1,

$$f[x_0,\dots,x_n] = \sum_{k=0}^{n} \frac{f(x_k)}{\prod\limits_{i=0, i\neq k}^{n} (x_k-x_i)}. \tag{2}$$

Това е търсеното явно представяне на разделената разлика чрез стойностите на f в точките $x_0,\dots,x_n$. Като използваме установеното вече равенство

$$\omega'(x_k) = \prod_{i=0, i\neq k}^{n} (x_k-x_i)$$

можем да запишем (2) и по-кратко:

$$f[x_0,\dots,x_n] = \sum_{k=0}^{n} \frac{f(x_k)}{\omega'(x_k)}.$$

От представянето (2) се вижда веднага, че разделената разлика е един линеен функционал, т.е. за всеки две функции f, g и число c е в сила формулата

$$(f+cg)[x_0,\dots,x_n] = f[x_0,\dots,x_n] + cg[x_0,\dots,x_n].$$

Друго следствие от (2) е, че разделената разлика не зависи от реда, в който се записват точките. Имаме

$$f[x_0,\dots,x_n] = f[x_{i_0},\dots,x_{i_n}]$$

за всяко разместване $(i_0, \dots i_n)$ на индексите $(0, \dots, n)$. Наистина, при разместване на индексите се променят само местата на събираемите в сумата (2).

Сега ще докажем, че ако $f(x) = a_0 x^n + \dots + a_{n-1}x + a_n$, то

$$f[x_0, \dots, x_n] = a_0.$$

С други думи, разделената разлика в $n+1$ точки на полином от степен n е равна на коефициента му пред x^n. Това твърдение следва веднага от факта, че ако $f \in \pi_n$, то f съвпада с интерполационния си полином на Лагранж в $n+1$ точки. Тогава

$$\begin{aligned} f[x_0, \dots, x_n] &= \text{коефициента пред } x^n \text{ в } L_n(f; x) \\ &= \text{коефициента пред } x^n \text{ в } f(x) \\ &= a_0. \end{aligned}$$

Един важен частен случай от това твърдение е следното свойство:

$$\text{Ако} \quad f \in \pi_{n-1}, \quad \text{то} \quad f[x_0, \dots, x_n] = 0.$$

Наистина, ако $f \in \pi_{n-1}$, то коефициентът пред x^n в $f(x)$ е равен на нула. И така, *разделената разлика в $n+1$ точки анулира всички полиноми от степен по-малка или равна на $n-1$*.

Сега вече сме готови да изведем формулата на Нютон за интерполационния полином. За целта да разгледаме разликата

$$L_{k+1}(f; x) - L_k(f; x),$$

където $L_{k+1}(f; x)$ интерполира f в точките $x_0, \dots, x_{k+1}$, а $L_k(f; x)$ интерполира f в точките $x_0, \dots, x_k$. Ясно е, че $L_{k+1}(f; x) - L_k(f; x)$ е алгебричен полином от степен $k+1$. Освен това

$$L_{k+1}(f; x_i) - L_k(f; x_i) = f(x_i) - f(x_i) = 0 \quad \text{за} \quad i = 0, \dots, k.$$

Следователно $x_0, \dots, x_k$ са всичките нули на полинома $L_{k+1}(f; x) - L_k(f; x)$. Тогава той може да се запише във вида

$$L_{k+1}(f; x) - L_k(f; x) = A(x - x_0) \dots (x - x_k), \tag{3}$$

където A е някаква константа. За да намерим A, нека да сравним коефициентите пред x^{k+1} в тъждеството (3). Вдясно този коефициент е A, а вляво – това е коефициентът пред x^{k+1} в $L_{k+1}(f; x)$. Но съгласно Теорема 1, коефициентът пред x^{k+1} в $L_{k+1}(f; x)$ е равен на разделената разлика $f[x_0, \dots, x_{k+1}]$. И така доказахме, че

$$A = f[x_0, \dots, x_{k+1}]$$

и следователно, от (3),

$$L_{k+1}(f;x) = L_k(f;x) + f[x_0,\ldots,x_{k+1}](x-x_0)\ldots(x-x_k). \tag{4}$$

Нека приложим сега тази връзка за $k = n-1,\ldots,2,1,0$. Получаваме следния явен израз за интерполационния полином на Лагранж

$$\begin{aligned} L_n(f;x) &= f(x_0) + f[x_0,x_1](x-x_0) + f[x_0,x_1,x_2](x-x_0)(x-x_1) + \cdots \\ &\quad + f[x_0,\ldots,x_n](x-x_0)\ldots(x-x_{n-1}). \end{aligned}$$

Това е *интерполационната формула на Нютон*. Понякога ще я записваме съкратено така

$$L_n(f;x) = \sum_{k=0}^{n} f[x_0,\ldots,x_k](x-x_0)\ldots(x-x_{k-1}), \tag{5}$$

като приемаме, че $(x-x_0)\ldots(x-x_{k-1}) = 1$ при $k=0$.

Сега ще изведем един израз за остатъка при интерполиране на f като използваме разделени разлики. Нека x е произволна фиксирана точка, различна от $x_0,\ldots,x_n$. Да означим с $L_{n+1}(f;t)$ полинома, който интерполира f в точките $x_0,\ldots,x_n$ и x. Нека $L_n(f;t)$ интерполира f в точките $x_0,\ldots,x_n$. От връзката (4) следва

$$L_{n+1}(f;t) = L_n(f;t) + f[x_0,\ldots,x_n,x](t-x_0)\ldots(t-x_n).$$

Това равенство е вярно за всяко t. Специално при $t=x$ имаме

$$L_{n+1}(f;x) = L_n(f;x) + f[x_0,\ldots,x_n,x](x-x_0)\ldots(x-x_n).$$

Но тъй като x е интерполационен възел за $L_{n+1}(f;t)$, то $L_{n+1}(f;x) = f(x)$. Следователно

$$f(x) = L_n(f;x) + f[x_0,\ldots,x_n,x](x-x_0)\ldots(x-x_n). \tag{6}$$

Равенството беше изведено при предположение, че $x \notin \{x_0,\ldots,x_n\}$. При $x = x_k,\ \ k=0,\ldots,n$, по определение имаме $f(x) = L_n(f;x)$.

Да отбележим, че представянето (6) е в сила за всяка функция f определена в точките $x_0,\ldots,x_n,x$.

Да сравним сега формулата (6) с известната ни вече формула

$$f(x) = L_n(f;x) + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)\ldots(x-x_n),$$

изведена при предположение, че f има непрекъсната $(n+1)$-ва производна. В първия случай остатъкът при интерполиране с $L_n(f;x)$ е записан като

$$f[x_0,\ldots,x_n,x]\omega(x), \quad \omega(x) := (x-x_0)\ldots(x-x_n),$$

а във втория, като

$$\frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x),$$

където ξ е някаква точка. Следователно разделената разлика на f в $n+2$ точки $x_0,\ldots,x_n,x$ е равна на $(n+1)$-вата производна в някаква междинна точка. Тъй като това свойство на разделената разлика е много важно, да го запишем точно:

Нека $f(x)$ има непрекъснати производни до k-тата включително в интервала $[a,b]$ и $x_0,\ldots,x_n$ са произволни различни точки в $[a,b]$. Тогава

$$f[x_0,\ldots,x_k]=\frac{f^{(k)}(\xi)}{k!}, \tag{7}$$

където ξ е някаква точка от интервала $(\min\{x_0,\ldots,x_k\},\max\{x_0,\ldots,x_k\})$.

От тази връзка, между другото, директно следва, че ако $f\in\pi_{k-1}$, то $f[x_0,\ldots,x_k]=0$ (защото $f^{(k)}(t)\equiv 0$).

От формулата на Нютон се вижда, че за да построим интерполационния полином $L_n(f;x)$, достатъчно е да намерим разделените разлики $f[x_0,\ldots,x_k]$, $k=0,\ldots,n$. Съществува много проста и удобна за компютърна реализация схема за изчисляване на разделените разлики. Тя се основава единствено на рекурентната връзка.

Схема за пресмятане на разделени разлики

x_i	f_i	$f[\cdot,\cdot]$	$f[\cdot,\cdot,\cdot]$	$f[\cdot,\cdot,\cdot,\cdot]$
x_0	$f(x_0)$			
		$f[x_0,x_1]$		
x_1	$f(x_1)$		$f[x_0,x_1,x_2]$	
		$f[x_1,x_2]$		$f[x_0,x_1,x_2,x_3]$
x_2	$f(x_2)$		$f[x_1,x_2,x_3]$	
		$f[x_2,x_3]$		$f[x_1,x_2,x_3,x_4]$
x_3	$f(x_3)$		$f[x_2,x_3,x_4]$	
		$f[x_3,x_4]$		
x_4	$f(x_4)$			

В първия стълб се записват възлите $\{x_i\}$, а във втория – стойностите $\{f(x_i)\}$. Таблицата се попълва стълб след стълб, като се използват намерените вече разлики в предишния стълб. Коефициентите $f[x_0, \ldots, x_k]$, $k = 0, \ldots, n$, във формулата на Нютон се намират по горния диагонал на таблицата.

Пример. Да построим полином $p(x)$ от степен 2, който удовлетворява интерполационните условия

$$p(0) = 1, \quad p(1) = 0, \quad p(2) = 3.$$

Решение. В този случай $x_0 = 0$, $x_1 = 1$, $x_2 = 2$. По интерполационната формула на Нютон

$$\begin{aligned} p(x) &= L_2(p; x) = p(x_0) + p[x_0, x_1](x - x_0) + p[x_0, x_1, x_2](x - x_0)(x - x_1) \\ &= p(0) + p[0, 1]x + p[0, 1, 2]x(x - 1). \end{aligned}$$

Коефициентите $p(x_0)$, $p[x_0, x_1]$, $p[x_0, x_1, x_2]$ се намират по горния диагонал на таблицата

$$\begin{array}{cccc} x_i & p(x_i) & & \\ 0 & 1 & & \\ & & -1 & \\ 1 & 0 & & 2 \\ & & 3 & \\ 2 & 3 & & \end{array}$$

Имаме $p(x_0) = 1$, $p[x_0, x_1] = -1$, $p[x_0, x_1, x_2] = 2$. Следователно

$$p(x) = 1 + (-1)x + 2x(x - 1) = 2x^2 - 3x + 1.$$

Можем да направим проверка за да се убедим, че намереният полином $p(x)$ удовлетворява исканите интерполационни условия.

4. КРАЙНИ РАЗЛИКИ. ИНТЕРПОЛАЦИОННИ ФОРМУЛИ С КРАЙНИ РАЗЛИКИ

Най-често използвани на практика са равноотдалечените възли при интерполиране на функции. В този случай може да се предложи една значително по-проста схема за построяване на интерполационния полином. Това става с използването на така наречените *крайни разлики.* Ние най-напред ще въведем това ново понятие и ще представим някои негови елементарни свойства.

Нека е дадена една редица от числа

$$f_0, f_1, f_2, \dots, f_m, \dots .$$

Ще интерпретираме тези числа като стойности на функция f в някакви точки

$$x_0, x_1, x_2, \dots, x_m, \dots .$$

Определение. *Крайна разлика на f в x_i от ред k* се бележи с $\Delta^k f_i$ и се определя индуктивно с рекурентната връзка

$$\Delta^k f_i := \Delta^{k-1} f_{i+1} - \Delta^{k-1} f_i, \quad k = 1, 2, \dots,$$

където $\Delta^1 f_i = \Delta f_i := f_{i+1} - f_i$ за всяко i.

В случай, че точките $\{x_i\}$ са равноотдалечени, съществува проста връзка между разделената и крайна разлика. Тя е представена в следната лема.

Лема 1. *Нека $x_j = x_0 + jh$, $j = 0, \dots, k$, и функцията $f(x)$ е определена в тези точки. Тогава*

$$f[x_0, \dots, x_k] = \frac{\Delta^k f_0}{k!\, h^k}. \tag{1}$$

Доказателство. Ще приложим индукция по броя на точките. За две точки имаме

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{f_{i+1} - f_i}{h} = \frac{\Delta f_i}{1!\, h}$$

и следователно твърдението е вярно. Да допуснем, че връзката (1) е в сила за произволни k равноотдалечени точки. Нека $x_j = x_0 + jh$, $j = 0 \dots, k$, са произволни $k+1$ точки. Като приложим рекурентната връзка за разделени разлики и индукционното предположение получаваме

$$\begin{aligned} f[x_0, \dots, x_k] &= \frac{f[x_1, \dots, x_k] - f[x_0, \dots, x_{k-1}]}{x_k - x_0} \\ &= \frac{1}{x_k - x_0} \left(\frac{\Delta^{k-1} f_1}{(k-1)!h^{k-1}} - \frac{\Delta^{k-1} f_0}{(k-1)!\, h^{k-1}} \right) = \frac{\Delta^k f_0}{k!\, h^k}. \end{aligned}$$

Лемата е доказана.

Чрез връзката (1) много от свойствата на разделената разлика се пренасят върху крайните разлики. Да отбележим някои от тях.

1. Крайната разлика е линеен функционал, т.е.

$$\Delta^n (f + \alpha g)_i = \Delta^n f_i + \alpha \Delta^n g_i$$

за всеки две функции f, g и число α.

2. Нека $f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$. Тогава

$$\Delta^n f_0 = n!\, h^n a_0$$

при всеки избор на $h > 0$ и точките $x_j = x_0 + jh$, $j = 0, \ldots, n$.

3. Крайната разлика от n-ти ред анулира всички полиноми от степен $n - 1$.

Съгласно определението, $\Delta f_i = f_{i+1} - f_i$. Оттук и рекурентната връзка следва, че всяка крайна разлика (от произволен ред) може да се представи като линейна комбинация на стойностите $\{f_i\}$. Например,

$$\begin{aligned} \Delta^2 f_0 &= \Delta f_1 - \Delta f_0 = (f_2 - f_1) - (f_1 - f_0) = f_2 - 2f_1 + f_0, \\ \Delta^3 f_0 &= \Delta^2 f_1 - \Delta^2 f_0 = f_3 - 3f_2 + 3f_1 - f_0. \end{aligned}$$

От тези примери се вижда, че коефициентите в разглежданото представяне са биномните коефициенти с алтернативно сменящи се знаци. Оказва се, че това наистина е така и то може да бъде строго доказано, например по индукция, като се използва рекурентната връзка за крайни разлики и свойствата на биномните коефициенти. Ние ще дадем тук едно друго доказателство, което се основава на връзката между разделена и крайна разлика.

Теорема 2. *За всяко естествено число n е в сила формулата*

$$\Delta^n f_0 = \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} f_i.$$

Доказателство. Нека $x_j = x_0 + jh$, $j = 0, \ldots, n$ и $f_i = f(x_i)$, $i = 0, \ldots, n$. Тогава по Лема 1,

$$\Delta^n f_0 = n!\, h^n f[x_0, \ldots, x_n].$$

По-нататък, като използваме представянето (3.2) на разделената разлика, получаваме

$$\Delta^n f_0 = n!\, h^n \sum_{i=0}^{n} \frac{f(x_i)}{\prod_{j=0, j\neq i} (ih - jh)}$$

$$= n! \sum_{i=0}^{n} (-1)^{n-i} \frac{f_i}{i!(n-i)!},$$

което е исканото равенство.

Тези сведения за крайните разлики са достатъчни за да се справим с нашата първоначална задача – представянето на интерполационния полином. И така, нека възлите $\{x_i\}_{i=0}^{n}$ са равноотдалечени и функцията f е определена в тях. Търсим полинома $L_n(f;x)$ от π_n, който интерполира f в $x_0, \dots, x_n$. Съгласно интерполационната формула на Нютон

$$L_n(f;x) = \sum_{k=0}^{n} f[x_0, \dots, x_k](x - x_0) \cdots (x - x_{k-1}).$$

Нека $x_i = x_0 + i\,h$, $i = 0, \dots, n$. Да направим смяна на променливата x с t по формулата $x = x_0 + th$. Тогава

$$(x - x_0) \cdots (x - x_{k-1}) = \prod_{i=0}^{k-1} (x_0 + th - x_0 - ih) = h^k t(t-1) \cdots (t-k+1).$$

Сега, като използваме и връзката между разделена и крайна разлика, получаваме

$$L_n(f;x) = L_n(f; x_0 + th) = \sum_{k=0}^{n} \frac{\Delta^k f_0}{k!} t(t-1) \cdots (t-k+1).$$

В литературата се среща означението $\binom{t}{k}$ при произволни реални стойности на параметъра t. С него се означава *биномната функция*, която се определя с равенството:

$$\binom{t}{k} := \begin{cases} \frac{t(t-1)\cdots(t-k+1)}{k!} & \text{при } k > 0 \\ 1 & \text{при } k = 0. \end{cases}$$

Следователно полученият израз за интерполационния полином може да се запише и така:

$$L_n(f; x_0 + th) = \sum_{k=0}^{n} \Delta^k f_0 \binom{t}{k}.$$

Това е *формулата на Нютон за интерполиране напред*. Тя се нарича така, защото възлите се привличат в нарастващ ред при изчисляване на коефициентите пред полиномите $\binom{t}{k}$. Да забележим, че стойността $f(x_0)$ участва във всички коефициенти, стойността в следващия възел x_1 участва във всички от втория до последния и т.н., стойността $f(x_n)$ участва само в последния коефициент. Следователно, ако искаме да изчислим приближено стойността на f в точка x, която е близко до x_0, то добре е да използваме формулата на Нютон за интерполиране напред, защото в тази формула участват съществено стойностите на f в точки близки до x_0 и като такива, те носят най-пълна информация за стойността на f в x. Следвайки тази логика би трябвало при приближаване на $f(x)$ за точки x, близки до последния възел x_n да използваме интерполационна формула, в която възлите се привличат в обратен ред: $x_n, x_{n-1}, \dots, x_0$. Да изведем и тази формула. По формулата на Нютон, приложена за възлите $x_n, x_{n-1}, \dots, x_0$ (в този ред), имаме

$$L_n(f;x) = \sum_{k=0}^{n} f[x_n, x_{n-1}, \dots, x_{n-k}](x - x_n)\cdots(x - x_{n-k+1}).$$

Като приложим смяната $x = x_n + th$ получаваме

$$L_n(f;x) = L_n(f;x_n + th) = \sum_{k=0}^{n} \frac{\Delta^k f_{n-k}}{k!\,h^k} \prod_{i=0}^{k-1}(x_n + th - x_n + ih)$$

и следователно

$$L_n(f;x_n + th) = \sum_{k=0}^{n} \Delta^k f_{n-k} \binom{t+k-1}{k}.$$

Това е *формулата на Нютон за интерполиране назад.*

По аналогичен начин могат да бъдат изведени интерполационни фомули, при които възлите се привличат в произволен друг ред. Например, ако точката x е близко до x_i, то добре е възлите да се подредят по следния начин: $x_i, x_{i+1}, x_{i-1}, x_{i+2}, x_{i-2}, \dots$.

Пресмятането на коефициентите на интерполационния полином с равноотдалечени възли се свежда към пресмятане на крайни разлики. Изчисленията могат да се организират по следната проста схема:

Схема за пресмятане на крайни разлики

$$
\begin{array}{cccccc}
x_i & f_i & & & \\
\vdots & \vdots & & & \\
x_{-3} & f_{-3} & & & \\
 & & \Delta f_{-3} & & \\
x_{-2} & f_{-2} & & \Delta^2 f_{-3} & \\
 & & \Delta f_{-2} & & \Delta^3 f_{-3} \\
x_{-1} & f_{-1} & & \Delta^2 f_{-2} & \\
 & & \Delta f_{-1} & & \Delta^3 f_{-2} \\
x_0 & f_0 & & \Delta^2 f_{-1} & \\
 & & \Delta f_0 & & \Delta^3 f_{-1} \\
x_1 & f_1 & & \Delta^2 f_0 & \\
 & & \Delta f_1 & & \Delta^3 f_0 \\
x_2 & f_2 & & \Delta^2 f_1 & \\
 & & \Delta f_2 & & \\
x_3 & f_3 & & & \\
\vdots & \vdots & & &
\end{array}
$$

В първите два стълба на тази таблица се попълват данните – интерполационните възли и стойности. След това таблицата се попълва стълб след стълб, като се използва рекурентната връзка за крайни разлики. Числата, получени по горния диагонал, започващ от f_0, са коефициентите на интерполационния полином при интерполиране напред, а тези по долния диагонал – на полинома при интерполиране назад. Да отбележим, че тук, за разлика от схемата за разделени разлики, се използва само операцията изваждане (при разделените разлики се изпълняваше и операцията деление). Това е едно съществено предимство на смятането с крайни разлики.

5. ИНТЕРПОЛАЦИОННА ЗАДАЧА НА ЕРМИТ

Досега се занимавахме с интерполационната задача на Лагранж, която се състоеше в построяването на алгебричен полином от степен $\leq n$, който в $n+1$ дадени различни точки $x_0, \dots, x_n$ приема дадени стойности $y_0, \dots, y_n$, съответно. Формулата на Лагранж, даваща решение на тази задача, играе първостепенна роля в числения анализ. Сега ще разгледаме една по-обща задача, при която се търси полином, който интерполира не само функцията, но и нейни производни. Да представим най-напред точната фомулировка.

Нека $x_0, \dots, x_n$ са дадени $n+1$ различни точки от реалната права. Нека $\nu_0, \dots, \nu_n$ са цели положителни числа и

$$\{y_{k\lambda},\ k = 0, \dots, n,\ \lambda = 0, \dots, \nu_k - 1\}$$

е таблица от произволни реални стойности. Означаваме $N := \nu_0 + \dots + \nu_n - 1$. Задачата е да се построи алгебричен полином P от степен N, който удовлетворява условията

$$P^{(\lambda)}(x_k) = y_{k\lambda},\ k = 0, \dots, n,\ \lambda = 0, \dots, \nu_k - 1. \tag{1}$$

Тя е известна като *интерполационна задача на Ермит.*

Теорема 1. *При всеки избор на интерполационните възли $\{x_k\}_0^n$ ($x_i \neq x_j$ при $i \neq j$) и при всяка таблица от стойности $\{y_{k\lambda}\}$ интерполационната задача на Ермит (1) има единствено решение.*

Доказателство. Условията (1) представляват една система от $N+1$ линейни уравнения с неизвестни – коефициентите $a_0, \dots, a_N$ на полинома $P(x)$. Тази система ще има единствено решение, ако нейната детерминанта D е различна от нула. Да допуснем, че $D = 0$. Тогава хомогенната система

$$P^{(\lambda)}(x_k) = 0, \quad k = 0, \dots, n,\ \lambda = 0, \dots, \nu_k - 1,$$

ще има някакво ненулево решение $P(x) = a_0 x^N + \dots + a_{N-1}x + a_N$ (т.е. с поне един коефициент a_i различен от нула). Но горните условия означават, че P има $N+1$ нули, броейки кратностите. От друга страна, $P \in \pi_N$. Следователно $P(x) \equiv 0$ и оттук $a_0 = \dots = a_N = 0$. Стигнахме до противоречие. Теоремата е доказана.

Остава да разгледаме важния за нас въпрос за построяване на решението. Ще започнем с един частен случай, при който $\nu_0 = \dots = \nu_n = 2$. Ще намерим в явен вид полинома от степен $2n+1$, който интерполира дадена функция f и нейната първа производна в $n+1$ точки $x_0 < \dots < x_n$. За целта ще използваме означенията

$$\omega(x) := (x - x_0) \dots (x - x_n), \quad \omega_k(x) := \frac{\omega(x)}{x - x_k}\ .$$

Да отбележим още тук, че

$$\omega'(x_k) = \omega_k(x_k) = (x_k - x_0)\dots(x_k - x_{k-1})(x_k - x_{k+1})\dots(x_k - x_n).$$

Теорема 2. *Нека $x_0, \dots, x_n$ са произволни различни (две по две) точки от реалната права. Тогава, при всеки избор на числата $y_0, \dots, y_n$ и $y'_0 \dots, y'_n$, полиномът*

$$\begin{aligned} P(x) &= \sum_{k=0}^{n} y_k \left\{1 - \frac{\omega''(x_k)}{\omega'(x_k)}(x - x_k)\right\} \left[\frac{\omega(x)}{(x - x_k)\omega'(x_k)}\right]^2 \\ &+ \sum_{k=0}^{n} y'_k \left\{\frac{\omega(x)}{(x - x_k)\omega'(x_k)}\right\}^2 (x - x_k) \end{aligned}$$

е от степен най-много $2n+1$ и удовлетворява условията

$$P(x_k) = y_k, \quad P'(x_k) = y'_k, \quad k = 0, \dots, n. \tag{2}$$

Доказателство. Съгласно Теорема 1 съществува и то единствен полином P от π_{2n+1}, който удовлетворява интерполационните условия (2). Ние ще търсим този полином във вида

$$P(x) = \sum_{k=0}^{n} y_k \Phi_{k0}(x) + \sum_{k=0}^{n} y'_k \Phi_{k1}(x),$$

където при всяко $k \in \{0, \dots, n\}$ базисните полиноми $\Phi_{k0}, \Phi_{k1} \in \pi_{2n+1}$ се определят от условието

$$\begin{cases} \Phi_{k0}(x_i) = \delta_{ki}, & \Phi'_{k0}(x_i) = 0, \\ \Phi_{k1}(x_i) = 0, & \Phi'_{k1}(x_i) = \delta_{ki} \end{cases} \tag{3}$$

за $i = 0, \dots, n$. Тук сме използвали *символа на Кронекер* δ_{ki},

$$\delta_{ki} := \begin{cases} 0 & \text{при} \quad k \neq i \\ 1 & \text{при} \quad k = i. \end{cases}$$

Очевидно условията (3) влекат веднага (2). Това се установява с директна проверка. Сега да построим полиномите Φ_{k0} и Φ_{k1}. Ще започнем с Φ_{k0}. От (3) се вижда, че $\Phi_{k0}(x)$ има двукратна нула в x_i за всяко $i \neq k$. Следователно $\Phi_{k0}(x)$ е от вида

$$\Phi_{k0}(x) = \omega_k^2(x)[A + B(x - x_k)],$$

където константите A и B са избрани така, че да удовлетворяват условията

$$\Phi_{k0}(x_k) = 1, \quad \Phi'_{k0}(x_k) = 0.$$

От първото условие

$$\Phi_{k0}(x_k) = \omega_k^2(x_k)A = 1$$

определяме A,

$$A = \frac{1}{\omega_k^2(x_k)} = \frac{1}{[\omega'(x_k)]^2}.$$

Заместваме получената стойност за A във второто условие

$$\Phi'_{k0}(x_k) = 2\omega_k(x_k)\omega'_k(x_k)A + \omega_k^2(x_k)B = 0$$

и определяме B,

$$B = -2\frac{\omega'_k(x_k)}{\omega_k^3(x_k)}.$$

Остава да забележим, че $2\omega'_k(x_k) = \omega''(x_k)$. Наистина, като диференцираме два пъти тъждеството

$$\omega_k(x)(x - x_k) = \omega(x)$$

получаваме

$$\omega''_k(x)(x - x_k) + 2\omega'_k(x) = \omega''(x),$$

откъдето при $x = x_k$ следва исканото равенство. Следователно

$$\begin{aligned}\Phi_{k0}(x) &= \omega_k^2(x)\left[\frac{1}{\omega_k^2(x_k)} - \frac{\omega''(x_k)}{\omega_k^3(x_k)}(x - x_k)\right] \\ &= \left[1 - \frac{\omega''(x_k)}{\omega'(x_k)}(x - x_k)\right]\left[\frac{\omega(x)}{(x - x_k)\omega'(x_k)}\right]^2.\end{aligned}$$

От условията (3) можем да намерим лесно явния вид на $\Phi_{k1}(x)$. Тъй като x_i е двукратна нула на $\Phi_{k1}(x)$ при $i \neq k$ и x_k е проста нула, то

$$\Phi_{k1}(x) = C\omega_k^2(x)(x - x_k).$$

Константата C определяме от условието $\Phi'_{k1}(x_k) = 1$. Получаваме

$$C\omega_k^2(x_k) = 1.$$

Оттук $C = 1/\omega_k^2(x_k) = 1/[\omega'(x_k)]^2$ и следователно

$$\Phi_{k1}(x) = \left[\frac{\omega(x)}{(x - x_k)\omega'(x_k)}\right]^2(x - x_k).$$

Теоремата е доказана.

Обикновено числата $\{y_k\}$ и $\{y'_k\}$ са стойности на някаква функция $f(x)$ и нейната производна $f'(x)$ във фиксирани точки $\{x_k\}$. Тогава интерполационният полином P се нарича *интерполационнен полином на Ермит* за функцията f. Като използваме означението за фундаменталните полиноми на Лагранж $l_{nk}(x)$ и Теорема 2, този полином се записва по следния начин

$$
\begin{aligned}
P(f;x) &= \sum_{k=0}^{n} f(x_k)\left[1-\frac{\omega''(x_k)}{\omega'(x_k)}(x-x_k)\right] l_{nk}^2(x) \\
&+ \sum_{k=0}^{n} f'(x_k) l_{nk}^2(x)(x-x_k).
\end{aligned}
$$

Числата $\nu_0,\dots,\nu_n$ се наричат *кратности* на възлите $x_0,\dots,x_n$. Ние решихме задачата на Ермит в случая на двукратни възли. Решението може да бъде дадено в явен вид и при произволни кратности $\{\nu_k\}_{k=0}^n$. Ние ще запишем по-долу интерполационния полином на Ермит в общия случай.

При дадени възли $\{x_k\}_{k=0}^n$ и кратности $\{\nu_k\}_{k=0}^n$ да означим с $\Omega(x)$ израза $(x-x_0)^{\nu_0}\cdots(x-x_n)^{\nu_n}$. Нека $N+1 := \nu_0+\cdots+\nu_n$. Полиномът

$$
H_N(f;x) = \sum_{k=0}^{n}\sum_{\lambda=0}^{\nu_k-1} f^{(\lambda)}(x_k) H_{k\lambda}(x), \tag{4}
$$

където

$$
H_{k\lambda}(x) = \frac{1}{\lambda!}\frac{\Omega(x)}{(x-x_k)^{\nu_k-\lambda}} \sum_{\mu=0}^{\nu_k-\lambda-1} \frac{1}{\mu!}\left\{\frac{(x-x_k)^{\nu_k}}{\Omega(x)}\right\}^{(\mu)}\Bigg|_{x=x_k} (x-x_k)^{\mu}
$$

е от степен $\le N$ и удовлетворява интерполационните условия (1) при $y_{k\lambda} = f^{(\lambda)}(x_k)$, $k=0,\dots,n$, $\lambda=0,\dots,\nu_k-1$.

За доказателството на това твърдение е достатъчно да се провери, че полиномите $H_{k\lambda}(x)$ удовлетворяват условията

$$
H_{k\lambda}^{(j)}(x_i) = \begin{cases} 0 & \text{при} \quad i\ne k,\ j=0,\dots,\nu_i-1, \\ 0 & \text{при} \quad i=k,\ j\ne\lambda, \\ 1 & \text{при} \quad i=k, j=\lambda. \end{cases}
$$

При $i\ne k$ равенството следва от факта, че $H_{k\lambda}$ има множител $(x-x_i)^{\nu_i}$. Остава да се докаже само, че $H_{k\lambda}^{(j)}(x_k)=\delta_{j\lambda}$. За целта да означим с $T_m(g;x)$ полинома на Тейлър от степен m за функцията g в точката x_k. По-точно,

$$
T_m(g;x) := \sum_{s=0}^{m} \frac{g^{(s)}(x_k)}{s!}(x-x_k)^s.
$$

Тъй като $g^{(s)}(x_k) = T_m^{(s)}(g;x_k)$ за $s=0,\dots,m$, то ясно е, че

$$
\{f(x)g(x)\}^{(s)}\Big|_{x=x_k} = \{f(x)T_m(g;x)\}^{(s)}\Big|_{x=x_k}
$$

за $0 \le s \le m$. Сега да забележим, че

$$H_{k\lambda}(x) = \frac{1}{\lambda!}\frac{(x-x_k)^\lambda}{g(x)} T_{\nu_k-\lambda-1}(g;x)$$

при $g(x) = (x-x_k)^{\nu_k}/\Omega(x)$. Следователно

$$\begin{aligned} H_{k\lambda}^{(j)}(x_k) &= \frac{1}{\lambda!}\left\{\frac{(x-x_k)^\lambda}{g(x)} T_{\nu_k-\lambda-1}(g;x)\right\}^{(j)}\Bigg|_{x=x_k} \\ &= \frac{1}{\lambda!}\left\{\frac{(x-x_k)^\lambda}{g(x)} T_{\nu_k-1}(g;x)\right\}^{(j)}\Bigg|_{x=x_k} \\ &= \frac{1}{\lambda!}\left\{\frac{(x-x_k)^\lambda}{g(x)} g(x)\right\}^{(j)}\Bigg|_{x=x_k} \\ &= \frac{1}{\lambda!}\left\{(x-x_k)^\lambda\right\}^{(j)}\Bigg|_{x=x_k} = \delta_{j\lambda}, \end{aligned}$$

което трябваше да докажем.

Накрая ще дадем оценка на грешката, която правим при приближението $f(x) \approx H_N(f;x)$.

Теорема 3. *Нека $a \le x_0 < \cdots < x_n \le b$, $\{\nu_k\}_0^n$ са произволни цели положителни числа и функцията f има непрекъсната $(N+1)$-ва производна в $[a,b]$, $N := \nu_0 + \cdots + \nu_n - 1$. Тогава за всяко $x \in [a,b]$ съществува число $\xi \in (\min\{x, x_0, \dots x_n\}, \max\{x, x_0, \dots x_n\})$ такова, че*

$$f(x) - H_N(f;x) = \frac{f^{(N+1)}(\xi)}{(N+1)!}(x-x_0)^{\nu_0}\dots(x-x_n)^{\nu_n}. \tag{5}$$

Доказателство. Твърдението (5) се доказва по същия начин, както доказахме съответната теорема за грешката при интерполиране по Лагранж. Образуваме си помощната функция

$$F(z) = f(z) - H_N(f;z) - C\Omega(z)$$

и избираме C така, че $F(z)$ да се анулира при $z = x$. Тогава $F(z)$ ще има $N+2$ нули: $x_0, \dots, x_n$, с кратности съответно $\nu_0, \dots, \nu_n$ и точката x. По теоремата на Рол, $F^{(N+1)}(z)$ ще има поне една нула, която се намира между най-малката и най-голямата нула на $F(z)$. Означаваме я с ξ. Тогава като изразим C от равенството $F^{(N+1)}(\xi) = 0$ и от условието $F(x) = 0$ получаваме (5). Теоремата е доказана.

6. РАЗДЕЛЕНИ РАЗЛИКИ С КРАТНИ ВЪЗЛИ

Понятието разделена разлика на функция в дадени точки $x_0, \dots, x_N$ беше въведено при предположение, че точките са различни, т.е. че $x_i \neq x_j$ при $i \neq j$. Съществува естествено обобщение на това понятие, което има смисъл и при произволна редица от точки. Тясната връзка между разделена разлика и интерполационната формула на Нютон ни подсказва, че бихме могли да използваме обобщените разделени разлики за построяване на интерполационния полином с кратни възли, т.е. за решаване на интерполационната задача на Ермит. Сега ще въведем обобщените разлики и ще докажем някои техни свойства.

Нека $\bar{x} = (x_0, \dots, x_N)$ е произволна редица от точки. За нас ще бъде по-удобно да предполагаме, че те са подредени във възходящ ред, $x_0 \leq x_1 \leq \cdots \leq x_N$. Нека f е достатъчно гладка функция (т.е. имаща достатъчен брой непрекъснати производни), която е определена в $x_0, \dots, x_N$. Тъй като точките $\{x_i\}$ не са задължително различни, да си мислим, че те са разделени на n групи от съвпадащи точки. По-точно, нека първите ν_1 точки съвпадат с точката t_1, следващите ν_2 съвпадат с t_2 и т.н., последните ν_n точки съвпадат с t_n , където $t_1 < \cdots < t_n$. Ще записваме това условие накратко така

$$\bar{x} = (x_0, \dots, x_N) \equiv ((t_1, \nu_1), \dots, (t_n, \nu_n)).$$

Тук (t, ν) означава, че точката t е записана последователно ν пъти в редицата. Ясно е, че $N + 1 = \nu_1 + \cdots + \nu_n$.

Ще казваме, че полиномът p *интерполира* f в точките $\bar{x}$, ако $p \in \pi_N$ и

$$p^{(j)}(t_i) = f^{(j)}(t_i), \quad i = 1, \dots, n, \quad j = 0, \dots, \nu_i - 1,$$

т.е. ако p интерполира f във възлите $((t_1, \nu_1), \dots, (t_n, \nu_n))$ в смисъл на Ермит.

Определение. *Разделена разлика* на функцията f в точките $x_0, \dots, x_N$ ще наричаме коефициента пред x^N в полинома $p(x)$, който интерполира f в същите точки $x_0, \dots, x_N$.

Тази обобщена разлика ще бележим отново с $f[x_0, \dots, x_N]$.

Да отбележим, че това определение е еквивалентно с даденото преди определение на обикновената разделена разлика в случая, когато точките $x_0, \dots, x_N$ са различни. Еквивалентността следва веднага от Теорема 3.1.

Като пример, да намерим обобщената разделена разлика $f[a, \dots, a]$ на f в точката a с кратност $N + 1$. Известно е, че полиномът

$$p(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \cdots + \frac{f^{(N)}(a)}{N!}(x - a)^N,$$

построен по формулата на Тейлър, удовлетворява условието

$$p^{(j)}(a) = f^{(j)}(a), \quad j = 0, \ldots, N.$$

С други думи, полиномът p интерполира f в точката a с кратност $N+1$. От явния вид на p се вижда, че коефициентът му пред x^N е точно равен на $f^{(N)}(a)/N!$. Следователно, съгласно даденото по-горе определение,

$$(1) \qquad f[x_0, \ldots, x_N] = \frac{f^{(N)}(a)}{N!} \quad \text{при } x_0 = \cdots = x_N = a.$$

Основната причина за въвеждането на обобщени разделени разлики се разкрива в следното твърдение.

Теорема 1. *Нека $\bar{x} = (x_0, \ldots, x_N)$, $a \le x_0 \le \cdots \le x_N \le b$, са произволни точки. Да предположим, че f има N непрекъснати производни в $[a, b]$. Тогава полиномът*

$$(2) \qquad p(\bar{x}, f; t) := \sum_{k=0}^{N} f[x_0, \ldots, x_k](t - x_0) \cdots (t - x_{k-1})$$

интерполира f в точките $\bar{x}$.

Доказателство. Ще приложим индукция по броя на точките. При $N = 0$ имаме $p(\bar{x}, f; t) = f(x_0)$ и твърдението е очевидно вярно. Да допуснем, че теоремата е вярна за произволни N точки. Нека $\bar{x} = (x_0, \ldots, x_N)$ е множество от произволни $N+1$ точки, $x_0 \le \cdots \le x_N$. Както вече знаем от лекцията за интерполационна задача на Ермит, съществува единствен полином $H(t)$ от π_N, който интерполира f в точките $\bar{x}$. Ще покажем, че $H(t) \equiv p(\bar{x}, f; t)$. За целта да отбележим най-напред, че съгласно индукционното предположение полиномът

$$p_1(t) := \sum_{k=0}^{N-1} f[x_0, \ldots, x_k](t - x_0) \cdots (t - x_{k-1})$$

интерполира f в точките $x_0, \ldots, x_{N-1}$. Тъй като

$$p(\bar{x}, f; t) = p_1(t) + f[x_0, \ldots, x_N](t - x_0) \cdots (t - x_{N-1}),$$

то и $p(\bar{x}, f; t)$ ще интерполира f в $x_0, \ldots, x_{N-1}$. Оттук следва, че полиномът $R(t) := H(t) - p(\bar{x}, f; t)$ ще се анулира в $x_0, \ldots, x_{N-1}$. И така, R има поне N нули. Но водещият коефициент (този пред t^N) в $p(\bar{x}, f; t)$ е $f[x_0, \ldots, x_N]$ по построение, а коефициентът пред t^N в $H(t)$ е също $f[x_0, \ldots, x_N]$, съгласно определението на обобщена разделена разлика. Следователно, коефициентът пред t^N в $R(t)$ е равен на нула. Това означава, че $R(t)$ е алгебричен полином от степен най-много $N-1$. Видяхме вече, че $R(t)$ има поне N нули. Следователно, по основната теорема на алгебрата, $R(t) \equiv 0$ и оттук, $H(t) \equiv p(\bar{x}, f; t)$. Индукцията е завършена и теоремата е доказана.

Теорема 1 показва, че *интерполационната формула на Нютон остава в сила и при произволни (не задължително различни) възли* $x_0, \ldots, x_N$. Този факт ни позволява да използваме класическата формула на Нютон за решаване на по-сложната задача на Ермит. Практическото приложение на формула (2) изисква ефективен метод за пресмятане на обобщените разделени разлики на дадена функция. Сега ще покажем, че простата схема за пресмятане на обикновени разделени разлики може да се пригоди за пресмятане на разделени разлики в общия случай. За целта, най-напред ще докажем една лема, която има и самостоятелно значение.

Лема 2. *Нека* $\xi, t_1, \ldots, t_m$ *са произволни точки и* f *е достатъчно гладка функция, определена в тях. Тогава*

$$\{(x-\xi)f(x)\}[\xi, t_1, \ldots, t_m] = f[t_1, \ldots, t_m]\ . \tag{3}$$

Доказателство: Да поясним, че лявата страна на (3) е разделена разлика на функцията $(x-\xi)f(x)$ в точките $\xi, t_1, \ldots, t_m$.

Нека p е полиномът от π_{m-1}, който интерполира f в $t_1, \ldots, t_m$. Тогава полиномът $q(x) := (x-\xi)p(x)$ ще интерполира функцията $(x-\xi)f(x)$ в точките $\xi, t_1, \ldots, t_m$. Съгласно определението на обобщена разделена разлика, лявата страна на (3) е коефициентът пред x^m в q, а дясната страна съвпада с коефициента пред x^{m-1} в p. Но тези два коефициента са еднакви поради връзката $q(x) = (x-\xi)p(x)$. Равенството (3) е доказано.

Вече сме готови да пристъпим към извеждането на рекурентна връзка за обобщените разделени разлики.

Теорема 3. *Нека* f *има* k *непрекъснати производни в* $[a, b]$. *Тогава за произволни точки* $x_0 \le \ldots \le x_k$ *от* $[a, b]$ *е в сила рекурентната връзка*

$$f[x_0, \ldots, x_k] = \begin{cases} \frac{f[x_1,\ldots,x_k] - f[x_0,\ldots,x_{k-1}]}{x_k - x_0}, & \text{ако } x_0 < x_k \\[2ex] \frac{f^{(k)}(x_0)}{k!}, & \text{ако } x_0 = x_k. \end{cases} \tag{4}$$

Доказателство: Случаят $x_0 = x_k$ следва от доказаната вече формула (1). Да предположим сега, че $x_0 < x_k$. Тъй като разделената разлика е линеен функционал, то

$$\begin{aligned} (x_k - x_0)f[x_0, \ldots, x_k] &= \{(x_k - x + x - x_0)f\}[x_0, \ldots, x_k] \\ &= \{(x_k - x)f\}[x_0, \ldots, x_k] + \{(x - x_0)f\}[x_0, \ldots, x_k]\ . \end{aligned}$$

Но въз основа на Лема 2,

$$\begin{aligned} \{(x_k - x)f\}[x_0, \ldots, x_k] &= -f[x_0, \ldots, x_{k-1}], \\ \{(x - x_0)f\}[x_0, \ldots, x_k] &= f[x_1, \ldots, x_k]\ . \end{aligned}$$

Следователно, при $x_0 < x_k$ имаме

$$(x_k - x_0) f[x_0, \ldots, x_k] = f[x_1, \ldots, x_k] - f[x_0, \ldots, x_{k-1}] \ ,$$

което е точно връзката в (4). Теоремата е доказана.

Изчисляването на обобщените разлики, а оттам и построяването на интерполационния полином на Ермит, може да се организира в проста схема, основаваща се на рекурентната връзка (4). Ще я покажем върху един пример.

Задача: Да се построи полиномът p от π_3, който удовлетворява интерполационните условия

$$p(0) = 1, \quad p'(0) = 0, \quad p''(0) = 2, \quad p(1) = -1 \ .$$

Решение: В този случай имаме $x_0 = x_1 = x_2 = 0$, $x_3 = 1$. Съгласно интерполационната формула на Нютон (с обобщени разделени разлики),

$$\begin{aligned} p(x) &= p(x_0) + p[x_0, x_1](x - x_0) + p[x_0, x_1, x_2](x - x_0)(x - x_1) \\ &+ p[x_0, x_1, x_2, x_3](x - x_0)(x - x_1)(x - x_2) \\ &= p(0) + p[0,0]x + p[0,0,0]x^2 + p[0,0,0,1]x^3 \ . \end{aligned}$$

За изчисляването на разделените разлики ще използваме таблицата по-долу, където първите 2 колони съдържат данните, а следващите се попълват въз основа на рекурентната връзка (4).

Таблица

x_i	$p(x_i)$	$p[\cdot,\cdot]$	$p[\cdot,\cdot,\cdot]$	$p[\cdot,\cdot,\cdot,\cdot]$
0	1			
		$\boxed{0}$		
0	1		$\boxed{1}$	
		0		$\boxed{-3}$
0	1		-2	
		-2		
1	-1			

Отбелязаните в квадаратче числа са търсените коефициенти. Получаваме

$$p(x) = 1 + 0.x + 1.x^2 - 3.x^3 = -3x^3 + x^2 + 1 \ .$$

Една непосредствена проверка показва, че намереният полином удовлетворява исканите интерполационни условия.

Непрекъснатост на разделената разлика. След като имаме определението на разделена разлика за всяка редица $x_0, \ldots, x_N$ от различни точки, бихме могли да разширим това определение и за произволни точки по непрекъснатост (т.е. чрез граничен преход). Например, можехме да определим разделена разлика на f в точка a с кратност 2 с равенството

$$f[a,a] = \lim_{h\to 0} f[a, a+h] \ ,$$

което е най-естественото разширение на това понятие. Тъй като

$$\lim_{h\to 0} f[a, a+h] = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = f'(a) \ ,$$

ако f е диференцируема функция в a, то при този подход бихме получили

$$f[a,a] = f'(a) \ .$$

Но това е точно резултатът, който вече получихме и при възприетото от нас определение. Оказва се, че двата подхода водят до един и същ резултат не само в този частен случай. Те са еквивалентни. Това твърдение следва от свойството *непрекъснатост* на въведената от нас обобщена разделена разлика.

Да докажем непрекъснатостта най-напред в един частен случай.

Лема 4. *Нека $x_i \to a$ за $i = 0, \ldots, N$. Тогава, ако f има непрекъсната N–та производна, то*

$$f[x_0, \ldots, x_N] \to f[a, \ldots, a] \ .$$

Доказателство. Нека $x_0, \ldots, x_N$ са произволни точки. Без ограничение на общността да приемем, че $x_0 \le \ldots \le x_N$. Нека p е полиномът от π_N, който интерполира f в $\bar{x} = (x_0, \ldots, x_N)$. По-точно, ако

$$\bar{x} = (x_0, \ldots, x_N) = ((t_1, \nu_1), \ldots, (t_n, \nu_n)) \ ,$$

то

$$p^{(j)}(t_i) = f^{(j)}(t_i), \quad i = 1, \ldots, n, \ j = 0, \ldots, \nu_i - 1 \ .$$

Следователно разликата $f(x) - p(x)$ има поне $N+1$ нули, броейки кратностите. Тогава, по теоремата на Рол, $f'(x) - p'(x)$ ще има поне N нули и т.н., $f^{(N)}(x) - p^{(N)}(x)$ ще има поне една нула ξ и тази нула лежи в интервала $[x_0, x_N]$. Имаме

$$f^{(N)}(\xi) = p^{(N)}(\xi) \ .$$

Но коефициентът пред x^N в $p(x)$ е равен на $f[x_0, \ldots, x_N]$. Следователно $p^{(N)}(\xi) = N!\ f[x_0, \ldots, x_N]$ и оттук

$$f[x_0, \ldots, x_N] = \frac{f^{(N)}(\xi)}{N!} . \tag{5}$$

И така, за произволни точки $x_0, \ldots, x_N$ ($x_0 \leq \ldots \leq x_N$) съществува точка ξ от $[x_0, x_N]$ такава, че е в сила равенството (5). Този факт ни е добре известен за обикновените разделени разлики. Сега вече лесно можем да докажем лемата. Наистина, да оставим в (5) x_i да клони към a, за $i = 0, \ldots, N$. Тогава поради неравенството $x_0 \leq \xi \leq x_N$ и по лемата за двамата милиционери, редицата от съответните точки ξ ще клони също към a. Тъй като и $f^{(N)}$ е непрекъсната по условие, от (5) получаваме

$$\lim_{x_i \to a} f[x_0, \ldots, x_N] = \lim_{\xi \to a} \frac{f^{(N)}(\xi)}{N!} = \frac{f^{(N)}(a)}{N!} = f[a, \ldots, a] \text{ (съгласно (1))}$$

Лемата е доказана.

Теорема 5. *Нека $\bar{y} = (y_0, \ldots, y_N)$ е произволен набор от точки и $\bar{x} \to \bar{y}$, т.е.*

$$|x_i - y_i| \to 0 \text{ за } i = 0, \ldots, N .$$

Тогава, ако f има непрекъсната N–та производна, то $f[x_0, \ldots, x_N] \to f[y_0, \ldots, y_N]$.

Доказателство. С други думи, разделената разлика е непрекъсната функция на своите аргументи. (Разбира се, тук се изисква функцията f да бъде достатъчно гладка.) Ще докажем теоремата по индукция. За $N = 1$ твърдението е очевидно при $y_0 < y_1$, а при $y_0 = y_1 = a$ имаме $f[y_0, y_1] = f'(a)$ и непрекъснатостта следва от разсъжденията в примера по-горе. Да допуснем сега, че $f[t_1, \ldots, t_N]$ е непрекъсната функция на $t_1, \ldots, t_N$ в областта $t_1 \leq t_2 \leq \ldots \leq t_N$. Ще докажем, че $f[t_0, \ldots, t_N]$ е непрекъсната функция в произволна фиксирана точка $\bar{y} = (y_0, \ldots, y_N)$, $y_0 \leq \ldots \leq y_N$, като използваме рекурентната връзка (4). При $y_0 = y_N$ твърдението беше доказано в Лема 4. Нека $y_0 < y_N$. Тъй като $\bar{x} \to \bar{y}$, то при $\bar{x}$ достатъчно близко до $\bar{y}$ ще имаме $x_0 < x_N$. Тогава по рекурентната връзка (4), като използваме индукционното предположение, получаваме

$$\begin{aligned}
\lim_{\bar{x} \to \bar{y}} f[x_0, \ldots, x_N] &= \lim_{\bar{x} \to \bar{y}} \frac{f[x_1, \ldots, x_N] - f[x_0, \ldots, x_{N-1}]}{x_N - x_0} \\
&= \frac{1}{y_N - y_0} \left\{ \lim_{\bar{x} \to \bar{y}} f[x_1, \ldots, x_N] - \lim_{\bar{x} \to \bar{y}} f[x_0, \ldots, x_{N-1}] \right\} \\
&= \frac{f[y_1, \ldots, y_N] - f[y_0, \ldots, y_{N-1}]}{y_N - y_0} = f[y_0, \ldots, y_N] .
\end{aligned}$$

Теоремата е доказана.

Да отбележим тук, че остатъкът при интерполиране в произволни точки $x_0, \ldots, x_N$ може да се представи във вида

$$f(x) - p(\bar{x}, f; x) = f[x_0, \ldots, x_N, x](x - x_0) \cdots (x - x_N).$$

Това следва чрез граничен преход от изведената вече формула в случая на различни възли.

Следващото твърдение ни дава правило за пресмятане на разделена разлика на произведение на две функции. То наподобява известното правило на Лайбниц за производна на произведение.

Лема на Попович. *За произволни точки $x_0, x_1, \ldots, x_n$ и достатъчно гладки функции f и g, определени в тях, е в сила представянето*

$$(fg)[x_0, \ldots, x_n] = \sum_{k=0}^{n} f[x_0, \ldots, x_k]\, g[x_k, \ldots, x_n] \,.$$

Доказателство. Ще приложим индукция по броя на точките. За една точка имаме $(fg)[x_0] = f(x_0)g(x_0) = f[x_0]g[x_0]$ и твърдението е очевидно вярно. Да допуснем, че то е вярно за произволни n точки. Нека $x_0, \ldots, x_n$ са произволни $n + 1$ точки. Представяме f по интерполационната формула на Нютон:

$$f(x) = f(x_0) + f[x_0, x](x - x_0).$$

Тогава

$$\begin{aligned} (fg)[x_0, \ldots, x_n] &= \{(f(x_0) + f[x_0, x](x - x_0))g(x)\}\, [x_0, \ldots, x_n] \\ &= f(x_0)g[x_0, \ldots, x_n] + \{f[x_0, x](x - x_0))g(x)\}\, [x_0, \ldots, x_n] \,. \end{aligned}$$

Но по Лема 2,

$$\{f[x_0, x](x - x_0))g(x)\}\, [x_0, \ldots, x_n] = \{f[x_0, x]g(x)\}\, [x_1, \ldots, x_n] \,.$$

Прилагаме индукционното предположение за последното произведение и получаваме

$$\begin{aligned} (fg)[x_0, \ldots, x_n] &= f(x_0)g[x_0, \ldots, x_n] + \sum_{k=1}^{n} f[x_0, x][x_1, \ldots, x_k]\, g[x_k, \ldots, x_n] \\ &= \sum_{k=0}^{n} f[x_0, \ldots, x_k]\, g[x_k, \ldots, x_n] \,. \end{aligned}$$

Тук използвахме равенството $f[x_0, x][t_1, \ldots, t_m] = f[x_0, t_1, \ldots, t_m]$, което може да се докаже лесно по индукция (по m). Доказателството е завършено.

7. СИСТЕМИ НА ЧЕБИШОВ. ИНТЕРПОЛИРАНЕ С ТРИГОНОМЕТРИЧНИ ПОЛИНОМИ.

Обща интерполационна задача. Нека $\varphi_0(x), \dots, \varphi_n(x)$ са линейно независими и непрекъснати функции в $[a, b]$. Линейните комбинации $a_0\varphi_0(x) + \dots + a_n\varphi_n(x)$ ще наричаме *обобщени полиноми* по системата $\{\varphi_i\}$. Ще разгледаме задачата за интерполиране с обобщени полиноми. При дадени възли $x_0 < \dots < x_n$ в $[a, b]$ и стойности $y_0, \dots, y_n$ търсим обообщен полином $\varphi(x) = a_0\varphi_0(x) + \dots + a_n\varphi_n(x)$, който да удовлетворява интерполационните условия

$$(1) \qquad a_0\varphi_0(x_k) + \dots + a_n\varphi_n(x_k) = y_k, \quad k = 0, \dots n.$$

Но (1) е линейна система по отношение на $a_0, \dots, a_n$. Следователно интерполационната задача (1) има единствено решение при всеки избор на $\{y_k\}$ тогава и само тогава, когато нейната детерминанта е различна от нула. Нищо повече не може да се каже в този най-общ случай.

Интерес представляват тези системи от функции $\{\varphi_i\}_0^n$, при които интерполационната задача (1) има единствено решение при всеки избор на възлите $x_0 < \dots < x_n$ в $[a, b]$ и при всеки избор на стойностите $y_0, \dots, y_n$. Например алгебричната система $\varphi_k(x) = x^k$, $k = 0, \dots n$, е такава. Сега ще разгледаме един клас от системи $\{\varphi_i\}_0^n$, които удовлетворяват това изискване и следователно се явяват естествени обобщения на класическите алгебрични полиноми.

Определение. Казваме, че функциите $\varphi_0(x), \dots, \varphi_n(x)$ образуват *система на Чебишов* в интервала I, ако всеки ненулев обобщен полином по тази система има най-много n различни нули в I.

Да припомним, че $a_0\varphi_0(x) + \dots + a_n\varphi_n(x)$ е *ненулев* обобщен полином, ако поне един от коефициентите му $\{a_i\}_0^n$ е различен от нула. Системите на Чебишов се наричат още *T-системи* или *чебишови системи*.

Да означим с $D[x_0, \dots, x_n]$ матрицата на системата (1).

Теорема 1. *Функциите $\varphi_0, \dots, \varphi_n$ образуват система на Чебишов в интервала I тогава и само тогава, когато*

$$\det D[x_0, \dots, x_n] \neq 0$$

при всеки избор на точките $x_0 < \dots < x_n$ в I.

Доказателство. Нека $\varphi_0, \dots, \varphi_n$ е система на Чебишов в I. Да допуснем, че $\det D[x_0, \dots, x_n] = 0$ при някои $x_0 < \dots < x_n$ от I. Тогава между стълбовете на матрицата $D[x_0, \dots, x_n]$ има линейна зависимост, т.е. съществуват числа $b_0, \dots, b_n$, поне едно от които е различно от нула и такива, че

$$(2) \qquad b_0\varphi_0(x_k) + b_1\varphi_1(x_k) + \dots + b_n\varphi_n(x_k) = 0 \quad \text{за} \quad k = 0, \dots, n.$$

Но тези равенства означават, че ненулевият обобщен полином $\varphi(x) := b_0\varphi_0(x)+\cdots+b_n\varphi_n(x)$ се анулира в $n+1$ различни точки, именно в $x_0,\ldots,x_n$. Това противоречи на определението на чебишова система. Следователно $\det D[x_0,\ldots,x_n] \neq 0$.

Обратно, нека $\det D[x_0,\ldots,x_n] \neq 0$ при всеки избор на $x_0 < \cdots < x_n$ в I. Да допуснем, че системата $\varphi_0,\ldots,\varphi_n$ не е чебишова. Тогава съществува ненулев обобщен полином $\varphi(x) := b_0\varphi_0(x) + \cdots + b_n\varphi_n(x)$ и $n+1$ различни точки $x_0 < \cdots < x_n$ в I такива, че $\varphi(x_k) = 0$ за $k = 0,\ldots,n$. Но това означава, че хомогенната система (2) има ненулево решение $b_0,\ldots,b_n$. Следователно нейната детерминанта е нула, т.е. $\det D[x_0,\ldots,x_n] = 0$, противоречие. Теоремата е доказана.

Едно непосредствено следствие от доказаното тук свойство на чебишовите системи е следната интерполационна теорема.

Теорема 2. *Нека функциите $\varphi_0(x),\ldots,\varphi_n(x)$ образуват система на Чебишов в интервала I. Тогава при дадени произволни възли $x_0 < \cdots < x_n$ от I и стойности $y_0,\ldots,y_n$ интерполационната задача*

$$a_0\varphi_0(x_k) + \cdots + a_n\varphi_n(x_k) = y_k, \quad k = 0,\ldots,n,$$

има единствено решение.

Действително, интерполационната задача има единствено решение тогава и само тогава, когато $\det D[x_0,\ldots,x_n] \neq 0$ и Теорема 2 следва веднага от доказаната Теорема 1.

Примери на T-системи:

1) Функциите $1, x, x^2,\ldots,x^n$ образуват T-система във всеки подинтервал на реалната права.

2) Функциите $x, x^3, x^5,\ldots,x^{2n+1}$ образуват T-система във всеки подинтервал на $(0,\infty)$.

3) Функциите $1, x^{\alpha_1},\ldots,x^{\alpha_n}$ образуват T-система във всеки подинтервал на $(0,\infty)$ при произволни реални числа $0 < \alpha_1 < \cdots < \alpha_n$.

4) Функциите $\frac{1}{x-x_0},\ldots,\frac{1}{x-x_n}$ образуват T-система във всеки подинтервал, който не съдържа точките $x_0,\ldots,x_n$.

5) Ако $p(x)$ е строго монотонна и непрекъсната функция в $[-1,1]$ и такава, че $p(-1) = -1$, $p(1) = 1$, то функциите $1, p(x), p^2(x),\ldots,p^n(x)$ образуват T-система в $[-1,1]$.

Интерполиране с тригонометрични полиноми. Всеки израз от вида

$$t_n(x) = a_0 + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx)$$

се нарича *тригонометричен полином* от ред n. Тригонометричните полиноми са 2π-периодични функции. Те са удобен апарат за приближение на функции, описващи периодични явления. Тук ще се занимаем със задачата за интерполиране на периодични функции с тригонометрични полиноми. Най-напред ще дадем една оценка за броя на нулите на тригонометричен полином в интервал с дължина 2π.

Лема 1. *Всеки ненулев тригонометричен полином от ред n има не повече от $2n$ различни нули в $[0, 2\pi)$.*

Доказателство. Да направим смяната $z = e^{ix}$ в $t_n(x)$. При $x \in [0, 2\pi)$ променливата z ще описва единичната окръжност. Тъй като

$$\cos kx = \frac{e^{ikx} + e^{-ikx}}{2} = \frac{z^k + z^{-k}}{2},$$

$$\sin kx = \frac{e^{ikx} - e^{-ikx}}{2i} = \frac{z^k - z^{-k}}{2i},$$

то

$$\begin{aligned} t_n(x) &= a_0 + \frac{1}{2}\sum_{k=1}^{n}[(a_k - ib_k)z^k + (a_k + ib_k)z^{-k}] \\ &= \sum_{k=-n}^{n} c_k z^k = z^{-n}\sum_{k=0}^{2n} c_{k-n} z^k =: z^{-n}P_{2n}(z) \end{aligned}$$

където $c_0 = a_0$,

$$\begin{aligned} c_k &= \frac{1}{2}(a_k - ib_k), \quad k = 1, \dots, n, \\ c_{-k} &= \frac{1}{2}(a_k + ib_k), \quad k = 1, \dots, n. \end{aligned}$$

Да допуснем сега, че ненулевият тригонометричният полином $t_n(x)$ се анулира в $2n+1$ различни точки от $[0, 2\pi)$. Тогава от горната връзка следва, че и алгебричният полином $P_{2n}(z)$ ще се анулира в $2n+1$ различни точки от единичната окръжност, тъй като $z^{-n} \neq 0$. Но P_{2n} е алгебричен полином от степен $2n$. Съгласно основната теорема на алгебрата, той има $2n$ нули в цялата комплексна равнина или е тъждествено равен на нула. Следователно $P_{2n} \equiv 0$. Оттук следва, че $c_k = 0$, $k = -n, \dots, n$, което от своя страна влече равенствата $a_k = 0$, $b_k = 0$ за всяко допустимо k. Получихме, че $t_n(x) \equiv 0$, което противоречи на условието. Лемата е доказана.

Непосредствено от лемата следва, че функциите

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \dots, \cos nx, \sin nx$$

образуват система на Чебишов във всеки интервал от вида $[\alpha, \alpha + 2\pi)$. Това ни дава възможност да формулираме следващото твърдение като частен случай на интерполацията с обобщени полиноми по чебишови системи, разгледана в началото на тази лекция.

Теорема 2. *Нека* $\alpha \le x_0 < \cdots < x_{2n} < \alpha + 2\pi$. *Тогава за всяка функция* f *определена в точките* $\{x_i\}_0^{2n}$ *съществува единствен тригонометричен полином* t_n *от ред* n *такъв, че*

$$t_n(x_i) = f(x_i), \quad i = 0, \dots, 2n. \tag{1}$$

Сега ще търсим явен израз за този интерполационен тригонометричен полином. Следвайки аналогията с формулата на Лагранж, ще търсим полинома във вида

$$t_n(x) = \sum_{k=0}^{2n} f(x_k)\lambda_k(x) \tag{2}$$

Твърдим, че

$$\lambda_k(x) = \prod_{i=0, i\neq k}^{2n} \frac{\sin \frac{x-x_i}{2}}{\sin \frac{x_k - x_i}{2}}. \tag{3}$$

Функциите λ_k удовлетворяват условията

$$\begin{aligned} \lambda_k(x_j) &= 0 \quad \text{за} \quad j \neq k\ , \\ \lambda_k(x_k) &= 1, \end{aligned}$$

откъдето веднага следва, че изразът (2) ще удовлетворява интерполационните условия (1). Остава само да се убедим, че $\lambda_k(x)$, а оттук и $t_n(x)$, е действително тригонометричен полином от ред n. За целта ще използваме индукция по n.

При $n = 1$ изразът $\lambda_k(x)$ е произведение от два синуса, т.е. от вида

$$\sin \frac{x-\alpha}{2} \sin \frac{x-\beta}{2}\ .$$

Въз основа на известни формули за тригонометрични преобразования получаваме последователно

$$\begin{aligned} \sin \frac{x-\alpha}{2} \sin \frac{x-\beta}{2} &= \frac{1}{2}\left(\cos \frac{\beta-\alpha}{2} - \cos\left(x - \frac{\alpha+\beta}{2}\right)\right) \\ &= \frac{1}{2}\cos\frac{\beta-\alpha}{2} - \frac{1}{2}\cos x \cos \frac{\alpha+\beta}{2} - \frac{1}{2}\sin x \sin \frac{\alpha+\beta}{2} \\ &= A_0 + A_1 \cos x + B_1 \sin x \end{aligned}$$

и следователно λ_k е тригонометричен полином от ред 1. Да допуснем сега, че всяко произведение от $n-1$ двойки синуси е тригонометричен полином от ред $n-1$. Да разгледаме произволен израз $\lambda_k(x)$ от вида (3). Очевидно той може да се запише по следния начин

$$\begin{aligned}\lambda_k(x) &= C\prod_{i=0,i\neq k}^{2n}\sin\frac{x-x_i}{2}\\ &= C\left\{\prod_{i=1}^{n-1}\sin\frac{x-\alpha_i}{2}\sin\frac{x-\beta_i}{2}\right\}\left\{\sin\frac{x-\alpha}{2}\sin\frac{x-\beta}{2}\right\},\end{aligned}$$

където C е константа. Но съгласно индукционното предположение и двата израза в големите скоби по-горе са тригонометрични полиноми (от ред $n-1$ и 1, съответно) и следователно

$$\lambda_k(x)=\left[a_0+\sum_{k=1}^{n-1}(a_k\cos kx+b_k\sin kx)\right][A_0+A_1\cos x+B_1\sin x].$$

Като извършим умножението, получаваме за λ_k израз, който е линейна комбинация на тригонометрични функции от вида

$$\cos kx\cos mx,\quad \cos kx\sin mx,\quad \sin kx\sin mx,$$

където $k+m\le n$. Но съгласно известните тригонометрични формули

$$\begin{aligned}\cos kx\,\sin mx &= \frac{1}{2}[\sin(k+m)x-\sin(k-m)x],\\ \cos kx\,\cos mx &= \frac{1}{2}[\cos(k-m)x+\cos(k+m)x],\\ \sin kx\,\sin mx &= \frac{1}{2}[\cos(k-m)x-\cos(k+m)x].\end{aligned}$$

Следователно $\lambda_k(x)$ се представя като линейна комбинация на

$$\sin kx,\ \cos kx,\quad k=0,\dots,n,$$

което означава, че $\lambda_k(x)$ е тригонометричен полином от ред n. Индукцията е завършена. И така, изразът $t_n(x)$ е тригонометричен полином от ред n и той удовлетворява интерполационните условия (1). По-нататък, при дадена функция f, ще бележим този полином с $\tau_n(f;x)$. Да формулираме доказаното твърдение.

Теорема 3. *Нека $\{x_k\}_{k=0}^{2n}$ са произволни точки, такива че $\alpha\le x_0< x_1<\cdots<x_{2n}<\alpha+2\pi$ за някакво α и f е произволна функция определена в тях. Тогава*

$$\tau_n(f;x)=\sum_{k=0}^{2n}f(x_k)\prod_{i=0,i\neq k}^{2n}\frac{\sin\frac{x-x_i}{2}}{\sin\frac{x_k-x_i}{2}}$$

е единственият тригонометричен полином от ред n, който интерполира f в $x_0, \ldots, x_{2n}$.

Когато възлите $\{x_k\}$ са равноотдалечени, интерполационният полином може да се запише в по-удобен вид. Сега ще дадем този запис. За целта най-напред ще докажем тригонометричното тъждество

$$\frac{1}{2} + \cos x + \cos 2x + \cdots + \cos nx = \frac{\sin\left(n+\frac{1}{2}\right)x}{2\sin\frac{x}{2}}. \tag{4}$$

Имаме

$$\begin{aligned}
&2\sin\frac{x}{2}\left\{\frac{1}{2} + \cos x + \cos 2x + \cdots + \cos nx\right\} \\
&= \sin\frac{x}{2} + 2\cos x\,\sin\frac{x}{2} + 2\cos 2x\,\sin\frac{x}{2} + \cdots + 2\cos nx\,\sin\frac{x}{2} \\
&= \sin\frac{x}{2} + \left(\sin\frac{3x}{2} - \sin\frac{x}{2}\right) + \left(\sin\frac{5x}{2} - \sin\frac{3x}{2}\right) + \cdots \\
&\quad + \left(\sin(n+\frac{1}{2})x - \sin(n-\frac{1}{2})x\right) \\
&= \sin\left(n+\frac{1}{2}\right)x,
\end{aligned}$$

което е еквивалентно на разглежданото тъждество.

Функцията

$$D_n(x) := \frac{\sin\left(n+\frac{1}{2}\right)x}{2\sin\frac{x}{2}}$$

се нарича *ядро на Дирихле*. Да обърнем внимание върху някои нейни свойства. От (4) се вижда веднага, че

$$D_n(0) = n + \frac{1}{2}.$$

Да означим с $\{x_k\}$ равноотдалечените възли

$$x_k = \frac{2k\pi}{2n+1}, \quad k = 0, \ldots, 2n.$$

Ядрото на Дирихле се анулира във всички възли x_k, които са различни от $x_0 = 0$. Наистина

$$D_n(x_k) = \frac{\sin\left(\frac{2n+1}{2}\,\frac{2k\pi}{2n+1}\right)}{2\sin\frac{k\pi}{2n+1}} = \frac{\sin k\pi}{2\sin\frac{k\pi}{2n+1}} = 0 \quad \text{за} \quad k = 1, 2, \ldots, 2n.$$

От тези две свойства следва, че функцията

$$\lambda_k(x) := \frac{2}{2n+1} D_n(x - x_k)$$

ще удовлетворява интерполационните условия

$$\lambda_k(x_i) = \delta_{ki}, \quad k, i = 0, \ldots, 2n.$$

Наистина

$$\lambda_k(x_k) = \frac{2}{2n+1} D_n(x_k - x_k) = \frac{2}{2n+1}\left(n + \frac{1}{2}\right) = 1,$$

а при $i \neq k$:

$$\begin{aligned}
\lambda_k(x_i) &= \frac{2}{2n+1} D_n(x_i - x_k) \\
&= \frac{2}{2n+1} D_n\left((i-k)\frac{2\pi}{2n+1}\right) \\
&= \frac{2}{2n+1} D_n(x_{|i-k|}) = 0.
\end{aligned}$$

Но $\lambda_k(x)$ е тригонометричен полином от ред n (това се вижда например от тъждеството (4)). Следователно $\lambda_k(x)$, $k = 0, \ldots, 2n$, могат да служат като базисни полиноми при интерполиране в точките $\{x_k\}$. Тогава формулата за интерполиране при равноотдалечени възли добива вида

$$\tau_n(f; x) = \sum_{k=0}^{2n} f(x_k) \frac{2}{2n+1} D_n(x - x_k),$$

или окончателно

$$\tau_n(f; x) = \frac{1}{2n+1} \sum_{k=0}^{2n} f(x_k) \frac{\sin\left((n + \frac{1}{2})(x - x_k)\right)}{\sin \frac{x - x_k}{2}}. \tag{5}$$

8. БЪРЗО ПРЕОБРАЗОВАНИЕ НА ФУРИЕ.

Знаем вече, че при дадени възли $0 \le x_0 < \cdots < x_{2n} < 2\pi$ и стойности $f(x_0), \ldots, f(x_{2n})$ съществува единствен тригонометричен полином $\tau(x)$ от ред n, който удовлетворява интерполационните условия

$$\tau(x_k) = f(x_k), \quad k = 0, \ldots, 2n.$$

Да представим $\tau(x)$ в обичайния му вид

$$\tau(x) = \frac{1}{2}a_0 + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx)$$

в случая когато нулите са равноотдалечени, т.е. при

$$x_k = \frac{2k\pi}{2n+1}, \quad k = 0, \ldots, 2n.$$

За целта да припомним, че съгласно формула (7.5),

$$\begin{aligned}
\tau(x) &= \frac{1}{2n+1}\sum_{k=0}^{2n} f(x_k)\frac{\sin\left(n+\frac{1}{2}\right)(x-x_k)}{\sin\frac{x-x_k}{2}} \\
&= \frac{2}{2n+1}\sum_{k=0}^{2n} f(x_k)\left\{\frac{1}{2} + \cos(x-x_k) + \cdots + \cos n(x-x_k)\right\}.
\end{aligned}$$

Оттук, като използваме, че

$$\cos m(x-x_k) = \cos mx \cos mx_k + \sin mx \sin mx_k$$

получаваме

$$\begin{aligned}
\tau(x) &= \frac{2}{2n+1}\sum_{k=0}^{2n} f(x_k)\left(\frac{1}{2} + \sum_{m=1}^{n} \cos mx_k \cos mx + \sin mx_k \sin mx\right) \\
&= \frac{1}{2}A_0 + \sum_{m=1}^{n}\left(A_m \cos mx + B_m \sin mx\right),
\end{aligned}$$

където

$$A_m := \frac{2}{2n+1}\sum_{k=0}^{2n} f(x_k)\cos\frac{2km\pi}{2n+1},$$

(1)

$$B_m := \frac{2}{2n+1}\sum_{k=0}^{2n} f(x_k)\sin\frac{2km\pi}{2n+1}.$$

Следователно намирането на интерполационния тригонометричен полином τ се свежда до пресмятане на коефициентите A_m и B_m по формулите (1).

Преобразованието (където i е имагинерната единица)

$$f(x) \to \hat{f}(t) := \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-itx}\,dx$$

се нарича *преобразование на Фурие* за f, а преобразованието

$$(f_0, f_1, \dots, f_{N-1}) \to (C_0, \dots, C_{N-1}),$$

където

$$C_m = \frac{1}{N}\sum_{k=0}^{N-1} f_k e^{-2\pi kim/N}, \quad m = 0, \dots, N-1,$$

се нарича *дискретно преобразование на Фурие.* Вижда се, че нашите коефициенти A_m и B_m са свързани с C_m при $N = 2n+1$. Имаме

$$A_m - iB_m = 2C_m\ . \tag{2}$$

Да отбележим тук, че *за пресмятането на всичките коефициенти* $\{C_m\}_{m=0}^{N-1}$ *са необходими* N^2 *умножения,* в които участват стойностите $f_0, \dots, f_{N-1}$.

Сега ще дадем един бърз алгоритъм за пресмятане на $\{C_m\}_{m=0}^{N-1}$ при всяко N, а оттук и метод за пресмятане на коефициентите $\{A_m\}$, $\{B_m\}$ въз основа на връзката (2). Този метод се нарича *бързо преобразование на Фурие.* Той е предложен през 1965 г. от Кули и Тъки.

Нека N се разлага на множители $N = pq$. Числата m и k се представят по единствен начин във вида

$$\begin{aligned} m &= m_1 q + m_0, \quad 0 \le m_0 < q \\ k &= k_1 p + k_0, \quad 0 \le k_0 < p. \end{aligned}$$

Не е трудно да се провери, че

$$\frac{mk}{N} = \frac{(m_1 q + m_0)k_1 p + mk_0}{N} = m_1 k_1 + \frac{m_0 k_1}{q} + \frac{mk_0}{N}.$$

Като използваме факта, че $e^{-2\pi ij} = 1$ за всяко цяло j, получаваме

$$\begin{aligned} C_m &= \frac{1}{N}\sum_{k=0}^{N-1} f_k e^{-2\pi i\frac{mk}{N}} \\ &= \frac{1}{pq}\sum_{k_0=0}^{p-1}\sum_{k_1=0}^{q-1} f_{k_1p+k_0} e^{-2\pi i\left(\frac{m_0k_1}{q}+\frac{mk_0}{N}\right)} \end{aligned}$$

$$= \frac{1}{p}\sum_{k_0=0}^{p-1}\left\{\frac{1}{q}\sum_{k_1=0}^{q-1} f_{k_1p+k_0}e^{-2\pi i\frac{m_0k_1}{q}}\right\}e^{-2\pi i\frac{mk_0}{N}}.$$

Следователно

$$C_m = \frac{1}{p}\sum_{k_0=0}^{p-1} C_{k_0}^{(1)} e^{-2\pi i\frac{mk_0}{N}}, \tag{3}$$

където с $C_{k_0}^{(1)}$ сме означили израза в големите скоби по-горе. Той е подобен на този за C_m, но със значително по-малко на брой събираеми. Формулата (3) представлява една рекурентна връзка, по която се пресмятат коефициентите C_m.

Когато N е степен на двойката ($N = 2^s$) редукцията на C_m към $C_{k_0}^{(1)}$, оттам към следващите $C_j^{(2)}$ и т.н., се извършва по прости и удобни формули, върху които се основава един бърз алгоритъм за пресмятане на C_m.

За пример да разгледаме случая $p = q \approx \sqrt{N}$. Тогава за пресмятането на всичките $C_{k_0}^{(1)}$ са необходими pq^2 умножения. Наистина, при всяко фиксирано k_0 $(0 \le k_0 \le p-1)$ и всяко фиксирано m_0 $(0 \le m_0 \le q-1)$, за пресмятането на $C_{k_0}^{(1)}$ по формулата

$$C_{k_0}^{(1)} = \frac{1}{q}\sum_{k_1=0}^{q-1} f_{k_1p+k_0}e^{-2\pi i\frac{m_0k_1}{q}}$$

се използват q умножения

$$f_{k_1p+k_0}\cdot e^{-2\pi i\frac{m_0k_1}{q}}, \quad k_1 = 0,\dots,q-1$$

или общо $p\cdot q\cdot q = pq^2$ умножения.

След това за пресмятане на всички C_m по формулата (3) са необходими още qp^2 умножения: При всяко фиксирано m, $(0 \le m \le N-1)$, се извършват p умножения (за $k_0 = 0,\dots,p-1$) или $Np = qp^2$ умножения. Общо получаваме

$$pq^2 + qp^2 = pq(p+q) \approx 2N\sqrt{N} = 2N^{3/2},$$

което, при големи N, е значително по-малко от N^2.

9. СПЛАЙН-ФУНКЦИИ. ИНТЕРПОЛИРАНЕ С КУБИЧНИ СПЛАЙНИ

Точността на приближение на дадена функция $f(x)$ в крайния интервал $[a, b]$ зависи съществено от дължината на интервала и степента на алгебричния полином. Тъй като компютърните пресмятания с полиноми от висока степен водят до известни проблеми, то желателно е да се използват полиноми от невисока степен. Тогава единственият шанс за увеличаване на точността на приближение идва от работа върху малки интервали. Ако интервалът $[a, b]$ е голям, той се разделя на малки подинтервали $[x_i, x_{i+1}]$, $i = 0, \dots, m$, и $f(x)$ се приближава в $[x_i, x_{i+1}]$ с алгебричен полином $p_i(x)$ от някаква ниска степен r. По този начин получаваме приближението

$$f(x) \approx P(x) := p_i(x) \text{ за } x \in (x_i, x_{i+1}).$$

Функцията $P(x)$ представя една на части полиномиална крива, която приближава графиката на f с определена точност. В общия случай $P(x)$ е прекъсната в точките $x_1, \dots, x_m$. Ако f описва гладък процес, то желателно е и приближаващата функция да бъде гладка. За да се постигне това, на полиномиалните части p_i се налага допълнителното условие да се свързват гладко, т.е. производните на $p_{i-1}(x)$ и $p_i(x)$ до определен ред да съвпадат в точката на свързване x_i. В резултат се получава една гладка крива, която приближава добре f. Такива криви се наричат сплайн-функции. Наименованието идва от един стар уред за чертане на гладки криви през зададени точки, наречен "сплайн".

Това е един от начините да се обясни появата на сплайн-функциите в математиката – като апарат, който е роден от нуждите на практиката. Интересните свойства на сплайн-функциите и дълбоките им връзки с други направления в математиката обаче показват, че появата на сплайн-функциите е обусловена от вътрешната логика на развитие на самата математика. Теорията на сплайн-функциите е една от най-бурно развиващите се области на анализа в последните 30 години.

Определение. Функцията $s(x)$, се нарича *сплайн-функция* от степен r с възли $x_1 < \dots < x_n$, ако:

1. $s(x)$ е полином от степен най-много r във всеки подинтервал (x_i, x_{i+1}), $\quad i = 0, \dots, n, \quad (x_0 = -\infty,\ x_{n+1} = \infty.)$;

2. $s(x), s'(x), \dots, s^{(r-1)}(x)$ са непрекъснати функции в $(-\infty, \infty)$.

Оттук нататък с $S_r(x_1, \dots, x_n)$ ще отбелязваме множеството от сплайн-функции от степен r с възли в точките $x_1, \dots, x_n$. Вместо сплайн-функция понякога ще пишем накратко "сплайн".

Непосредствено от определението следват свойствата:

1. Ако $s \in S_r(x_1, \dots, x_n)$, то s' е сплайн от степен $r-1$ със същите възли.

2. Ако $s \in S_r(x_1, \dots, x_n)$, то r-тата производна на s е на части постоянна функция с евентуални скокове в точките $x_1, \dots, x_n$. Обратно, r-тата примитивна функция на една на части постоянна функция със скокове в точките $x_1, \dots, x_n$ е сплайн от степен r с възли в $x_1, \dots, x_n$.

Отсечената степенна функция

$$(x-\xi)_+^r := \begin{cases} (x-\xi)^r, & \text{ако } x \geq \xi \\ 0, & \text{ако } x < \xi \end{cases}$$

е прост пример на сплайн-функция. Тя е сплайн от степен r с единствен възел в точката ξ. Наистина, $(x-\xi)_+^r$ съвпада с полинома $(x-\xi)^r$ при $x \geq \xi$ и с полинома $p(x) \equiv 0$ при $x < \xi$. Освен това $\left\{(x-\xi)_+^r\right\}^{(i)}$ е непрекъсната в точката $x = \xi$ за $i = 0, \dots, r-1$ (и приема там стойност 0). Отсечената степенна функция играе основна роля в теорията на сплайн-функциите. Както ще видим по-долу, всеки сплайн се представя като линейна комбинация на подходящи отсечени степенни функции.

Теорема 1. *Всяка сплайн-функция $s(x)$ от класа $S_r(x_1, \dots, x_n)$ се представя по единствен начин във вида*

$$s(x) = p(x) + \sum_{k=1}^{n} c_k (x - x_k)_+^r \,, \tag{1}$$

където p е полином от степен r, а $c_1, \dots, c_n$ са реални числа. Нещо повече,

$$c_k = \frac{s^{(r)}(x_k+0) - s^{(r)}(x_k-0)}{r!}, \quad k = 1, \dots, n. \tag{2}$$

Доказателство. Да поясним най-напред, че тук сме използвали означението

$$f(x+0) := \lim_{h \to 0, h>0} f(x+h).$$

Аналогично се определя $f(x-0)$.

Нека $s(x) \in S_r(x_1, \dots, x_n)$. Тогава s съвпада с някакъв полином P_k от степен r в подинтервала (x_k, x_{k+1}), $k = 0, \dots, n$. Тъй като $s^{(j)}(x)$ е непрекъсната функция в точката x_k, то $P_{k-1}^{(j)}(x_k) = P_k^{(j)}(x_k)$ за $j = 0, \dots, r-1$. Следователно

$$P_k(x) = P_{k-1}(x) + c_k (x - x_k)^r \quad \text{за всяко } x \,, \tag{3}$$

където c_k е някаква константа. Това е една рекурентна връзка за полиномите $\{P_k\}$, от която следва веднага представянето

$$P_k(x) = P_0(x) + \sum_{i=1}^{k} c_i (x - x_i)^r \,.$$

Като вземем предвид уговорката, че $s(x)$ съвпада с $P_k(x)$ при $x \in (x_k, x_{k+1})$ и определението на отсечената степенна функция, от горното равенство получаваме

$$s(x) = P_0(x) + \sum_{i=1}^{n} c_i (x - x_i)_+^r ,$$

което е търсеното представяне.

Остава да покажем, че коефициентите c_k се определят еднозначно. Наистина, от връзката (3) чрез диференциране r пъти в точката x_k получаваме

$$P_k^{(r)}(x_k) = P_{k-1}^{(r)}(x_k) + c_k\, r!$$

и оттук

$$c_k\, r! = s^{(r)}(x_k + 0) - s^{(r)}(x_k - 0),$$

което съвпада с формулата (2).

Полиномът $p(x)$ в представянето (1) се определя също еднозначно – той съвпада с полинома $P_0(x)$. Теоремата е доказана.

Тъй като всеки израз от вида (1) е сплайн от класа $S_r(x_1, \dots, x_n)$, то от теоремата следва, че $S_r(x_1, \dots, x_n)$ съвпада с множеството от всички функции от вида (1). Следователно размерността на линейното пространство $S_r(x_1, \dots, x_n)$ е равна на $r + n + 1$ (което е броят на отсечените степенни функции и алгебричните функции $1, x, \dots x^r$ участващи в представянето (1)).

Сега ще се спрем на задачата за интерполиране със сплайн-функции. Ще разгледаме интерполационната задача на Лагранж за сплайните от трета степен, наричани още *кубични сплайни*. Те се използват често на практика.

Нека $f(x)$ е реална непрекъсната функция в $[a, b]$. Искаме да построим кубичен сплайн $s(x)$ с възли в $x_1, \dots, x_n$, който интерполира $f(x)$ в точките $x_0, \dots, x_{n+1}$, където $a = x_0 < x_1 < \dots < x_{n+1} = b$. Да построим s значи да намерим полиномите от трета степен $\{P_i(x)\}$, които представят $s(x)$ в интервалите (x_i, x_{i+1}), $i = 0, \dots, n$, съответно. От интерполационните условия

$$s(x_i) = f(x_i), \quad i = 0, \dots, n + 1,$$

следват условия за полинома P_i:

$$(4) \qquad P_i(x_i) = f(x_i), \quad P_i(x_{i+1}) = f(x_{i+1}), \quad i = 0, \dots, n.$$

Да отбележим, че от горните равенства веднага следва

$$P_{i-1}(x_i) = P_i(x_i), \quad i = 1, \dots, n,$$

което гарантира, че $s(x)$ е непрекъсната функция в $[a, b]$. Да припомним, че всеки кубичен полином се определя напълно с четири интерполационни условия. За сега всяка кубична част $P_i(x)$ интерполира $f(x)$ само в две точки – x_i и x_{i+1}. Следователно имаме на разположение още две интерполационни условия. По-долу ще подберем тези условия така, че s да бъде не само непрекъсната, но да има непрекъснати първа и втора производна, т.е. да бъде кубичен сплайн. Различните методи за интерполиране с на части кубични функции се различават само по това как сме избрали тези две допълнителни интерполационни условия. За сега да поискаме $P_i(x)$ да удовлетворява условията

$$(5) \qquad P_i'(x_i) = d_i, \quad P_i'(x_{i+1}) = d_{i+1}, \quad i = 0, \dots, n ,$$

където $d_0, \dots, d_{n+1}$ са параметри, чиито избор ще уточним по-късно. Горните условия гарантират, че $s'(x)$ е непрекъсната функция в $[a, b]$. За да определим $P_i(x)$ от интерполационните условия на Ермит (4) и (5), ще използваме формулата на Нютон

$$\begin{aligned} P_i(x) &= P_i(x_i) + P_i[x_i, x_i](x - x_i) + P_i[x_i, x_i, x_{i+1}](x - x_i)^2 \\ &+ P_i[x_i, x_i, x_{i+1}, x_{i+1}](x - x_i)^2(x - x_{i+1}). \end{aligned}$$

Определяме коефициентите от таблицата за пресмятане на разделените разлики. Да въведем означението $\Delta_i := x_{i+1} - x_i$. Имаме

$$\begin{aligned} P_i(x_i) &= f(x_i) , \\ P_i[x_i, x_i] &= d_i , \\ P_i[x_i, x_{i+1}] &= f[x_i, x_{i+1}] , \\ P_i[x_i, x_i, x_{i+1}] &= \frac{f[x_i, x_{i+1}] - d_i}{\Delta_i} , \\ P_i[x_i, x_i, x_{i+1}, x_{i+1}] &= \frac{d_{i+1} - 2f[x_i, x_{i+1}] + d_i}{(\Delta_i)^2} . \end{aligned}$$

Да отбележим, че всички разделени разлики на P_i в четири точки са еднакви и равни на водещия коефициент на $P_i(x)$.

Избирайки по различен начин параметрите $\{d_i\}_0^{n+1}$ ние получаваме различни интерполиращи функции. Да отбележим един специален частен случай.

На части кубична ермитова интерполация. Избираме

$$d_i = f'(x_i), \quad i = 0, \dots, n + 1.$$

В този случай $P_i(x)$ зависи само от поведението на $f(x)$ в $[x_i, x_{i+1}]$. Ако с s_0 означим получената при горните условиа крива, то за $x \in [x_i, x_{i+1}]$ получаваме

$$|f(x) - s_0(x)| = |(x - x_i)^2(x - x_{i+1})^2| . |f[x_i, x_i, x_{i+1}, x_{i+1}, x]|$$

$$\leq \left(\frac{\Delta_i}{2}\right)^4 \max_{\xi\in[x_i,x_{i+1}]} \frac{\left|f^{(4)}(\xi)\right|}{4!}$$

при предположение, че f има непрекъсната четвърта производна в $[a,b]$. Следователно за всяко x от $[a,b]$ имаме

$$|f(x)-s_0(x)| \leq \max_{\xi\in[a,b]} \left|f^{(4)}(\xi)\right| \frac{\left(\max\limits_{0\leq i\leq n} \Delta_i\right)^4}{384}.$$

Кубична сплайнова интерполация. Както вече отбелязахме, функцията $s(x)$, определена от условията (4) и (5), е не само непрекъсната, но и с непрекъсната първа производна при всеки избор на параметрите d_i. Сега ще покажем, че винаги е възможно да се изберат параметрите $\{d_i\}$ така, че функцията $s(x)$ да има и непрекъсната втора производна, т.е. да бъде кубична сплайн-функция.

Нашето изискване $s''(x)$ да е непрекъсната е еквивалентно с условията

$$P''_{i-1}(x_i) = P''_i(x_i), \quad i=1,\dots,n. \tag{6}$$

Като използваме изведеното представяне на P_{i-1} и P_i по формулата на Нютон, получаваме съответно:

$$P''_i(x_i) = 2P_i[x_i,x_i,x_{i+1}] - 2P_i[x_i,x_i,x_{i+1},x_{i+1}]\Delta_i\ ,$$

$$P''_{i-1}(x_i) = 2P_{i-1}[x_{i-1},x_{i-1},x_i] + 4P_{i-1}[x_{i-1},x_{i-1},x_i,x_i]\Delta_{i-1}.$$

Заместваме разделените разлики с получените по-горе изрази и (6) добива вида

$$\frac{f[x_{i-1},x_i]-d_{i-1}}{\Delta_{i-1}} + 2\frac{d_i-2f[x_{i-1},x_i]+d_{i-1}}{\Delta_{i-1}}$$
$$= \frac{f[x_i,x_{i+1}]-d_i}{\Delta_i} - \frac{d_{i+1}-2f[x_i,x_{i+1}]+d_i}{\Delta_i}\ .$$

Привеждаме под общ знаменател и стигаме до равенството

$$\Delta_i d_{i-1} + 2(\Delta_{i-1}+\Delta_i)d_i + \Delta_{i-1}d_{i+1} = b_i, \quad i=1,\dots,n, \tag{7}$$

където

$$b_i = 3(f[x_{i-1},x_i]\Delta_i + f[x_i,x_{i+1}]\Delta_{i-1}), \quad i=1,\dots,n.$$

Да предположим за момент, че d_0 и d_{n+1} са избрани по някакъв начин. Тогава от (7) получаваме система от n уравнения с n неизвестни за определянето на $d_1, \ldots, d_n$. Тази система има доминиращ главен диагонал (което значи, че диагоналният елемент е по-голям по абсолютна стойност от сумата на абсолютните стойности на всички останали извъндиагонални елементи на реда). Лесно се показва, че всяка матрица с доминиращ главен диагонал има ненулева детерминанта. Следователно системата (7) има винаги единствено решение при всеки избор на d_0 и d_{n+1}.

Следните два начина за избиране на параметрите d_0 и d_{n+1} се използват най-често.

I) Ако $f'(a)$ и $f'(b)$ се знаят, то естествено е да се вземе

$$d_0 = f'(a), \quad d_{n+1} = f'(b).$$

По този начин се получава така наречената *пълна кубична сплайнова интерполация.*

II) Друг начин за избор на d_0 и d_{n+1} е да се добавят уравненията

$$\begin{aligned} s''(a) &= P_0''(x_0) = 0, \\ s''(b) &= P_n''(x_{n+1}) = 0, \end{aligned}$$

които заедно с (7) образуват една линейна система от $n+2$ уравнения с $n+2$ неизвестни за определянето на $d_0, \ldots, d_{n+1}$. Така получаваме *естествената кубична сплайнова интерполация.*

Кубични сплайни с възли $x_0, x_1, \ldots, x_{n+1}$, които съвпадат с полиноми от първа степен в интервалите $(-\infty, x_0)$ и (x_{n+1}, ∞) се наричат *естествени кубични сплайни.* Вижда се, че естествените кубични сплайни s удовлетворяват условията $s''(a) = s''(b) = 0$. Те се наричат *естествени* поради естественото им поведение – те описват достатъчно добре поведението на еластична пръчка, която е промушена през халки, намиращи се в точките $\{x_i, f(x_i)\}_0^{n+1}$. Ясно е, че такава пръчка ще заеме положение на права линия преди първата и след последната халка. Уредът за чертане "сплайн", за който вече споменахме, се състои от еластична пръчка и приспособления за прикрепването ѝ към произволни точки от чертожната дъска.

Ще докажем едно екстремално свойство на естествените сплайни. То показва, че те са най-гладките (в известен смисъл) функции, сред всички други, които интерполират дадена таблица.

Нека $\bar{x} = (x_0, \ldots, x_{n+1})$ са дадени точки, $a = x_0 < \cdots < x_{n+1} = b$, и $\bar{y} = (y_0, \ldots, y_{n+1})$ са дадени стойности. Да означим с $F(\bar{x}, \bar{y})$ класа от всички функции f с непрекъсната в $[a, b]$ втора производна, които удовлетворяват интерполационните условия:

$$f(x_k) = y_k, \quad \text{за } k = 0, \ldots, n+1.$$

Теорема на Холидей. *При дадени $\bar{x}$ и $\bar{y}$, нека $s(x)$ е единственият естествен кубичен сплайн с възли в $x_1, \dots, x_n$, който принадлежи на класа от интерполиращи функции $F(\bar{x}, \bar{y})$. Тогава*

$$\int_a^b s''^2(x)\,dx \le \int_a^b f''^2(x)\,dx \qquad \text{за всяко } f \in F(\bar{x}, \bar{y}) \; .$$

Равенството се достига само при $f \equiv s$.

Доказателство. Нека f е произволна функция от $F(\bar{x}, \bar{y})$ и $s(x)$ е произволен кубичен сплайн от $F(\bar{x}, \bar{y})$ с възли в $x_0, \dots, x_{n+1}$. Тогава интегралът

$$\sigma := \int_a^b \left[f''(x) - s''(x)\right] s''(x)\,dx$$

се пресмята много лесно. Наистина, интегрирайки по части получаваме

$$\begin{aligned} \sigma &= \sum_{i=1}^{n+1} \int_{x_{i-1}}^{x_i} \left[f''(x) - s''(x)\right] s''(x)\,dx \\ &= \sum_{i=1}^{n+1} s''(x)\left[f'(x) - s'(x)\right]\Big|_{x_{i-1}}^{x_i} - \sum_{i=1}^{n+1} \int_{x_{i-1}}^{x_i} \left[f'(x) - s'(x)\right] s'''(x)\,dx \; . \end{aligned}$$

Но s е кубичен сплайн. Следователно s съвпада с полином от трета степен в подинтервала (x_{i-1}, x_i) и тогава $s'''(x)$ е константа в (x_{i-1}, x_i). Да означим тази константа с c_i. Получаваме

$$\sigma = \sum_{i=1}^{n+1} s''(x)\left[f'(x) - s'(x)\right]\Big|_{x_{i-1}}^{x_i} - \sum_{i=1}^{n+1} c_i\left[f(x) - s(x)\right]\Big|_{x_{i-1}}^{x_i} \; .$$

Последната сума е нула, защото $f, s \in F(\bar{x}, \bar{y})$ и следователно

$$f(x_j) - s(x_j) = y_j - y_j = 0 \quad \text{за } j = 0, \dots, n+1.$$

Освен това функцията $s''(x)\left[f'(x) - s'(x)\right]$ е непрекъсната в точките $\{x_j\}_0^{n+1}$. Тогава при сумирането в първата сума всички изрази съдържащи стойностите на тази функция във вътрешните точки $x_1, \dots, x_n$ взаимно ще се съкратят, защото участват с противоположни знаци. Ще останат само стойностите в първата и последна точка. Окончателно получаваме

$$\sigma = \int_a^b \left[f''(x) - s''(x)\right] s''(x)\,dx$$

$$= s''(b)\left[f'(b) - s'(b)\right] - s''(a)\left[f'(a) - s'(a)\right] \; . \tag{8}$$

Да забележим сега, че ако s е естественият кубичен сплайн от $F(\bar{x}, \bar{y})$, то $s''(a) = s''(b) = 0$ и следователно $\sigma = 0$. С други думи, функциите $f''(x) -$

$s''(x)$ и $s''(x)$ са ортогонални в този случай. Но ако две функции f_1 и f_2 са ортогонални, то очевидно

$$\int_a^b f_1^2(x)\,dx \le \int_a^b [f_1(x) + f_2(x)]^2\,dx \ ,$$

като равенство се достига само при $f_2(x) \equiv 0$. Като приложим този факт за $f_1 = s''(x)$ и $f_2 = f''(x) - s''(x)$, получаваме

$$\int_a^b s''^2(x)\,dx \le \int_a^b [f''(x) - s''(x) + s''(x)]^2\,dx = \int_a^b f''^2(x)\,dx \ .$$

Равенството се достига само при $f_2 = f'' - s'' \equiv 0$. Тогава $f - s$ е полином от първа степен. Тъй като $f - s$ се анулира в $x_0, \dots, x_{n+1}$ по условие, то очевидно $f \equiv s$. И така, равенството се достига само при $f \equiv s$. Теремата е доказана.

Същата теорема остава в сила, ако F е класът от функции удовлетворяващи интерполационните условия

$$f(x_i) = y_i, \quad i = 0, \dots, n+1, \quad f'(a) = y_0', \ f'(b) = y_{n+1}',$$

а s е кубичният сплайн, осъществяващ пълната кубична сплайнова интерполация. Наистина, в този случай

$$\begin{aligned} f'(a) - s'(a) &= 0, \\ f'(b) - s'(b) &= 0 \end{aligned}$$

и от (8) отново следва, че функциите $f''(x) - s''(x)$ и $s''(x)$ са ортогонални.

10. B-СПЛАЙНИ

Вече показахме, че всеки сплайн от степен $r-1$ с възли $x_1 < \cdots < x_n$ може да бъде представен като линейна комбинация на полином p от π_{r-1} и отсечените степенни функции

$$(x-x_1)_+^{r-1}, \ldots, (x-x_n)_+^{r-1}.$$

Такова представяне на сплайна не е удобно при работа с компютър по следната причина. Ако n е много голямо, то стойността на сплайна $s(x)$ в точката $x \in (x_i, x_{i+1})$ се записва като сума на голям брой изрази (по-точно на $i+r$ израза), а по същество $s(x)$ е полином от степен $r-1$ в (x_i, x_{i+1}) и би трябвало да се запише като линейна комбинация на r линейно независими функции. При смятането с голям брой изрази се получават евентуални грешки при закръгляне на числата в компютъра, които се натрупват и могат да доведат до съществена неточност в крайния резултат. Сега ще въведем един друг базис в пространството от сплайни, който няма този недостатък.

Определение. *Разделената разлика* на отсечената степенна функция $(x-t)_+^{r-1}$ по отношение на x в точките $x_0 < \cdots < x_r$ се нарича B-сплайн от степен $r-1$ с възли $x_0, \ldots, x_r$.

Да въведем означението $B(x_0, \ldots, x_r; t)$ за този B-сплайн. Съгласно определението,

$$B(x_0, \ldots, x_r; t) = (\cdot - t)_+^{r-1}[x_0, \ldots, x_r].$$

Лесно се вижда, че изразът $B(x_0, \ldots, x_r; t)$ е наистина сплайн от степен $r-1$ с възли $x_0, \ldots, x_r$. За целта да си припомним, че за всяко f

$$f[x_0, \ldots, x_r] = c_0 f(x_0) + c_1 f(x_1) + \cdots + c_r f(x_r),$$

където $c_k = \frac{1}{\omega'(x_k)}$ и $\omega(x) := (x-x_0)\ldots(x-x_r)$. Следователно

$$B(x_0, \ldots, x_r; t) = \sum_{k=0}^{r} c_k (x_k - t)_+^{r-1},$$

което показва, че изразът $B(x_0, \ldots, x_r; t)$ е линейна комбинация на отсечените степенни функции $\{(x_k - t)_+^{r-1}\}_{k=0}^{r}$, т.е. $B(x_0, \ldots, x_r; t)$ е сплайн от степен $r-1$ с възли $x_0, \ldots, x_r$.

Ще докажем няколко интересни свойства на B-сплайните.

Теорема 1. *При всяко $r \geq 1$ имаме:*

а) $B(x_0, \ldots, x_r; t) = 0$ *за всяко* $t \leq x_0$ *и всяко* $t \geq x_r$,

б) $B(x_0, \ldots, x_r; t) > 0$ *при* $t \in (x_0, x_r)$.

Доказателство. а) Нека $t \le x_0$. Тогава $x_k - t \ge 0$ за всяко k. Следователно $(x_k - t)_+^{r-1} = (x_k - t)^{r-1}$ и $B(x_0, \ldots, x_r; t) = (\cdot - t)^{r-1}[x_0, \ldots, x_r]$. Но $(x - t)^{r-1}$ е полином от степен $r - 1$. Тогава неговата разделена разлика в произволни $r + 1$ точки е равна на нула. И така, $B(x_0, \ldots, x_r; t) = 0$ за $t \le x_0$. Да предположим сега, че $t \ge x_r$. Тогава $x_k - t \le 0$ за $k = 0, \ldots, r$ и следователно $(x_k - t)_+^{r-1} = 0$ за $k = 0, \ldots, r$. Оттук и от определението на B-сплайн получаваме

$$B(x_0, \ldots, x_r; t) = \sum_{k=0}^{r} c_k (x_k - t)_+^{r-1} = 0 \quad \text{при } t \ge x_r.$$

б) Нека t е фиксирана точка в (x_0, x_r). Да означим с $P_r(x) = bx^r + \cdots$ полинома от степен r, който интерполира функцията $Q_t(x) := (x - t)_+^{r-1}$ в точките $x_0, \ldots, x_r$. Ясно е, че $P_r(x)$ не може да съвпада тъждествено с $(x - t)^{r-1}$ или с нулата в някакъв подинтервал на $(-\infty, \infty)$. В противен случай полиномът $P_r(x)$ би съвпадал тъждествено с полинома $(x - t)^{r-1}$ или с нулата, което очевидно не е вярно.

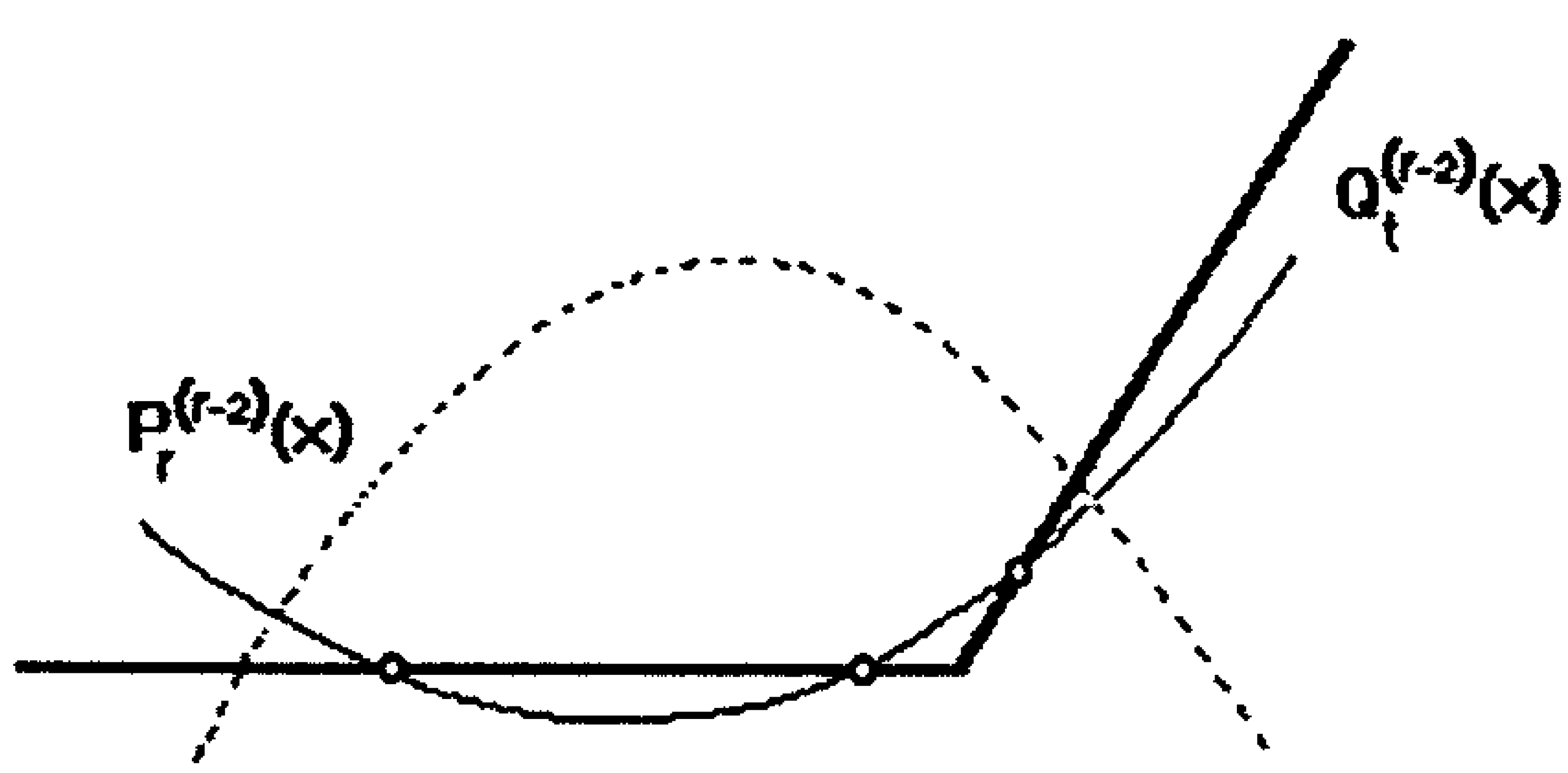

Рис. 3

Да разгледаме разликата $P_r(x)$–$Q_t(x)$. Тя има поне r+1 нули: $x_0, x_1, \ldots, x_r$. Съгласно направената по-горе бележка, тези нули са изолирани (т.е. в произволно малка околност на x_i има точки, в които функцията е различна от нула). Тогава по теоремата на Рол, между всеки две нули на $P_r(x) - Q_t(x)$ ще има поне една нула на $P_r'(x) - Q_t'(x)$ или по-точно, между всеки две нули на $P_r(x) - Q_t(x)$ ще има точка, в която производната $P_r'(x) - Q_t'(x)$ си сменя знака. И така, $P_r'(x) - Q_t'(x)$ ще има поне r различни нули (смени на знака). Продължавайки по същия начин заключаваме, че $P_r''(x) - Q_t''(x)$ ще има поне $r - 1$ нули и т.н., $P_r^{(r-2)}(x) - Q_t^{(r-2)}(x)$ ще има поне 3 нули (смени на знака). Но

$$P_r^{(r-2)}(x) = \frac{r!}{2}\, bx^2 + \cdots,$$

т.е. $P_r^{(r-2)}(x)$ е парабола, а $Q_t^{(r-2)}(x)$ е монотонно растяща начупена линия (вж. Рис. 3). Вижда се, че параболата $P_r^{(r-2)}(x)$ не би могла да пресича $Q_t^{(r-2)}(x)$ в повече от две точки, ако тя е "обърната надолу", т.е. ако водещият ѝ коефициент $r!\,b/2$ е отрицателен. Ако този коефициент е нула, то $P_r^{(r-2)}(x)$ е линейна функция и също не би могла да пресича $Q_t^{(r-2)}(x)$ в повече от две точки. Следователно $r!\,b/2 > 0$ и оттук $b > 0$.
Но b е коефициентът пред x^r в интерполационния полином $P_r(x)$. Съгласно едно от свойствата на разделената разлика b съвпада с разделената разлика на интерполираната функция $(x-t)_+^{r-1}$ в $x_0, \ldots, x_r$. С други думи

$$b = B(x_0, \ldots, x_r; t) > 0,$$

което трябваше да докажем.

Нека

$$\cdots < x_{i-1} < x_i < x_{i+1} < \cdots$$

е дадена крайна или безкрайна редица от различни точки. С тази редица ще свързваме съответстващата ѝ редица от B-сплайни

$$B_{i,r-1}(t) := B(x_i, \ldots, x_{i+r}; t) \text{ за } \quad \forall i.$$

Ще покажем, че функциите $B_{i,r-1}(t)$, $\;i = m, m+1, \ldots, m+N$, са линейно независими в $(-\infty, \infty)$ при всеки избор на m и N. Наистина, да допуснем противното. Тогава съществува линейна комбинация

$$f(t) = \sum_{i=m}^{m+N} \alpha_i B_{i,r-1}(t),$$

която се анулира за всяко t от $(-\infty, \infty)$, но поне един от коефициентите $\{\alpha_i\}$ е различен от нула. Да изберем t от интервала (x_m, x_{m+1}). При това t ще имаме

$$f(t) = \alpha_m B_{m,r-1}(t),$$

защото $B_{i,r-1}(t) = 0$ за $i > m$. Тъй като по Теорема 1 $B_{m,r-1}(t) > 0$, то от условието $f(t) = 0$ следва, че $\alpha_m = 0$. По същия начин сега показваме, че $\alpha_{m+1} = 0$ и т.н., докато стигнем до заключението, че всички коефициенти $\{\alpha_i\}_{i=m}^{m+N}$ са равни на нула, което противоречи на направеното предположение. Твърдението е доказано.

Сега ще построим един нов базис за пространството от сплайни, като използваме B-сплайните. За целта ще ни бъде необходима следната лема.

Лема 2. *При всеки избор на точките* $\xi_1 < \cdots < \xi_r$ *функциите* $(\xi_1 - x)^{r-1}, \ldots, (\xi_r - x)^{r-1}$ *са линейно независими в* $(-\infty, \infty)$.

Доказателство. Допускаме противното. Тогава съществува линейна комбинация

$$f(x) = \sum_{i=1}^{r} a_i(\xi_i - x)^{r-1},$$

която се анулира за всяко $x \in (-\infty, \infty)$, но поне едно a_i е различно от нула. Тъй като $f(x)$ е алгебричен полином на x, който се анулира тъждествено, то и неговите производни ще се анулират тъждествено, т.е.

$$f(x) = f'(x) = \cdots = f^{r-1}(x) = 0 \quad \text{за} \quad x \in (-\infty, \infty)$$

Да фиксираме някакво x в $(-\infty, \xi_1)$ и да означим $y_i = \xi_i - x, \quad i = 1, \ldots, r$. Тогава

$$f(x) = 0 \quad \Rightarrow \quad a_1 y_1^{r-1} + \cdots + a_r y_r^{r-1} = 0$$

$$f'(x) = 0 \quad \Rightarrow \quad a_1 y_1^{r-2} + \cdots + a_r y_r^{r-2} = 0$$

....

$$f^{(r-1)}(x) = 0 \quad \Rightarrow \quad a_1.1 + \cdots + a_r.1 = 0$$

Получихме, че $a_1, \ldots, a_r$ удовлетворяват една линейна система. Но детерминантата на тази система е Вандермондова, следователно различна от нула. Тогава системата допуска само нулево решение $a_1 = \cdots = a_r = 0$. Стигнахме до противоречие. Лемата е доказана.

От Теорема 1 следва, че всеки B-сплайн $B_{i,r-1}(t)$ е различен от нула само в крайния интервал (x_i, x_{i+r}). Този интервал се нарича носител на $B_{i,r-1}(t)$. И така, B-сплайните са сплайн-функции с краен носител. Сега ще покажем, че няма други сплайни от степен $r-1$, които имат "по-малък" носител от B-сплайните.

Лема 3. *Нека* $x_1 < \cdots < x_r$ *и* $f \in S_{r-1}(x_1, \ldots, x_r)$. *Ако* $f(t) = 0$ *за всяко* $t \notin [x_1, x_r]$, *то* $f(t) \equiv 0$ *в* $(-\infty, \infty)$.

Доказателство. Сплайнът f може да бъде представен във вида

$$f(t) = p(t) + \sum_{k=1}^{r} c_k(t - x_k)_+^{r-1},$$

където p е полином от π_{r-1}. Нека $t \in (-\infty, x_1)$. Тогава $f(t) = p(t) = 0$. Следователно $p \equiv 0$. И така

$$f(t) = \sum_{k=1}^{r} c_k(t - x_k)_+^{r-1}$$

за всяко t. Да изберем сега произволни точки $\xi_1 < \cdots < \xi_r$ от интервала (x_r, ∞) и да въведем означенията

$$p_j(x) := (\xi_j - x)^{r-1}, \quad j = 1, \ldots, r.$$

От Лема 2 следва, че $p_1, \ldots, p_r$ са линейно независими полиноми. Тъй като техният брой е r, то те образуват базис в π_{r-1}. Да разгледаме функционала

$$L(p) := \sum_{k=1}^{r} c_k p(x_k).$$

От условието $f(\xi_j) = 0$ следва $L(p_j) = 0, \quad j = 1, \ldots, r$. Тъй като $\{p_j\}_1^r$ образуват базис в π_{r-1}, то

$$L(q) = 0 \quad \text{за всяко} \quad q \in \pi_{r-1}.$$

Нека q_j е полиномът от π_{r-1}, който удовлетворява интерполационните условия

$$q_j(x_k) = \delta_{kj}, \quad k = 1, \ldots, r.$$

При $q = q_j$ получаваме

$$0 = L(q_j) = \sum_{k=1}^{r} c_k q_j(x_k) = c_j$$

за $j = 1, \ldots, r$. Следователно $f(t) \equiv 0$. Твърдението е доказано.

И така, нека подчертаем още веднъж: B-сплайните от степен $r-1$ имат минимален носител в пространството от сплайни от степен $r-1$.

Вече сме готови да докажем основния резултат в тази лекция.

Теорема 4. *Нека $a < x_{r+1} < \cdots < x_n < b$ са фиксирани точки. Да изберем произволни други $2r$ точки $x_1 < \cdots < x_r < a$ и $b < x_{n+1} < \cdots < x_{n+r}$. Нека $B_i(t) := B(x_i, \ldots, x_{i+r}; t), \quad i = 1, \ldots, n$. B-сплайните $B_1, \ldots, B_n$ образуват базис в пространството $S_{r-1}(x_{r+1}, \ldots, x_n)$ върху интервала $[a, b]$.*

Доказателство. Вече знаем, че размерността на пространството $S_{r-1}(x_{r+1}, \ldots, x_n)$ е равна на n. Тъй като $B_i \in S_{r-1}(x_{r+1}, \ldots, x_n)$ за $i = 1, \ldots, n$ (и техният брой е също n), то остава да покажем само, че $B_1, \ldots, B_n$ са линейно независими функции в $[a, b]$.

Допускаме противното. Тогава съществува линейна комбинация

$$f(t) = \sum_{i=1}^{n} \alpha_i B_i(t),$$

която се анулира тъждествено в $[a, b]$, но поне един от коефициентите $\{\alpha_i\}$ е различен от нула. От израза за f се вижда, че

$$f(t) \equiv 0 \quad \text{в} \quad (-\infty, x_1),$$

$$f(t) \equiv 0 \quad \text{в} \quad (x_{r+n}, \infty).$$

Освен това, по допускане $f(t) \equiv 0$ в $[a, b]$. Но f съвпада с алгебричен полином в (x_r, x_{r+1}). От това, че $f(t) \equiv 0$ в $[a, x_{r+1}]$ следва, че $f(t) \equiv 0$ в целия подинтервал $[x_r, x_{r+1}]$. По същия начин се вижда, че $f(t) \equiv 0$ и в $[x_n, x_{n+1}]$. Следователно $f \equiv 0$ в $[x_r, x_{n+1}]$ и тогава тя има вида, показан на Рис. 4.

Рис. 4

Да разгледаме функциите

$$f_1(t) = \begin{cases} 0 & \text{при} \quad t > x_r \\ f(t) & \text{при} \quad t \le x_r \end{cases},$$

$$f_2(t) = \begin{cases} 0 & \text{при} \quad t < x_{n+1} \\ f(t) & \text{при} \quad t \ge x_{n+1} \end{cases}.$$

Очевидно $f(t) = f_1(t) + f_2(t)$. Но $f_1 \in S_{r-1}(x_1, \dots, x_r)$ и $f_1(t) \equiv 0$ за $t \notin [x_1, x_r]$. Тогава от Лема 3 (за минималния носител) следва, че $f_1(t) \equiv 0$ в $(-\infty, \infty)$ и оттук $f(t) \equiv 0$ в $(-\infty, a]$. По същия начин се вижда, че $f_2 \equiv 0$ и оттук, $f(t) \equiv 0$ в $[b, \infty)$. Следователно $f(t) \equiv 0$ в $(-\infty, \infty)$. Но, както отбелязахме в началото, функциите $B_1, \dots, B_n$ са линейно независими в $(-\infty, \infty)$. Тогава $\alpha_1 = \cdots = \alpha_n = 0$. Стигаме до противоречие. Теоремата е доказана.

И така, всяка сплайн-функция f от $S_{r-1}(x_{r+1}, \dots, x_n)$ може да бъде представена по единствен начин във вида

$$f(t) = \sum_{i=1}^{n} \alpha_i B_i(t). \tag{3}$$

Имайки предвид крайния носител на $B_i(t)$, това е много удобно представяне на f за работа с компютър, тъй като при фиксирано t, слайнът $f(t)$ е всъщност линейна комбинация само на r последователни B-сплайни, които съдържат t в своя носител. Едно друго предимство на представянето (3) е, че съществува проста схема за пресмятане стойността на B_i в дадена точка. Тази схема се основава на следната рекурентна връзка.

Основна рекурентна връзка: *За всяко $r \geq 2$ и $t \in (-\infty, \infty)$ е в сила равенството*

$$B_{i,r-1}(t) = \frac{t - x_i}{x_{i+r} - x_i} B_{i,r-2}(t) + \frac{x_{i+r} - t}{x_{i+r} - x_i} B_{i+1,r-2}(t). \tag{4}$$

Доказателство. Ще използваме известното правило на Стефенсен за намиране на разделена разлика на произведение от функции

$$(f \cdot g)[x_0, \ldots, x_n] = \sum_{k=0}^{n} f[x_0, \ldots, x_k]\, g[x_k, \ldots, x_n].$$

Да изберем $f(x) = x - t$ и $g(x) = (x - t)_+^{r-2}$. Очевидно

$$f(x)g(x) = (x - t)_+^{r-1} \quad \text{за } x \in (-\infty, \infty)$$

и следователно

$$\begin{aligned} B_{i,r-1}(t) &= (f.g)[x_i, \ldots, x_{i+r}] \\ &= f(x_i)g[x_i, \ldots, x_{i+r}] + f[x_i, x_{i+1}]g[x_{i+1}, \ldots, x_{i+r}], \end{aligned}$$

тъй като $f[x_i, \ldots, x_{i+k}] = 0$ за $k \geq 2$. По-нататък, като вземем предвид, че $f[x_i, x_{i+1}] = 1$ и приложим рекурентната връзка за разделените разлики получаваме

$$\begin{aligned} B_{i,r-1}(t) &= f(x_i)\frac{g[x_{i+1}, \ldots, x_{i+r}] - g[x_i, \ldots, x_{i+r-1}]}{x_{i+r} - x_i} + g[x_{i+1}, \ldots, x_{i+r}] \\ &= \left(1 + \frac{f(x_i)}{x_{i+r} - x_i}\right) g[x_{i+1}, \ldots, x_{i+r}] - \frac{f(x_i)}{x_{i+r} - x_i} g[x_i, \ldots, x_{i+r-1}] \\ &= \frac{x_{i+r} - t}{x_{i+r} - x_i} B_{i+1,r-2}(t) + \frac{t - x_i}{x_{i+r} - x_i} B_{i,r-2}(t), \end{aligned}$$

което е исканото равенство.

Да отбележим, че коефициентите пред $B_{i,r-2}(t)$ и $B_{i+1,r-2}(t)$ в горната рекурентна връзка са положителни при $t \in (x_i, x_{i+1})$ и тяхната сума е равна на 1. Следователно формулата (4) представя $B_{i,r-1}(t)$ като изпъкнала комбинация на $B_{i,r-2}(t)$ и $B_{i+1,r-2}(t)$.

Формулата (4) е в основата на следната схема за пресмятане на стойностите на B-сплайните.

$$
\begin{array}{ccccccc}
B_{00}(t) & & & & & & \\
 & \searrow & & & & & \\
 & & B_{01}(t) & & & & \\
 & \nearrow & & \searrow & & & \\
B_{10}(t) & & & & B_{02}(t) & & \\
 & \searrow & & \nearrow & & \searrow & \\
 & & B_{11}(t) & & & & B_{03}(t) \\
 & \nearrow & & \searrow & & \nearrow & \\
B_{20}(t) & & & & B_{12}(t) & & \\
 & \searrow & & \nearrow & & \searrow & \\
 & & B_{21}(t) & & & & B_{13}(t) \\
 & \nearrow & & \searrow & & \nearrow & \\
B_{30}(t) & & & & B_{22}(t) & & \\
 & \searrow & & \nearrow & & & \\
 & & B_{31}(t) & & & & \\
 & \nearrow & & & & & \\
B_{40}(t) & & & & & &
\end{array}
$$

Първата колона на тази таблица се попълва, като се използва определението на $B_{i,0}(t)$,

$$
B_{i,0}(t) = \begin{cases} \frac{1}{x_{i+1}-x_i} & \text{за} \quad t \in [x_i, x_{i+1}) \\ 0 & \text{за} \quad t \notin [x_i, x_{i+1}) \end{cases} .
$$

Следващите колони се попълват една след друга като се използват данните от предишната колона и рекурентната връзка (4).

11. НАЙ-ДОБРИ ПРИБЛИЖЕНИЯ В ЛИНЕЙНИ НОРМИРАНИ ПРОСТРАНСТВА

Нека F е дадено линейно пространство. Да въведем в F *разстояние*, т.е. на всеки два елемента f, g от F съпоставяме число $\rho(f, g)$, което удовлетворява следните изисквания:

1) $\rho(f,g) \geq 0$, като равенство се достига тогава и само тогава когато $f = g$;

2) $\rho(f,g) = \rho(g,f)$ (симетричност);

3) $\rho(f,g) \leq \rho(f,h) + \rho(h,g)$ за $\forall f, g, h \in F$.

Линейно пространство, в което е въведено разстояние (метрика) се нарича *линейно метрично пространство*. Ще формулираме задачата за приближаване в линейното метрично пространство F.

Нека $\varphi_0, \ldots, \varphi_n$ са произволни линейно независими елементи от F. Да означим с Ω_n множеството от всички линейни комбинации на $\{\varphi_k\}_0^n$, т.е.

$$\Omega_n := \left\{ \sum_{k=0}^{n} a_k \varphi_k \ : \ (a_0, \ldots, a_n) \in \mathbb{R}^{n+1} \right\}.$$

Величината

$$E_n(f) := \inf\{\rho(f, \varphi) \ : \ \varphi \in \Omega_n\}$$

се нарича *най-добро приближение на* f с елементи от Ω_n. Ако съществува елемент φ_f от Ω_n, за който горното равенство се достига, т.е. за който

$$\rho(f, \varphi_f) = \inf\{\rho(f, \varphi) \ : \ \varphi \in \Omega_n\},$$

този елемент φ_f се нарича *елемент на най-добро приближение* за f.

След тази формулировка на задачата за приближаване възникват следните основни въпроси:

Съществува ли елемент на най-добро приближение?

Ако такъв елемент съществува, то той единствен ли е?

Как може да бъде построен елементът на най-добро приближение?

Съществува достатъчно широк клас от линейни метрични пространства, за които може да бъде даден отговор на въпроса за съществуване на елемент на най-добро приближение. Това са така наречените *линейни нормирани пространства*. Да припомним накратко определението на нормирано пространство.

Нека F е дадено линейно пространство. Казваме, че в F е въведена норма, ако на всеки елемент f от F е съпоставено число $\|f\|$ (наречено *норма* на f) и това съответствие удовлетворява условията:

1) $\|f\| \geq 0$ (равенството се достига тогава и само тогава когато $f = 0$);

2) $\|\lambda f\| = |\lambda|\|f\|$ за $\forall\lambda$;
3) $\|f+g\| \le \|f\| + \|g\|$ за $\forall f, g \in F$.

Линейното пространство F, в което е въведена норма се нарича линейно нормирано пространство.

Всяка норма $\|\cdot\|$ поражда разстояние по следния начин:

$$\rho(f,g) := \|f-g\|.$$

Не е трудно да се провери, че така определеното разстояние $\rho(f,g)$ действително удовлетворява изброените по-горе свойства. Ще оставим тази проверка за упражнение.

Всяка норма в F може да се разглежда като функция на f, определена в F.

Теорема 1. *Нормата е непрекъсната функция по отношение на разстоянието, породено от нея.*

Доказателство. Най-напред да докажем неравенството

$$|\,\|f\| - \|g\|\,| \le \|f-g\|.$$

Наистина,

$$\|f\| = \|f-g+g\| \le \|f-g\| + \|g\|$$

и оттук следва $\|f\|-\|g\| \le \|f-g\|$. Аналогично $\|g\|-\|f\| \le \|g-f\| = \|f-g\|$. Следователно $|\,\|f\| - \|g\|\,| \le \rho(f,g)$. Оттук се вижда, че ако $\rho(f,g) \to 0$, то $\|f\| \to \|g\|$, което показва, че $\|f\|$ е непрекъсната функция на f.

Да разгледаме линейното пространство

$$\mathbb{R}^n = \{f = (f_1,\dots,f_n) \;:\; f_1,\dots,f_n - \text{ реални числа }\}$$

от n-мерни реални вектори. Тук всяка норма е всъщност функция на n променливи – координатите $f_1,\dots,f_n$ на f.

Теорема 2. *Всяка норма в $\mathbb{R}^n$ е непрекъсната функция относно координатите на елемента.*

Доказателство. Наистина, да означим с

$$\mathbf{e}_k = (\underbrace{0,\dots,1}_{k},0,\dots,0), \qquad k = 1,\dots,n,$$

базисните вектори в $\mathbb{R}^n$. Тогава всеки вектор $f = (f_1,\dots,f_n)$ от $\mathbb{R}^n$ се записва във вида $f = f_1\mathbf{e}_1 + \dots + f_n\mathbf{e}_n$ и следователно

$$|\,\|f\| - \|g\|\,| \le \|f-g\| = \|\sum_{i=1}^{n}(f_i-g_i)\mathbf{e}_i\| \le \sum_{i=1}^{n} |\,f_i - g_i\,|\,\|\mathbf{e}_i\|.$$

Оттук се вижда, че $\|f\| \to \|g\|$ при $f_i \to g_i, \quad i = 1, \ldots, n$. Теоремата е доказана.

В едно линейно пространство F може да бъде въведена норма по различни начини. Например, в $\mathbb{R}^n$ се използват често нормите:

$$\|f\|_\infty := \max_{1\le i\le n} |\, f_i \,|,$$

$$\|f\|_1 := |\, f_1 \,| + \cdots + |\, f_n \,|,$$

$$\|f\|_2 := \left(\sum_{i=1}^{n} f_i^2\right)^{1/2} \text{ – евклидова норма.}$$

Последната се нарича *евклидова*, защото поражда евклидовото разстояние

$$d(f,g) := \|f-g\|_2 = \left\{\sum_{k=1}^{n}(f_k - g_k)^2\right\}^{1/2}.$$

Определение: Казваме, че две норми $\nu(f)$ и $\mu(f)$ са *еквивалентни* в F, ако съществуват положителни числа m и M такива, че

$$m\mu(f) \le \nu(f) \le M\mu(f)$$

за всяко $f \in F$.

Теорема 3. *Всеки две норми в $\mathbb{R}^n$ са еквивалентни.*

Доказателство. Достатъчно е да докажем, че всяка норма ν е еквивалентна с евклидовата норма $\|\cdot\|_2$. За целта да въведем единичната сфера

$$S := \{(f_1, \ldots, f_n) \;:\; \sum_{i=1}^{n} f_i^2 = 1\}.$$

S е ограничено и затворено множество. Освен това, съгласно Теорема 2, $\nu(f) = \nu(f_1, \ldots, f_n)$ е непрекъсната функция на $f_i, \quad -\infty < f_i < \infty$. По теоремата на Вайерщрас, $\nu(f)$ достига своята минимална стойност в S. Следователно съществува елемент f^* от S такъв, че

$$m := \inf\{\nu(f) \;:\; (f_1, \ldots, f_n) \in S\} = \nu(f^*).$$

Очевидно $m \ge 0$. Нещо повече, $m > 0$. Наистина, допускането $m = 0$ води до $\nu(f^*) = 0$ и следователно $f^* = 0$, т.е. $f_1^* = \cdots = f_n^* = 0$, противоречие със заключението, че $f^* \in S$.

И така, $\nu(f) \ge m > 0$ за всяко $f \in S$.

Нека f е произволен ненулев елемент на F. Тогава $f/\|f\|_2 \in S$ и съгласно току-що доказаното неравенство ще имаме

$$\nu(f) = \nu\left(\frac{f}{\|f\|_2}\|f\|_2\right) = \|f\|_2\nu\left(\frac{f}{\|f\|_2}\right) \geq m\|f\|_2 .$$

Доказахме, че $m\|f\|_2 \leq \nu(f)$ за $\forall f \in F$. Аналогично, като изберем

$$M := \sup\{\nu(f) : (f_1, \ldots, f_n) \in S\},$$

получаваме

$$\nu(f) \leq M\|f\|_2 \quad \text{за} \quad \forall f \in F.$$

Теоремата е доказана.

Сега ще формулираме едно важно следствие от теоремата за еквивалентност на нормите.

Следствие 4. *Всяко кълбо $D_r = \{(f_1, \ldots, f_n) : \|f\| \leq r < \infty\}$ в $\mathbb{R}^n$ е ограничено и затворено множество.*

Доказателство. Нека f е елемент от кълбото D_r. Тогава $\|f\| \leq r$ и следователно съществува константа M такава, че

$$\|f\|_\infty \leq Mr.$$

Оттук следва неравенството $|f_i| \leq Mr$, което показва, че множеството D_r е ограничено. Сега ще покажем затвореността . Нека $\{f^{(k)}\}$ е произволна редица от елементи $f^{(k)} \in D_r$, която клони към някакъв елемент g от $\mathbb{R}^n$. Имаме

$$\|g\| \leq \|g - f^{(k)} + f^{(k)}\| \leq \|f^{(k)} - g\| + \|f^{(k)}\| \leq \|f^{(k)} - g\| + r.$$

Като извършим граничен преход при $k \to \infty$ получаваме $\|g\| \leq r$, което означава, че $g \in S_r$. Следователно D_r е затворено множество.

Твърденията, доказани тук за $\mathbb{R}^n$, са верни за всяко крайномерно линейно пространство $\Omega_n = \{\varphi = a_1\varphi_1 + \cdots + a_n\varphi_n : (a_1, \ldots, a_n) \in \mathbb{R}^n\}$ тъй като всеки елемент φ от Ω_n може да се отъждестви с вектора $(a_1, \ldots, a_n)$ от своите коефициенти.

Теорема 5. *Нека F е линейно нормирано пространство. Нека $\varphi_1, \ldots, \varphi_n$ са линейно независими елементи на F и Ω_n е линейното подпространство, породено от тях. Тогава, за всяко $f \in F$ съществува елемент на най-добро приближение от Ω_n по отношение на разстоянието, породено от нормата на F.*

Доказателство. Нека $\varphi \in \Omega_n$ и $\|\varphi\| > 2\|f\| =: r$. Тогава

$$\|f - \varphi\| \geq | \, \|\varphi\| - \|f\| \, | > 2\|f\| - \|f\| = \|f\| = \|f - 0\| \geq E_n(f).$$

Следователно

$$\inf_{\varphi\in\Omega_n} \|f-\varphi\| = \inf\{\|f-\varphi\| : \varphi\in\Omega_n, \|\varphi\|\le r\} = \inf_{(a_1,\dots,a_n)\in D_r} \|f-\sum_{k=1}^{n} a_k\varphi_k\|.$$

Но $\|f-\varphi\|$ е непрекъсната функция на коефициентите $a_1,\dots,a_n$ на φ, а D_r е ограничено и затворено множество. По теоремата на Вайерщрас

$$\inf_{\varphi\in S_r} \|f-\varphi\| = \min_{\varphi\in S_r} \|f-\varphi\| = \|f-\varphi_f\|$$

за някое $\varphi_f \in \Omega_n$. Теоремата е доказана.

Следващата теорема се отнася до единствеността на елемента на най-добро приближение.

Определение: Казваме, че нормираното пространство F е *строго нормирано*, ако от равенството

$$\|f+g\| = \|f\| + \|g\|$$

следва, че елементите f и g са линейно зависими.

Теорема 6. *Ако F е строго нормирано пространство, то за всяко $f\in F$ съществува единствен елемент на най-добро приближение от Ω_n.*

Доказателство. Допускаме противното. Тогава има $f\in F$ и елементи p и q от Ω_n, за които

$$\|f-p\| = \|f-q\| = E_n(f) := \inf\{\|f-\varphi\| : \varphi\in\Omega_n\}$$

и $p\neq q$. Но

$$(1)\qquad \|f-\frac{p+q}{2}\| = \frac{1}{2}\|(f-p)+(f-q)\| \le \frac{1}{2}(\|f-p\|+\|f-q\|) = E_n(f).$$

От друга страна, $\|f-(p+q)/2\| \ge E_n(f)$ по определение. Следователно в (1) имаме само равенства. В частност,

$$\|(f-p)+(f-q)\| = \|f-p\| + \|f-q\|.$$

Тогава от строгата нормираност на F следва, че $f-p=\alpha(f-q)$. Ако $\alpha=1$ това равенство води до $p=q$, противоречие. Ако $\alpha\neq 1$, то $f=(p-\alpha q)/(1-\alpha)$ и следователно $f\in\Omega_n$, което от своя страна влече $f=p=q$ и стигаме отново до противоречие с предположението, че $p\neq q$. Теоремата е доказана.

12. РАВНОМЕРНО ПРИБЛИЖЕНИЕ НА ФУНКЦИИ С АЛГЕБРИЧНИ ПОЛИНОМИ

Нека $[a, b]$ е даден краен интервал. Да разгледаме линейното пространство от всички непрекъснати в $[a, b]$ функции. Да въведем норма в това пространство, определена с равенството

$$\|f\| := \max_{x \in [a,b]} |f(x)| . \tag{1}$$

Лесно се вижда, че (1) наистина е норма. Тя се нарича *равномерна (или Чебишова) норма*. Оттук нататък, с $C[a, b]$ ще бележим нормираното по този начин просранство от непрекъснати в $[a, b]$ функции. Както вече знаем, всяка норма поражда разстояние. Равномерната норма поражда *равномерното разстояние*

$$\rho(f, g) := \|f - g\| = \max_{x \in [a,b]} |f(x) - g(x)| . \tag{2}$$

В метризираното по този начин пространство $C[a, b]$ да разгледаме задачата за най-добро приближение на непрекъснати функции с алгебрични полиноми.

Величината

$$E_n(f) := \inf_{p \in \pi_n} \|f - p\|$$

ще наричаме *най-добро равномерно приближение* на f с полиноми от степен n. Ако инфимумът се достига за някакъв полином p^* от π_n, т.е. ако $\|f - p^*\| = E_n(f)$, то p^* се нарича *полином на най-добро равномерно приближение* за f в π_n.

Тъй като $C[a, b]$ е нормирано пространство и π_n е линейно подпространство на $C[a, b]$, то въпросът за съществуване на полином на най-добро равномерно приближение за всяка непрекъсната функция f се решава като следствие от общата теорема за приближение в линейни нормирани пространства. В сила е следната теорема.

Теорема на Борел. *За всяка функция f от $C[a, b]$ и всяко цяло неотрицателно число n съществува полином на най-добро равномерно приближение за f от степен n.*

Твърдението следва като частен случай от Теорема 11.5.

Въпросът за единственост не може да бъде решен с общата теорема за приближаване в строго нормирани пространства, защото $C[a, b]$ не е строго нормирано. Това се вижда от следния пример. Нека приемем, за простота, че $[a, b] = [0, 1]$. При $f_1(x) := 1$ и $f_2(x) := x$ имаме

$$\|f_1\| = \|f_2\| = 1, \quad \|f_1 + f_2\| = 2 .$$

Следователно $\|f_1 + f_2\| = \|f_1\| + \|f_2\|$, но очевидно f_1 и f_2 не са линейно зависими.

И така, въпросът за единственост на полинома на най-добро равномерно приближение не е елементарен. Да пристъпим към неговото решаване. Най-напред ще обърнем внимание на едно свойство, забелязано от белгийския математик Вале-Пусен.

Лема на Вале-Пусен. *Нека $Q \in \pi_n$. Нека съществуват $n+2$ точки $x_0 < \ldots < x_{n+1}$ в $[a,b]$ и положителни числа $\lambda_0, \ldots, \lambda_{n+1}$ такива, че*

$$f(x_i) - Q(x_i) = (-1)^i \varepsilon \lambda_i, \qquad i = 0, \ldots, n+1 \ , \tag{3}$$

където $\varepsilon = 1$ или $\varepsilon = -1$. Тогава

$$E_n(f) \geq \lambda := \min_{0 \leq i \leq n+1} \lambda_i \ .$$

Доказателство: Да допуснем противното. Тогава съществува полином $P \in \pi_n$ такъв, че

$$\|f - P\| = E_n(f) < \lambda \ .$$

При това, очевидно $P \neq Q$. Ситуацията е представена геометрично на Рис. 5. Свойството (3) означава, че графиката на Q се нагъва около графиката на f, като разликата $f(x) - Q(x)$ приема стойности $\{\lambda_i\}$ с алтернативно сменящи се знаци (започвайки с плюс, ако $\varepsilon = 1$, или с минус, ако $\varepsilon = -1$). Графиката на полинома P лежи в ивицата с ширина 2λ и централна крива f.

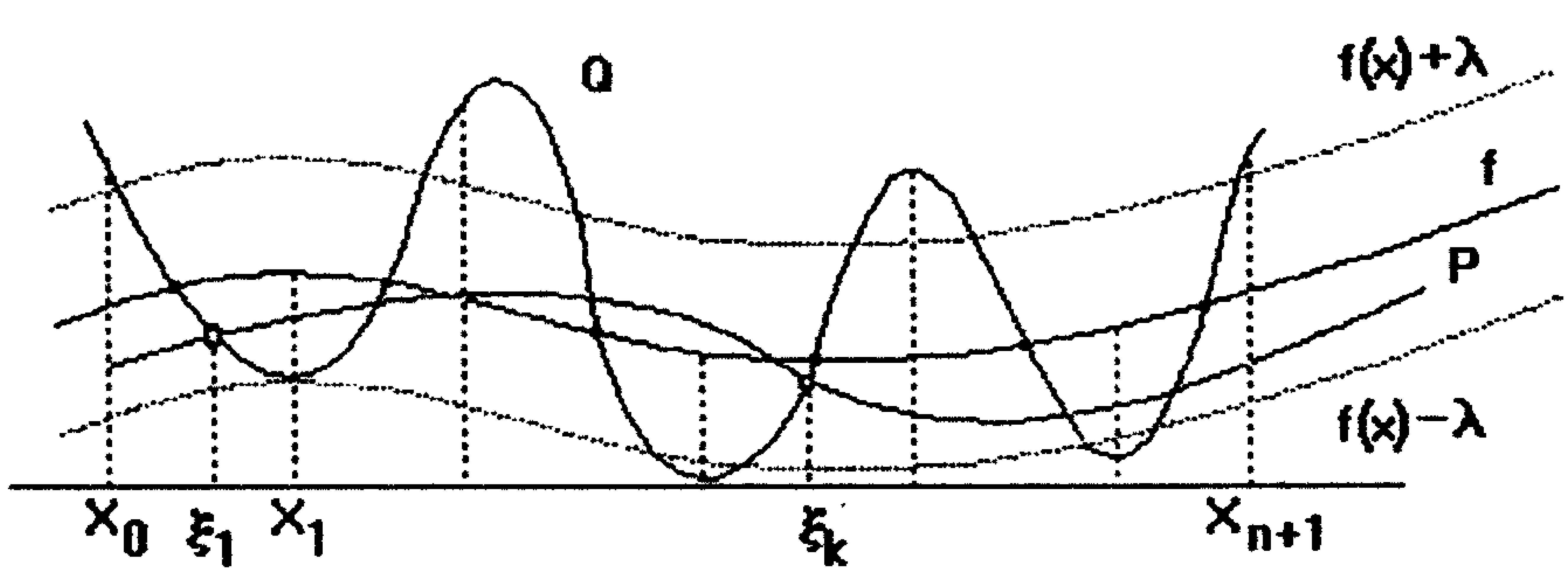

Рис. 5

Тъй като графиката на Q преграждa тази ивица $n+1$ пъти (между точките x_0 и x_1, между x_1 и $x_2, \ldots$, между x_n и x_{n+1}), то графиката на P пресича тази на Q поне в $n+1$ различни точки: $\xi_1, \ldots, \xi_{n+1}$. Тогава $P(\xi_i) - Q(\xi_i) = 0$ за $i = 1, \ldots, n+1$. Но $P - Q \in \pi_n$. Следователно $P \equiv Q$. Стигнахме до противоречие. Следователно $E_n(f) \geq \lambda$. Лемата е доказана.

Пълната характеристика на полинома на най-добро равномерно приближение е дадена от великия руски математик Пафнутий Лвович Чебишов (1821-1894) в знаменитата му *теорема за алтернанса.* Тази теорема е основополагаща в Теория на приближенията.

Теорема на Чебишов за алтернанса. *Нека f е произволна непрекъсната функция в крайния и затворен интервал $[a, b]$. Необходимо и достатъчно условие полиномът P от π_n да бъде полином на най-добро равномерно приближение за f от n-та степен в $[a, b]$ е да съществуват $n+2$ точки $\{x_i\}_{i=0}^{n+1}$ от $[a, b]$ такива, че $a \le x_0 < x_1 < \ldots < x_{n+1} \le b$ и*

$$(4) \qquad f(x_i) - P(x_i) = (-1)^i \varepsilon \; \|f - P\|, \qquad i = 0, \ldots, n+1 \; ,$$

където $\varepsilon = 1$ или $\varepsilon = -1$.

Доказателство: Условието (4) означава, че разликата $f(x) - P(x)$ достига максималната си по модул стойност в $n+2$ точки, като си сменя алтернативно знака. В такъв случай казваме, че f и P осъществяват *алтернанс* в $n+2$ точки.

Достатъчност на условието (4). Нека P удовлетворява (4). Тогава по Лемата на Вале-Пусен,

$$E_n(f) \ge \|f - P\| \; .$$

Но, по определение, $E_n(f) \le \|f - Q\|$ за всяко $Q \in \pi_n$. Следователно $E_n(f) = \|f - P\|$ и P е полином на най-добро равномерно приближение.

Необходимост. Нека P е полином на най-добро приближение за f от степен n. Ще покажем, че съществуват $n+2$ точки на алтернанс, в които P удовлетворява условието (4). Лесно се вижда, че P има поне 2 точки на алтернанс. Наистина, графиката на P лежи в ивицата, определана от $f(x) - E_n(f)$ и $f(x) + E_n(f)$ и тя докосва поне една от тези две гранични линии. Нашата цел е да докажем, че тя се допира и до двете. Да допуснем, че графиката на P се докосва само до едната, например, до горната граница $f(x) + E_n(f)$ и не се допира до долната $f(x) - E_n(f)$. Тогава

$$f(x) - E_n(f) < P(x) \le f(x) + E_n(f)$$

за всяко $x \in [a, b]$ и следователно съществува константа $c > 0$ такава, че

$$f(x) - E_n(f) < P(x) - c < f(x) + E_n(f) \; .$$

Това показва, че $|f(x) - (P(x) - c)| < E_n(f)$ в $[a, b]$. С други думи, полиномът $P(x) - c$ приближава f по-добре от P. Стигнахме до противоречие.

Да допуснем сега, че разликата $f(x) - P(x)$ има най-много $m+2$ точки на алтернанс в $[a, b]$, където $m < n$. Нека $\{x_i\}_0^{m+1}$ са $m+2$ такива точки, т.е. $a \le x_0 < \ldots < x_{m+1} \le b$ и

$$f(x_i) - P(x_i) = (-1)^i \varepsilon \, E_n(f), \quad i = 0, \ldots, m+1 \; ,$$

с някакво $\varepsilon = 1$ или $\varepsilon = -1$. За всеки интервал $[x_{i-1}, x_i]$ определяме точките $\bar{x}_{i-1}$ и $\underline{x}_i$ по следния начин: $\bar{x}_{i-1}$ е точната горна граница на точките x от интервала $[x_{i-1}, x_i]$, за които

(5) $$f(x) - P(x) = (-1)^{i-1}\varepsilon\ E_n(f)\ ,$$

а $\underline{x}_i$ е точната долна граница на точките x от $[x_{i-1}, x_i]$, за които

(6) $$f(x) - P(x) = (-1)^{i}\varepsilon\ E_n(f)\ .$$

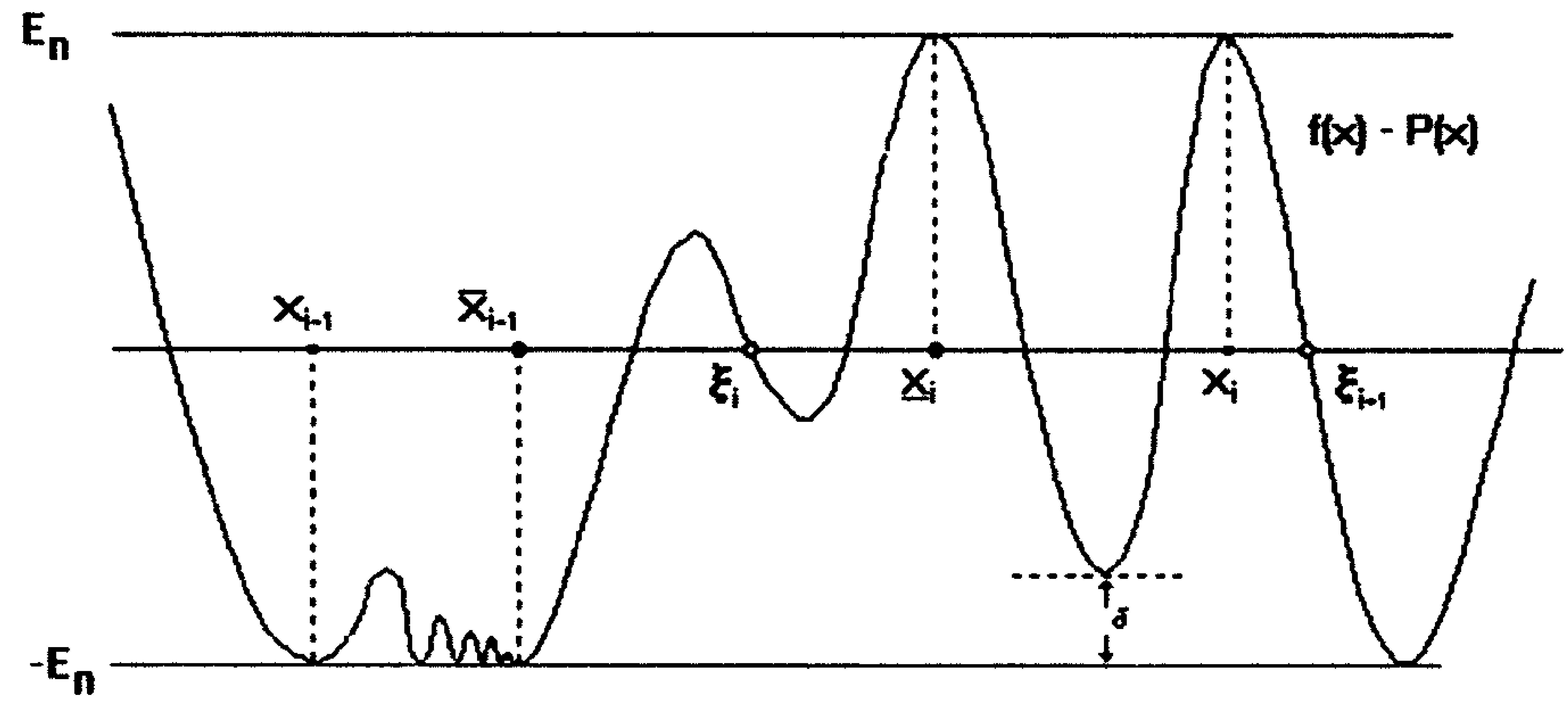

Рис. 6

Тъй като f е непрекъсната функция, то равенствата (5) и (6) остават в сила и за граничните точки $\bar{x}_{i-1}$ и $\underline{x}_i$, съответно. Следователно $\bar{x}_{i-1} < \underline{x}_i$ и $f(x) - P(x)$ (бидейки непрекъсната и с различни знаци в две точки) се анулира в някаква точка ξ_i от $(\bar{x}_{i-1}, \underline{x}_i)$. (Виж Рис. 6.) Това заключение важи за всяко $i = 1, 2, \dots, m+1$. Полагаме $\xi_0 := a$, $\xi_{m+2} := b$. Да разгледаме поведението на разликата $f(x) - P(x)$ в интервала $[\xi_i, \xi_{i+1}]$. От избора на точките $\{\xi_i\}$ се вижда, че $f(x) - P(x)$ приема максималната си по модул стойност $E_n(f)$ само със знак $(-1)^i\varepsilon$ в $[\xi_i, \xi_{i+1}]$. Следователно

(7) $$-E_n(f) < (-1)^i\varepsilon\ [f(x) - P(x)] \le E_n(f)$$

за всяко x от $[\xi_i, \xi_{i+1}]$. И това е вярно за $i = 0, \dots, m+1$. Тъй като интервалите $[\xi_i, \xi_{i+1}]$ са краен брой, то съществува число $\delta > 0$ такова, че

(8) $$-E_n(f) + \delta < (-1)^i\varepsilon\ [f(x) - P(x)] \le E_n(f)$$

за $x \in [\xi_i, \xi_{i+1}]$ и всяко i. Да въведем полинома

$$Q(x) := \lambda(x-\xi_1)\dots(x-\xi_{m+1}) ,$$

където λ е число, избрано така, че да бъдат изпълнени условията:

1) $|Q(x)| \le \delta/2$ за всяко $x \in [a,b]$;
2) $sign\ Q(x) = (-1)^i\varepsilon$ за $x \in (\xi_i, \xi_{i+1})$, $i = 0,\dots,m+1$.

Очевидно тези условия са удовлетворени при

$$\lambda = \frac{(-1)^{m+1}\varepsilon\ \delta}{2\ M}, \text{ където } M := \max_{x\in[a,b]} |(x-\xi_1)\dots(x-\xi_{m+1})| \ .$$

Тъй като $m < n$ по условие, то $Q \in \pi_n$ и следователно $P + Q \in \pi_n$. Сега ще покажем, че полиномът $P + Q$ приближава f по-добре от P. За целта, да разгледаме разликата $f(x)-(P(x)+Q(x))$ в интервала $[\xi_i, \xi_{i+1}]$. От условията

$$-\frac{\delta}{2} \le Q(x) \le \frac{\delta}{2} ,$$

$$sign\ Q(x) = sign\ [f(x_i) - P(x_i)] = (-1)^i\varepsilon$$

и неравенството (8) следва, че

$$-E_n(f) + \frac{\delta}{2} < (-1)^i\varepsilon\ [f(x) - (P(x) + Q(x))] < E_n(f)$$

за всяко $x \in [\xi_i, \xi_{i+1}]$, $i = 0,\dots,m+1$. Тъй като сумата от подинтервали $\{[\xi_i, \xi_{i+1}]\}$ покрива $[a,b]$, то

$$\|f - (P+Q)\| < E_n(f) \ .$$

Стигнахме до противоречие с определението на $E_n(f)$. Следователно $m \ge n$ и доказателството на теоремата е завършено.

Единствеността на полинома на най-добро равномерно приближение следва лесно от теоремата на Чебишов.

Следствие 1. *За всяка непрекъсната в $[a,b]$ функция f съществува единствен полином на най-добро равномерно приближение от степен n.*

Доказателство. Допускаме противното. Тогава съществуват непрекъсната функция f и полиноми P и Q от π_n, за които

$$\|f - P\| = \|f - Q\| = E_n(f) \ . \tag{9}$$

При това, $P \ne Q$. Полиномът $(P+Q)/2$ е също полином на най-добро равномерно приближение на f, защото

$$\begin{aligned} E_n(f) &\le \left\| f - \frac{P+Q}{2} \right\| = \frac{1}{2}\,\|(f-P) + (f-Q)\| \\ &\le \frac{1}{2}\,\|f-P\| + \frac{1}{2}\,\|f-Q\| = E_n(f) \ . \end{aligned}$$

По теоремата на Чебишов за алтернанса, съществуват поне $n+2$ точки $\{x_i\}_0^{n+1}$ такива, че

$$f(x_i) - \frac{P(x_i)+Q(x_i)}{2} = (-1)^i \varepsilon \; E_n(f) \quad (\varepsilon = 1 \text{ или } \varepsilon = -1) \; .$$

Следователно

$$\left| \frac{f(x_i) - P(x_i)}{2} + \frac{f(x_i) - Q(x_i)}{2} \right| = E_n(f) \; . \tag{10}$$

Но

$$|f(x_i) - P(x_i)| \;\leq\; E_n(f) \; ,$$

$$|f(x_i) - Q(x_i)| \;\leq\; E_n(f) \; ,$$

съгласно предположението (9). Тогава, за да бъде изпълнено (10), числата $f(x_i) - P(x_i)$ и $f(x_i) - Q(x_i)$ трябва да имат еднакъв знак и да са равни на $E_n(f)$ по абсолютна стойност, т.е.

$$f(x_i) - P(x_i) = f(x_i) - Q(x_i), \quad i = 0, \dots, n+1 \; .$$

Оттук следва $P(x_i) = Q(x_i)$ за $i = 0, \dots, n+1$, което влече $P \equiv Q$. Стигнахме до противоречие с условието $P \not\equiv Q$. Твърдението е доказано.

Съществуват много малко функции f, за които може да бъде намерен в явен вид полинома на най-добро равномерно приближение. Един интересен пример в това отношение е функцията x^n. Оказва се, че при всяко n полиномът P_{n-1} на най-добро равномерно приближение в $[-1,1]$ за $f(x) = x^n$ от $(n-1)-$ва степен се записва в явен вид и е тясно свързан с полинома на Чебишов $T_n(x)$. Съгласно теоремата за алтернанса, P_{n-1} се определя напълно от условието да съществуват $(n-1)+2$ точки $x_0, \dots, x_n$ в $[-1,1]$ такива, че

$$x_i^n - P_{n-1}(x_i) = (-1)^i \varepsilon \max_{x \in [-1,1]} |x^n - P_{n-1}(x)|, \quad i = 0, \dots, n \; . \tag{11}$$

Но $x^n - P_{n-1}(x)$ е полином от $n-$та степен с водещ коефициент 1. Следователно (11) ще бъде изпълнено, ако построим полином от вида

$$x^n + a_1 x^{n-1} + \ldots + a_n \; ,$$

който приема максималната си стойност в $[-1,1]$ в $n+1$ точки с алтернативно сменящи се знаци. Но ние вече знаем за съществуването на такъв полином. Показахме, че полиномът на Чебишов $T_n(x)$ удовлетворява условията:

$$\begin{aligned}
\max_{x\in[-1,1]} |T_n(x)| &= 1 , \\
T_n(x) &= 2^{n-1}x^n + \dots , \\
T_n(\eta_k) &= (-1)^k, \quad k = 0, \dots, n ,
\end{aligned}$$

където $\eta_k = \cos \frac{k\pi}{n}$.

Следователно полиномът

$$\frac{1}{2^{n-1}} T_n(x) = x^n - P_{n-1}(x)$$

удовлетворява исканите условия (11) в точките $\xi_0, \dots, \xi_n$. Тогава,

$$P_{n-1}(x) = x^n - \frac{1}{2^{n-1}} T_n(x)$$

е полиномът на най-добро равномерно приближение от степен $n-1$ за x^n в $[-1, 1]$.

13. ТЕОРЕМА НА ВАЙЕРЩРАС

Видяхме, че за всяка непрекъсната в $[a,b]$ функция $f(x)$ и всяко фиксирано естествено число n, съществува полином на най-добро равномерно приближение p_n от степен n. Най-доброто приближение означаваме с $E_n(f)$. Възниква естественият въпрос: дали $E_n(f)$ клони към нула, когато n клони към безкрайност? С други думи, дали графиката на полинома на най-добро приближение p_n се приближава все по-плътно до графиката на f при $n \to \infty$? Отговор на този въпрос е дал още Вайерщрас. Той е доказал, че всяка непрекъсната функция е граница на редица от алгебрични полиноми. Или, по-нагледно, колкото и тясна ивица да изберем около графиката на една непрекъсната функция f, може да се намери алгебричен полином p, чиято графика влиза също в тази ивица.

За да се подготвим за доказателството на този резултат, ще въведем някои понятия.

Модул на непрекъснатост. Нека f е функция, определена в $[a,b]$. Величината

$$\omega(f;\delta) := \sup\{\ |f(x)-f(y)|\ :\ x,y \in [a,b],\ |x-y| \le \delta\ \}$$

се нарича *модул на непрекъснатост* на f в $[a,b]$. Този модул е определен за всяко $\delta \in [0, b-a]$. Той характеризира напълно непрекъснатите функции по следния начин: Функцията f е непрекъсната в $[a,b]$ тогава и само тогава, когато $\omega(f;\delta) \to 0$ при $\delta \to 0$. Наистина, ако f е непрекъсната в крайния и затворен интервал $[a,b]$, тя е равномерно непрекъсната и следователно, за всяко $\varepsilon > 0$, съществува число $\delta > 0$ такова, че $|x-y| < \delta$ влече $|f(x)-f(y)| < \varepsilon$ за всяко x, y от $[a,b]$. Оттук следва, че $\omega(f;\delta) \to 0$ при $\delta \to 0$. Обратно, ако $\omega(f;\delta) \to 0$ при $\delta \to 0$, то за всяко $\varepsilon > 0$ може да се намери $\delta = \delta(\varepsilon) > 0$ такова, че $\omega(f;\delta(\varepsilon)) < \varepsilon$. Оттук следва, че $|f(x) - f(y)| < \varepsilon$ при $|x-y| < \delta$, т.е. f е равномерно непрекъсната в $[a,b]$.

Да отбележим някои важни свойства на модула на непрекъснатост.

1) Ако $0 \le \delta_1 \le \delta_2$, то $\omega(f;\delta_1) \le \omega(f;\delta_2)$ (монотонност на ω).

Това свойство следва непосредствено от определението на $\omega(f;\delta)$.

2) За всяко реално число $\lambda \ge 0$ е в сила неравенството

$$\omega(f;\lambda\delta) \le (1+\lambda)\omega(f;\delta)\ .$$

Доказателство. Да предположим най-напред, че λ е цяло число. Нека $\lambda = k$. Нека $x < y$ са произволни точки от $[a,b]$ такива, че $|x-y| \le k\delta$. Разделяме интервала $[x,y]$ на k равни части с помощта на точките

$$x_i = x + i\ (y-x)/k, \qquad i = 0,\ldots,k.$$

Очевидно
$$x_i - x_{i-1} = \frac{y-x}{k} \leq \delta \ , \quad i = 1, \dots, k \ .$$
Имаме
$$|f(x) - f(y)| = \left|\sum_{i=1}^{k}(f(x_i) - f(x_{i-1}))\right| \leq \sum_{i=1}^{k}|f(x_i) - f(x_{i-1})| \leq k\,\omega(f;\delta) \ .$$

И така, ако $|x - y| \leq k\,\delta$, то $|f(x) - f(y)| \leq k\,\omega(f;\delta)$. Следователно
$$\omega(f; k\,\delta) \leq k\,\omega(f;\delta)$$
и неравенството от 2) е доказано при $\lambda = k$.

Нека сега λ е произволно реално положително число. Тогава, $[\lambda] < \lambda < [\lambda] + 1$, където $[\lambda]$ е цялата част на λ. От току-що доказаното неравенство и от монотонността на $\omega(f;\delta)$ следва
$$\omega(f;\lambda\delta) \leq \omega(f;([\lambda]+1)\delta) \leq ([\lambda]+1)\,\omega(f;\delta) \leq (\lambda+1)\,\omega(f;\delta)$$
и твърдението 2) е доказано.

Полиноми на Бернщайн. Ние ще докажем резултата на Вайерщрас като използваме доказателството, предложено от С. Н. Бернщайн (1880–1968). За всяка непрекъсната функция f в $[a,b]$, той построява в явен вид алгебрични полиноми, които клонят към тази функция относно равномерното разстояние.

Нека $f(t)$ е произволна функция, определена в интервала $[0,1]$. *Полиномът на Бернщайн* от степен n за функцията f се бележи с $B_n(f;t)$ и се определя с равенството
$$B_n(f;t) := \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k}t^k(1-t)^{n-k} \ .$$

Очевидно $B_n \in \pi_n$, $B_n(f;0) = f(0)$, $B_n(f;1) = f(1)$. Освен това
$$\begin{aligned} B_n(cf;t) &= c\,B_n(f;t) \\ B_n(f+g;t) &= B_n(f;t) + B_n(g;t) \ . \end{aligned}$$

Последните две свойства показват, че $B_n(f;t)$ е един *линеен оператор* в пространството от функции, определени в $[0,1]$.

Да отбележим още, че полиномите
$$\varphi_{nk}(t) := \binom{n}{k}t^k(1-t)^{n-k}$$

са положителни в $(0,1)$. Оттук следва, че ако $f(t) \geq 0$ за всяко $t \in [0,1]$, то $B_n(f;t) \geq 0$ за всяко $t \in [0,1]$. Това свойство се нарича *положителност* на оператора $B_n(f;t)$. От него веднага следва *монотонността* на $B_n(f;t)$, т.е.

$$f(t) \leq g(t) \quad \text{в} \quad [0,1] \quad \text{влече} \quad B_n(f;t) \leq B_n(g;t) \quad \text{в} \quad [0,1] .$$

Да намерим $B_n(f;x)$ за функциите $f(t)=1$, $f(t)=t$ и $f(t)=t^2$. Те ще ни потрябват по-нататък. Ще докажем следните равенства:

а) $B_n(1;x) = 1$;

б) При $f(t)=t$, $B_n(f;x) = x$;

в) При $f(t)=t^2$, $B_n(f;x) = x^2 + \frac{x(1-x)}{n}$.

Доказателство. Прилагаме последователно два пъти оператора $\frac{1}{n}\, p\, \frac{d}{dp}$ към тъждеството

$$(A) \qquad \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n$$

и получаваме съответно

$$(B) \qquad \sum_{k=0}^{n} \binom{n}{k} \frac{k}{n} p^k q^{n-k} = \frac{p}{n} n(p+q)^{n-1} = p(p+q)^{n-1},$$

$$(C) \qquad \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^2 p^k q^{n-k} = \frac{p}{n}\Big[(p+q)^{n-1} + (n-1)p(p+q)^{n-2}\Big].$$

Изразите за $B_n(f;x)$ следват от (A), (B) и (C) при $p=x$ и $q=1-x$.

Бележка. Следващата лема дава представяне на $B_n(f;t)$ по степените на t. За целта, да означим с $\Delta^k f_0$ крайната разлика на f в точките $0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{k}{n}$.

Лема 1.

$$B_n(f;t) = \sum_{k=0}^{n} \Delta^k f_0 \cdot \binom{n}{k} t^k .$$

Доказателство. Като използваме биномната формула на Нютон за $(1-t)^{n-k}$, получаваме

$$B_n(f;t) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} t^k \sum_{j=0}^{n-k} (-1)^{n-k-j} \binom{n-k}{j} t^{n-k-j} ,$$

което след смяната $m = n-j$, се преобразува в

$$B_n(f;t) = \sum_{k=0}^{n} \sum_{m=k}^{n} (-1)^{m-k} \binom{n-k}{n-m} \binom{n}{k} f\left(\frac{k}{n}\right) t^m .$$

Тъй като

$$\binom{n-k}{n-m}\binom{n}{k}=\binom{n}{m}\binom{m}{k},$$

след като разменим реда на сумиране, получаваме

$$B_n(f;t)=\sum_{m=0}^{n}\binom{n}{m}\left\{\sum_{k=0}^{m}(-1)^{m-k}\binom{m}{k}f\left(\frac{k}{n}\right)\right\}t^m,$$

което е исканото представяне на $B_n(f;t)$. Лемата е доказана.

Бележка. От Лема 1 следва например, че ако f е полином от степен m, то $B_n(f;t)\in\pi_m$ за всяко n. (*Упътване*: използвайте свойството на крайните разлики да анулират алгебричните полиноми.)

Сега можем и по още един начин да намерим лесно $B_n(f;x)$ за функциите $f(t)=1$, $f(t)=t$ и $f(t)=t^2$. Наистина,

$$B_n(1;t)=\sum_{k=0}^{n}\binom{n}{k}t^k(1-t)^{n-k}=\{t+(1-t)\}^n=1^n=1\,.$$

При $f(t)=t$, от Лема 1, имаме

$$B_n(f;t)=\Delta f_0\cdot\binom{n}{1}t+f(0)\binom{n}{0}1=\left(\frac{1}{n}-0\right)nt=t\,.$$

При $f(t)=t^2$, отново по Лема 1,

$$\begin{aligned}B_n(f;t)&=\Delta^2 f_0\cdot\binom{n}{2}t^2+\Delta f_0\cdot\binom{n}{1}t+f(0)\binom{n}{0}.1\\&=\left(\left(\frac{2}{n}\right)^2-2\left(\frac{1}{n}\right)^2+0\right)\frac{n(n-1)}{2}t^2+\frac{t}{n}\\&=\frac{2}{n^2}\,\frac{n(n-1)}{2}t^2+\frac{t}{n}=\frac{n-1}{n}t^2+\frac{t}{n}=t^2+\frac{t(1-t)}{n}\,.\end{aligned}$$

От тези примери се вижда, че $B_n(f;t)\to f(t)$ при $n\to\infty$ за $f(t)=1,t,t^2$. По-долу ще видим, как от сходимостта на полиномите на Бернщайн за тези специални функции, следва сходимост за всяка непрекъсната функция в $[0,1]$.

Теорема 2. *Нека f е произволна функция дефинирана в $[0,1]$. Тогава за всяко n и всяко $t\in[0,1]$ имаме*

$$|f(t)-B_n(f;t)|\le\frac{3}{2}\omega\left(f;\frac{1}{\sqrt{n}}\right)\,.$$

Доказателство. Нека t е произволна точка от $[0,1]$. Да отбележим, че

$$f(t)=\sum_{k=0}^{n}f(t)\binom{n}{k}t^k(1-t)^{n-k}=\sum_{k=0}^{n}f(t)\varphi_{nk}(t)\,.$$

Тогава

$$\begin{aligned}
|f(t) - B_n(f;t)| &= \left| \sum_{k=0}^{n} \left[f(t) - f\left(\frac{k}{n}\right) \right] \varphi_{nk}(t) \right| \\
&\le \sum_{k=0}^{n} \left| f(t) - f\left(\frac{k}{n}\right) \right| \varphi_{nk}(t) \\
&\le \sum_{k=0}^{n} \omega\left(f; \left|t - \frac{k}{n}\right|\right) \varphi_{nk}(t) .
\end{aligned}$$

Но от свойство 2) на $\omega(f;\delta)$ имаме

$$\omega\left(f; \left|t - \frac{k}{n}\right|\right) = \omega\left(f; \frac{|t-\frac{k}{n}|}{\delta}\delta\right) \le \left(\frac{1}{\delta}\left|t - \frac{k}{n}\right| + 1\right)\omega(f;\delta)$$

за всяко $\delta > 0$. Тогава

$$|f(t) - B_n(f;t)| \le \frac{\omega(f;\delta)}{\delta} \sum_{k=0}^{n} \left|t - \frac{k}{n}\right| \varphi_{nk}(t) + \omega(f;\delta) .$$

По неравенството на Коши-Буняковски-Шварц

$$\begin{aligned}
\sum_{k=0}^{n} \left|t - \frac{k}{n}\right| \varphi_{nk}(t) &\le \left(\sum_{k=0}^{n} \left(t - \frac{k}{n}\right)^2 \varphi_{nk}(t)\right)^{\frac{1}{2}} \left(\sum_{k=0}^{n} \varphi_{nk}(t)\right)^{\frac{1}{2}} \\
&= \sqrt{B_n((t-x)^2;t)} .
\end{aligned}$$

Тук $B_n\left((t-x)^2;t\right)$ е стойността на полинома на Бернщайн в точката t за функцията $(t-x)^2$ като функция на x (t е параметър). Окончателно получаваме

$$|f(t) - B_n(f;t)| \le \frac{\omega(f;\delta)}{\delta}\sqrt{B_n((t-x)^2;t)} + \omega(f;\delta) .$$

Но поради линейността на оператора $B_n(f;t)$ и като вземем пред вид намерените в а), б) и в) изрази за $B_n(f;t)$ при $f(x) = 1, x, x^2$, получаваме

$$\begin{aligned}
B_n\left((t-x)^2;t\right) &= B_n(t^2 - 2tx + x^2;t) \\
&= t^2B_n(1;t) - 2tB_n(x;t) + B_n(x^2;t) \\
&= t^2 - 2t.t + t^2 + \frac{t(1-t)}{n} = \frac{t(1-t)}{n} .
\end{aligned}$$

Следователно за всяко $\delta > 0$ и всяко $t \in [0,1]$ имаме

$$|f(t) - B_n(f;t)| \le \omega(f;\delta)\left\{\frac{1}{\delta}\sqrt{\frac{t(1-t)}{n}} + 1\right\} .$$

Но $t(1-t) \leq \frac{1}{4}$ за $t \in [0,1]$. Следователно

$$|f(t) - B_n(f;t)| \leq \omega(f;\delta) \left\{ \frac{1}{2\delta\sqrt{n}} + 1 \right\} .$$

Избирайки $\delta = \frac{1}{\sqrt{n}}$ получаваме

$$|f(t) - B_n(f;t)| \leq \frac{3}{2}\, \omega\left(f; \frac{1}{\sqrt{n}}\right) .$$

Сега вече сме готови да дадем доказателство на теоремата на Вайерщрас за произволен краен интервал $[a,b]$.

Теорема на Вайерщрас. *Нека $[a,b]$ е произволен краен интервал и $f(x)$ е непрекъсната функция в $[a,b]$. Тогава, за всяко $\varepsilon > 0$ съществува алгебричен полином $P(x)$ такъв, че*

$$\max_{x\in[a,b]} |f(x) - P(x)| \leq \varepsilon ,$$

Доказателство: Ще използваме резултата от Теорема 2. За целта да въведем функцията

$$h(t) := f(a + t(b-a)) ,$$

определена за $t \in [0,1]$. Тъй като f е непрекъсната в $[a,b]$, то $h(t)$ е непрекъсната в $[0,1]$ и следователно

$$\lim_{\delta\to 0} \omega(h,\delta) = 0.$$

Тогава съществува n такова, че

$$\frac{3}{2}\, \omega\left(h; \frac{1}{\sqrt{n}}\right) < \varepsilon .$$

По Теорема 2,

$$|h(t) - B_n(h;t)| \leq \frac{3}{2}\, \omega\left(h; \frac{1}{\sqrt{n}}\right) < \varepsilon$$

за това n, т.е. полиномът $B_n(h;t)$ приближава h в $[0,1]$ с точност ε. Сега, като се върнем обратно към променливата x със смяната $t = \frac{x-a}{b-a}$, получаваме

$$\left| h\left(\frac{x-a}{b-a}\right) - B_n\left(h; \frac{x-a}{b-a}\right) \right| < \varepsilon \qquad \text{за } x \in [a,b] .$$

Остава да забележим, че $f(x) = h\left(\frac{x-a}{b-a}\right)$ и $P(x) = B_n\left(h; \frac{x-a}{b-a}\right)$ е алгебричен полином от степен n. И така, намерихме полином P, за който

$$|f(x) - P(x)| < \varepsilon \qquad \text{при } x \in [a,b] .$$

Доказателството е завършено.

Аналогичен резултат е в сила и за приближаване на непрекъснати периодични функции с тригонометрични полиноми. Той може да бъде доказан директно, като се използва конкретен тригонометричен оператор. Ние ще изведем тук този резултат, като следствие от алгебричния случай.

Втора теорема на Вайерщрас. *Нека $f(t)$ е произволна непрекъсната 2π–периодична функция. Тогава за всяко $\varepsilon > 0$ съществува тригонометричен полином $\tau(t)$ такъв, че*

$$\max_{t\in(-\infty,\infty)} |f(t) - \tau(t)| \le \varepsilon \ . \tag{1}$$

Доказателство. Тъй като тригонометричните полиноми са също 2π-периодични функции, то достатъчно е да докажем (1) в интервал с дължина 2π, например, в $[-\pi, \pi]$.

Ще се уговорим да пишем

$$h(t) \approx \tau(t) \ ,$$

когато искаме да кажем, че съществува тригонометричен полином, който приближава h в $[-\pi,\pi]$ с точност ε.

Сега ще използваме факта, че всяка функция f може да бъде представена като сума от четна и нечетна функция по формулата

$$f(t) = \frac{f(t) + f(-t)}{2} + \frac{f(t) - f(-t)}{2} \ . \tag{2}$$

Да припомним, че $F(t)$ е *четна*, ако $F(t) = F(-t)$ и *нечетна*, ако $F(t) = -F(-t)$ за всяко t.

Да предположим най-напред, че $f(t)$ е четна периодична функция. Да разгледаме функцията

$$g(x) := f(\arccos x) \ , \qquad x \in [-1,1] \ .$$

Вижда се, че когато x пробягва $[-1,1]$, променливата $t = \arccos x$ пробягва $[0,\pi]$. Функцията $g(x)$ е непрекъсната в $[-1,1]$. Тогава, по Теоремата на Вайерщрас, съществува алгебричен полином $P(x)$ такъв, че $|g(x)-P(x)| \le \varepsilon$ в $[-1,1]$.

Като се върнем към променливата t със смяната $x = \cos t$, получаваме

$$|g(\cos t) - P(\cos t)| \le \varepsilon \qquad \text{в } [0,\pi] \ . \tag{3}$$

Не е трудно да се види, че $P(\cos t)$ е тригонометричен полином. Наистина, $P(x)$ може да се представи във вида

$$P(x) = \sum_{k=0}^{n} a_k T_k(x) \ ,$$

където $T_k(x) = \cos(k \arccos x)$ са полиномите на Чебишов. Тогава

$$P(\cos t) = \sum_{k=0}^{n} a_k \cos kt ,$$

т.е. $P(\cos t)$ е тригонометричен полином от ред n. При това, $P(\cos t)$ е четна функция на t (защото $\cos kt$ са четни функции). Тъй като $f(t) = g(\cos t)$ е също четна по предположение, то неравенството (3) ще бъде изпълнено и в симетричния интервал $[-\pi, 0]$. И така, получихме, че ако $f(t)$ е четна функция, то съществува тригонометричен полином, който приближава f в $[-\pi, \pi]$ с точност ε и съгласно нашата уговорка този резултат се записва така

$$f(t) \approx \tau(t) .$$

Но тогава, като умножим двете страни със $\sin^2 t$, получаваме

$$f(t)\sin^2 t \approx \tau(t) . \tag{4}$$

Да предположим сега, че $f(t)$ е нечетна функция. Тогава $f(t)\sin t$ е четна и съгласно доказаното по-горе,

$$f(t)\sin t \approx \tau(t) .$$

Оттук, като умножим двете страни със $\sin t$ стигаме отново до (4). И така, (4) е изпълнено и за всяка нечетна функция. Тъй като една произволна функция се представя чрез (2) като сума от четна и нечетна функция, то заключаваме, че (4) е в сила за произволна непрекъсната 2π-периодична функция.

Нашата цел е да покажем, че $f(t) \approx \tau(t)$. Ние ще изведем това приближение от (4). Разглеждаме функцията $f\left(\theta - \frac{\pi}{2}\right)$. Тя е също непрекъсната и $2\pi-$периодична. Следователно, по (4),

$$f\left(\theta - \frac{\pi}{2}\right)\sin^2(\theta) \approx \tau(\theta) .$$

Оттук, след смяната $t = \theta - \frac{\pi}{2}$, получаваме

$$f(t)\cos^2 t \approx \tau(t) .$$

Сега, като съберем това неравенство с (4) стигаме до заключението

$$\begin{aligned} f(t)\cos^2 t + f(t)\sin^2 t &\approx \tau(t) \\ \Rightarrow\ f(t)(\cos^2 t + \sin^2 t) &\approx \tau(t) \\ \Rightarrow\ f(t) &\approx \tau(t) \qquad (\text{защото } \cos^2 t + \sin^2 t = 1) . \end{aligned}$$

Теоремата е доказана.

14. ОРТОГОНАЛНИ ПОЛИНОМИ

По-нататък ние многократно ще използваме различни свойства на така наречените ортогонални полиноми. Затова тук ще дадем някои основни предварителни сведения за тях.

Нека $[a,b]$ е даден интервал, краен или безкраен и функцията $\mu(x)$ е определена и неотрицателна в $[a,b]$. Да предположим още, че

$$\int_{\alpha}^{\beta} \mu(x)\,dx > 0$$

за всеки подинтервал $[\alpha,\beta]$ на $[a,b]$. Всяка функция $\mu(x)$ с тези свойства се нарича *теглова функция* или *тегло* в $[a,b]$. Да определим скаларно произведение (f,g) на две функции $f(x)$ и $g(x)$ по следния начин:

$$(f,g) = \int_a^b \mu(x)f(x)g(x)\,dx.$$

Разбира се, тук предполагаме, че f и g са определени в $[a,b]$ и интегралът по-горе съществува.

Определение. Функциите $f(x)$ и $g(x)$ са *ортогонални* в $[a,b]$ с тегло $\mu(x)$, ако $(f,g)=0$.

Определение. Ще казваме, че $P_0(x), P_1(x), P_2(x), \ldots$ е (крайна или безкрайна) *редица от ортогонални полиноми* в $[a,b]$ с тегло $\mu(x)$, ако

а) $P_i \in \pi_i, \quad \forall i,$

б) $(P_i,P_i) \neq 0, \quad \forall i,$

в) $(P_i,P_j) = 0$ при $i \neq j$.

Да отбележим някои свойства на ортогоналните полиноми.

Свойство 1. Всяка крайна подредица $P_0(x), \ldots, P_n(x)$ на

$$P_0(x), P_1(x), P_2(x), \ldots$$

е линейно независима система от функции.

Доказателство. Допускаме противното. Тогава съществува число n и коефициенти $a_0, \ldots, a_n$, поне един от които е различен от нула и такива, че полиномът $f(x) = a_0P_0(x) + \cdots + a_nP_n(x)$ е тъждествено равен на нула. Тогава

(1) $$(f,P_i) = 0 \quad \text{за всяко} \quad i.$$

От друга страна,

$$(f, P_i) = \sum_{k=0}^{n} a_k(P_k, P_i) = a_i(P_i, P_i)$$

и знаем, че поне един от коефициентите $a_0, \dots, a_n$ е различен от нула. Това, заедно с условието б), води до противоречие с (1). Твърдението е доказано.

Свойство 2. Ако полиномът $f(x)$ е от степен по-малка или равна на n, то f се представя по единствен начин във вида

$$f(x) = a_0P_0(x) + \cdots + a_nP_n(x)$$

с някакви реални коефициенти $a_0, \dots, a_n$.

Това е просто следствие от Свойство 1. Тъй като π_n е $(n+1)$-мерно линейно пространство и $P_0, \dots, P_n$ са (съгласно Свойство 1) $n+1$ линейно независими елемента от π_n, то всеки елемент на π_n се представя като тяхна линейна комбинация. Лесно могат да се получат и явни изрази за a_k, $k = 0, \dots, n$. Наистина, имаме

$$\begin{aligned}(f, P_k) &= a_0(P_0, P_k) + \cdots + a_k(P_k, P_k) + \cdots + a_n(P_n, P_k)\\ &= a_k(P_k, P_k).\end{aligned}$$

Оттук

$$a_k = \frac{(f, P_k)}{(P_k, P_k)}.$$

Струва си да се отбележи още, че от условията а), б) и в) следва, че $P_n(x)$ е алгебричен полином от степен точно n, т.е. $P_n(x)$ е от вида

(2) $$P_n(x) = \alpha_n x^n + \text{полином от } \pi_{n-1},$$

където $\alpha_n \neq 0$. Да допуснем противното. Тогава $P_n \in \pi_{n-1}$ и съгласно Свойство 2, $P_n(x)$ може да се запише във вида

$$P_n(x) = a_0P_0(x) + \cdots + a_{n-1}P_{n-1}(x)$$

с някакви константи $a_0, \dots, a_{n-1}$. Но горната връзка означава, че $P_0, \dots, P_n$ са линейно зависими, което противоречи на Свойство 1.

Свойство 3. Нека $f(x)$ е произволен полином от степен по-малка или равна на $n-1$. Тогава $(f, P_n) = 0$.

Доказателство. Тъй като $f \in \pi_{n-1}$, то по Свойство 2,

$$f(x) = a_0P_0(x) + \cdots + a_{n-1}P_{n-1}(x).$$

Оттук, съгласно в),

$$(f, P_n) = a_0(P_0, P_n) + \cdots + a_{n-1}(P_{n-1}, P_n) = 0.$$

Свойство 4. При всяко естествено число n, полиномът $P_n(x)$ има n различни нули, които лежат в отворения интервал (a, b).

Доказателство. Да допуснем, че полиномът $P_n(x)$ има само k смени на знака си в (a, b) и $k < n$. Нека $\{\xi_i\}_{i=1}^k$, $a < \xi_1 < \cdots < \xi_k < b$, са точките, в които $P_n(x)$ си сменя знака. Избираме произволна точка t от интервала (ξ_k, b) в която $P_n(t) \neq 0$ и построяваме полинома

$$Q(x) = P_n(x)(x - \xi_1) \dots (x - \xi_k) \; .$$

Очевидно $Q(x)P_n(x)$ е алгебричен полином, който не е равен на нула и $Q(x)P_n(x) \geq 0$ в $[a, b]$. Следователно

$$(Q, P_n) = \int_a^b \mu(x)Q(x)P(x)\, dx > 0.$$

От друга страна, $Q \in \pi_{n-1}$ (защото $k < n$) и по Свойство 3, $(Q, P_n) = 0$. Стигнахме до противоречие. Следователно $k \geq n$. Тъй като $P_n(x)$ си сменя знака в ξ_i, $i = 1, \dots, k$ то $\xi_1, \dots \xi_k$ са нули за $P_n(x)$. Но $P_n \in \pi_n$. Тогава съгласно основната теорема на алгебрата $P_n(x)$ има най-много n реални нули в (a, b). Следователно k е точно равно на n, т.е. $\xi_1, \dots, \xi_n$ са всичките нули на $P_n(x)$, а те по определение лежат в (a, b) и са различни.

Свойство 5. За всяка система от ортогонални полиноми

$$P_0(x), P_1(x), \dots$$

съществува тричленна рекурентна връзка от вида

$$P_{n+1}(x) = (A_n x - B_n)P_n(x) + C_n P_{n-1}(x), \quad n = 1, 2, \dots \; , \tag{3}$$

където A_n, B_n, C_n са константи.

Доказателство. По Свойство 2,

$$xP_n(x) = a_0 P_0(x) + \cdots + a_n P_n(x) + a_{n+1} P_{n+1}(x)$$

с някакви константи $a_0, \dots, a_{n+1}$. Умножаваме това равенство с $P_i(x)$ и интегрираме. Получаваме

$$\int_a^b \mu(x) x P_n(x) P_i(x)\, dx = a_i(P_i, P_i), \quad i = 0, \dots, n+1 \; , \tag{4}$$

като при това интегралът вляво е равен на нула при $i = 0, \ldots, n-2$, съгласно Свойство 3, защото $xP_i(x) \in \pi_{n-1}$ в този случай. Следователно $a_i = 0$ при $i = 0, \ldots, n-2$ и

$$xP_n(x) = a_{n-1}P_{n-1}(x) + a_nP_n(x) + a_{n+1}P_{n+1}(x) \ . \tag{5}$$

Оттук получаваме исканата връзка. От (4) намираме

$$\begin{aligned} a_{n-1} &= \frac{(xP_n, P_{n-1})}{(P_{n-1}, P_{n-1})} \ , \\ a_n &= \frac{(xP_n, P_n)}{(P_n, P_n)} \ , \end{aligned}$$

докато от (5) се вижда, че $a_{n+1} = \alpha_n/\alpha_{n+1}$, ако

$$P_k(x) = \alpha_k x^k + \text{ полином от } \ \pi_{k-1}, \quad k = 0, 1, \ldots \ .$$

Свойство 6. За ортогоналния полином $P_n(x)$ с тегло $\mu(x)$ в интервала $[a, b]$ и за всеки полином $Q_n(x) \in \pi_n$, имащ същия коефициент пред x^n, както и $P_n(x)$, е в сила неравенството

$$\int_a^b \mu(x)P_n^2(x)\,dx \le \int_a^b \mu(x)Q_n^2(x)\,dx,$$

като равенството е в сила само при $P_n(x) \equiv Q_n(x)$.

Доказателство. Съгласно условията имаме

$$\begin{aligned} P_n(x) &= \alpha_n x^n + q_{n-1}(x) \\ Q_n(x) &= \alpha_n x^n + r_{n-1}(x) \ , \end{aligned}$$

където $\alpha_n \ne 0$ и $q_{n-1}, r_{n-1} \in \pi_{n-1}$. Тогава

$$\begin{aligned} \int_a^b \mu(x)Q_n^2(x)\,dx &= \int_a^b \mu(x)\left[\alpha_n x^n + r_{n-1}(x)\right]^2 dx \\ &= \int_a^b \mu(x)[P_n(x) + r_{n-1}(x) - q_{n-1}(x)]^2\,dx \\ &= \int_a^b \mu(x)P_n^2(x)\,dx \\ &+ 2\int_a^b \mu(x)P_n(x)\left[r_{n-1}(x) - q_{n-1}(x)\right]\,dx \\ &+ \int_a^b \mu(x)[r_{n-1}(x) - q_{n-1}(x)]^2\,dx \ . \end{aligned}$$

Тъй като $r_{n-1} - q_{n-1} \in \pi_{n-1}$, то $\int_a^b \mu(x)P_n(x)\left[r_{n-1}(x) - q_{n-1}(x)\right] dx = 0$. Получихме, че

$$\int_a^b \mu(x)Q_n^2(x)\,dx - \int_a^b \mu(x)P_n^2(x)\,dx \geq 0,$$

като равенството е изпълнено, ако $r_{n-1}(x) \equiv q_{n-1}(x)$, т.е., ако $Q_n(x) \equiv P_n(x)$. Свойството е доказано.

Накрая остава да изясним един основен въпрос – въпроса за съществуване и евентуално построяване на редица от ортогонални полиноми при даден интервал $[a, b]$ и тегло $\mu(x)$. Нека $[a, b]$ е произволен интервал и $\mu(x)$ е произволно тегло в $[a, b]$. Ще поискаме $\mu(x)$ да удовлетворява допълнителното условие

$$\int_a^b \mu(x)x^k\,dx < \infty, \quad k = 0, 1, \ldots,$$

ако интервалът $[a, b]$ е безкраен. За да построим една редица от ортогонални полиноми можем да постъпим по следния начин.

1. Избираме произволна редица от числа α_0, α_1, ... , всяко от които е различно от нула. Това ще бъдат коефициентите пред x^n в $P_n(x)$, съответно за $n = 0, 1, \ldots$. Следователно $P_0(x) \equiv \alpha_0$.

2. При $n = 1, 2, \ldots$ построяваме полинома

$$P_n(x) = \alpha_n x^n + \text{полином от } \pi_{n-1}$$

така, че да удовлетворява условията

$$(P_n, P_i) = 0, \quad i = 0, \ldots n-1\ . \tag{6}$$

Очевидно условията (6) ще бъдат изпълнени, ако $(P_n, f) = 0$ за всяко $f \in \pi_{n-1}$. Следователно цялата работа се свежда до построяване на полином $P_n(x)$ от степен n с фиксиран коефициент α_n пред x^n, който е ортогонален на всички полиноми от класа π_{n-1}. Тази задача има и самостоятелен интерес, затова ще я отделим като теорема.

Теорема 1. *При даден интервал $[a, b]$, тегло $\mu(x)$ и коефициент α_n, съществува единствен полином от вида*

$$P_n(x) = \alpha_n x^n + \textit{полином от } \pi_{n-1},$$

който е ортогонален на всички полиноми от степен по-малка или равна на $n-1$.

Доказателство. Ще приложим индукция по n. За $n = 0$ полиномът $P_0(x)$ е определен еднозначно от условието $P_0(x) = \alpha_0$. Да допуснем, че сме определили еднозначно $P_0, P_1, \ldots, P_{n-1}$. Тъй като те образуват ортогонална система от полиноми, то всеки полином от степен $n-1$ се представя като тяхна линейна комбинация. Следователно $P_n(x)$ се записва във вида

$$P_n(x) = \alpha_n x^n + b_{n-1}P_{n-1}(x) + \cdots + b_0 P_0(x).$$

От условията за ортогоналност (6) получаваме

$$(P_n, P_i) = \alpha_n(x^n, P_i) + b_i(P_i, P_i),$$

откъдето, при дадено α_n, коефициентите b_i се определят еднозначно. Теоремата е доказана.

И така, при даден интервал $[a, b]$, тегло $\mu(x)$ и фиксирани коефициенти $\alpha_0, \alpha_1, \ldots$ пред най-високите степени, съществува единствена система от ортогонални полиноми.

Пример. Да се докаже, че полиномите

$$L_n(x) = \frac{1}{2^n n!}\frac{d^n(x^2-1)^n}{dx^n}, \qquad n = 0, 1, \ldots,$$

образуват ортогонална система в $[-1, 1]$ при тегло $\mu(x) \equiv 1$.

Решение. Трябва да проверим условията а), б) и в). Очевидно $(x^2-1)^n$ е полином от степен точно $2n$. Следователно неговата n-та производна е полином от степен точно n. Оттук следва а) и б). Остава да докажем в). Ние ще покажем, че

$$\int_{-1}^{1} L_n(x) f(x)\, dx = 0$$

за всеки полином $f \in \pi_{n-1}$. Наистина, като означим функцията $\frac{1}{2^n n!}(x^2-1)^n$ за простота с $\varphi(x)$, след многократно интегриране по части получаваме

$$\begin{aligned}
\int_{-1}^{1} L_n(x) f(x)\, dx &= \int_{-1}^{1} f(x)\varphi^{(n)}(x)\, dx = \int_{-1}^{1} f(x)\, d\varphi^{(n-1)}(x) \\
&= f(x)\varphi^{(n-1)}(x)\Big|_{-1}^{1} - \int_{-1}^{1} f'(x)\varphi^{(n-1)}(x)\, dx \\
&\ldots \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&= \sum_{k=1}^{n} (-1)^{k-1} f^{(k-1)}\varphi^{(n-k)}\Big|_{-1}^{1} + (-1)^n \int_{-1}^{1} f^{(n)}(x)\varphi(x)\, dx.
\end{aligned}$$

Очевидно последният израз е равен на нула, защото $f^{(n)}(x) \equiv 0$ и $\varphi^{(n-k)}(x)$ се анулира при $x = \pm 1$ за $k = 1, \ldots, n$.

Полиномите $L_n(x)$ се наричат *полиноми на Льожандър.*

Коефициентът $\frac{1}{2^n n!}$ е поставен за да бъде изпълнено условието

$$L_n(1) = 1, \quad n = 0, 1, 2, \ldots$$

Известните ни вече *полиноми на Чебишов от първи род*

$$T_n(x) = \cos n \arccos x$$

са ортогонални в $[-1, 1]$ с тегло $\mu(x) = \frac{1}{\sqrt{1-x^2}}$.

Полиномите на Чебишов от втори род

$$U_n(x) := T'_{n+1}(x)$$

са ортогонални в $[-1, 1]$ с тегло $\mu(x) = \sqrt{1 - x^2}$.

Полиномите на Льожандър, на Чебишов от първи и втори род са специални случаи на така наречените *полиноми на Якоби* $\{P_n^{(\alpha,\beta)}\}$, които са ортогонални в $[-1, 1]$ с тегло $(1 - x)^\alpha (1 + x)^\beta$ при $\alpha, \beta > -1$. Те се дават с формулата

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n\, n!}(1 - x)^{-\alpha}(1 + x)^{-\beta} \frac{d^n}{dx^n}\{(1 - x)^{n+\alpha}(1 + x)^{n+\beta}\} .$$

Полиномите на Ермит

$$H_n(x) := (-1)^n e^{x^2} \frac{d^n}{dx^n}(e^{-x^2})$$

са ортогонални в $(-\infty, \infty)$ с тегло e^{-x^2}.

15. ПРИБЛИЖЕНИЕ В ХИЛБЕРТОВО ПРОСТРАНСТВО

Едно линейно пространство H се нарича хилбертово, ако в него е въведено *скаларно произведение.* Това значи, че на всеки два елемента f, g от H се съпоставя числото (f, g), което удовлетворява условията:

1) $(f, f) \geq 0$, като $(f, f) = 0 \Leftrightarrow f = 0$;

2) $(f, g) = (g, f)$;

3) $(\alpha f + \beta g, h) = \alpha(f, h) + \beta(g, h)$.

Тук разглеждаме само случая, когато (f, g) е реално число.

Строгото определение на хилбертово пространство изисква още за всяко n в H да съществуват n линейно независими елемента /т.е. H да е безкрайномерно пространство/ и "пълнота" на H – свойство, което няма да използваме тук и затова няма да го поясняваме.

Всяко хилбертово пространство може да се нормира, като се въведе норма по следния начин:

$$\|f\| := \sqrt{(f, f)} \ . \tag{1}$$

За да се убедим, че (1) наистина определя норма в H, ще докажем найнапред някои свойства на скаларното произведение.

Неравенство на Коши-Буняковски-Шварц: За всеки два елемента f и g от хилбертовото пространство H е изпълнено неравенството

$$|(f, g)| \leq \sqrt{(f, f)}\sqrt{(g, g)} \ .$$

Равенството се достига тогава и само тогава, когато f и g са линейно зависими.

Доказателство. За всяко реално t имаме

$$(f + tg, f + tg) = (f, f) + 2t(f, g) + t^2(g, g) \geq 0 \ .$$

Но последният израз е полином от втора степен на t. Следователно неговата дискриминанта е неположителна, т.е.

$$[(f, g)]^2 \leq (f, f)\,(g, g) \ .$$

Неравенството е доказано. Ако $f = \alpha g$, то очевидно неравенството става равенство. Вярно е и обратното – ако имаме равенство, то f и g са линейно зависими. Наистина, в противен случай бихме получили, от една страна, $(f - \alpha g, f - \alpha g) > 0$ при всяко α, а от друга, ако означим с ε знака на (f, g),

$$(f - \alpha g, f - \alpha g) = (f, f) - 2\alpha(f, g) + \alpha^2(g, g) = \left(\sqrt{(f, f)} - \varepsilon\alpha\sqrt{(g, g)}\right)^2 = 0$$

при $\alpha = \varepsilon\sqrt{(f,f)}/\sqrt{(g,g)}$. Доказателството е завършено.

Неравенство на триъгълника: За всяко f и g от H имаме

$$\sqrt{(f+g,f+g)} \leq \sqrt{(f,f)} + \sqrt{(g,g)}\ , \tag{2}$$

като равенство се достига тогава и само тогава, когато f и g са линейно зависими.

Доказателство. Като приложим доказаното вече неравенство на Коши-Буняковски-Шварц, получаваме

$$\begin{aligned}(f+g,f+g) &= (f,f)+2(f,g)+(g,g)\\ &\leq (f,f)+2\sqrt{(f,f)(g,g)}+(g,g)\\ &= \left\{\sqrt{(f,f)}+\sqrt{(g,g)}\right\}^2\ ,\end{aligned}$$

откъдето следва (2). Равенството се достига тогава и само тогава, когато $[(f,g)]^2 = (f,f)(g,g)$. Но това е вярно, както вече видяхме, само при линейна зависимост на f и g.

Като се възползваме от означението (1), можем да запишем неравенството (2) във вида

$$\|f+g\| \leq \|f\| + \|g\|\ .$$

Това показва, че въведеното чрез (1) съответствие $f \to \|f\|$ удовлетворява неравенството на триъгълника. Останалите две свойства от определението на норма ($\|f\| > 0$ за $f \neq 0$ и $\|\lambda f\| = |\lambda|\|f\|$) са очевидно също изпълнени. Следователно (1) определя норма в H.

Нормата (1), от своя страна, поражда разстоянието

$$d(f,g) := \|f-g\| = \sqrt{(f-g,f-g)}\ .$$

Оттук нататък, когато говорим за конкретно хилбертово пространство, ще предполагаме, че то е нормирано и метризирано по описания по-горе начин.

Нека $\varphi_0, \varphi_1, \ldots, \varphi_n$ са произволни, фиксирани, линейно независими елементи на H. Да означим

$$\Omega_n := \left\{\ \sum_{i=0}^{n} a_i\varphi_i\ :\ (a_0,\ldots,a_n) \in \mathbb{R}^{n+1}\ \right\}\ .$$

Ще разгледаме задачата за приближаване на елементи f от H с елементи от Ω_n. Най-напред да забележим, че H е *строго нормирано* пространство. Това следва от формулираното по-горе неравенство на триъгълника (2). Следователно, въз основа на общата теорема за приближаване в линейно нормирано пространство:

За всяко f от H, съществува единствен елемент на най-добро приближение от Ω_n.

Остава да разгледаме важния въпрос за построяване на елемента на най-добро приближение. Най-напред ще дадем една негова характеризация.

Казваме, че f е *ортогонален* на g (пишем $f \perp g$), ако $(f, g) = 0$.

Теорема 1. *Нека H е произволно хилбертово пространство и $f \in H$. Елементът p от Ω_n е елемент на най-добро приближение за f с елементи от Ω_n тогава и само тогава, когато*

$$(f - p, \varphi) = 0 \qquad \textit{за всяко } \varphi \textit{ от } \Omega_n \ . \tag{3}$$

Доказателство. Да предположим, че p е елемент на най-добро приближение, т.е.

$$\|f - p\| = \inf \{ \ \|f - \varphi\| \ : \ \varphi \in \Omega_n \ \} =: \varepsilon_n(f) \ .$$

Тогава, за произволно $\varphi \in \Omega_n$ и $\varphi \neq 0$, функцията

$$\begin{aligned} r(\lambda) \ &:= \ \|f - p + \lambda\varphi\|^2 = (f - p + \lambda\varphi, f - p + \lambda\varphi) \\ &= \ \varepsilon_n^2(f) + 2\lambda(f - p, \varphi) + \lambda^2(\varphi, \varphi) \end{aligned}$$

ще има минимум при $\lambda = 0$. Това влече $r'(0) = 0$. Но $r'(0) = 2(f - p, \varphi)$. Следователно $(f - p, \varphi) = 0$ за всяко $\varphi \in \Omega_n$.

Обратно, да допуснем, че $p \in \Omega_n$ и p удовлетворява условията за ортогоналност (3). Нека φ е произволен друг елемент от Ω_n. Тогава $\delta := p - \varphi \in \Omega_n$ и следователно

$$\begin{aligned} \|f - \varphi\|^2 \ &= \ \|f - p + p - \varphi\|^2 = (f - p + \delta, f - p + \delta) \\ &= \ \|f - p\|^2 + 2(f - p, \delta) + \|\delta\|^2 \\ &= \ \|f - p\|^2 + \|\delta\|^2 \qquad \text{(защото } (f - p) \perp \delta) \\ &\geq \ \|f - p\|^2 \ . \end{aligned}$$

И така, ако p удовлетворява (3), то

$$\|f - p\| \leq \|f - \varphi\| \ \text{ за всяко } \varphi \in \Omega_n \ .$$

При това, равенство се достига само при $\delta = 0$, т.е. при $\varphi = p$. Теоремата е доказана.

Сега ще построим елемента на най-добро приближение за f, като използваме характеризацията (3). Ще търсим p във вида

$$p = a_0\varphi_0 + a_1\varphi_1 + \ldots + a_n\varphi_n \ .$$

Тъй като $f-p \perp \varphi_i$ за $i = 0, 1, \ldots, n$, то коефициентите $\{a_i\}$ удовлетворяват условията:

$$
\begin{aligned}
&a_0(\varphi_0,\varphi_0)+a_1(\varphi_1,\varphi_0)+\ldots+a_n(\varphi_n,\varphi_0)=(f,\varphi_0)\\
(4)\qquad &a_0(\varphi_0,\varphi_1)+a_1(\varphi_1,\varphi_1)+\ldots+a_n(\varphi_n,\varphi_1)=(f,\varphi_1)\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&a_0(\varphi_0,\varphi_n)+a_1(\varphi_1,\varphi_n)+\ldots+a_n(\varphi_n,\varphi_n)=(f,\varphi_n)\ .
\end{aligned}
$$

Но това е линейна система от $n+1$ уравненея с $n+1$ неизвестни. Да означим с $D(\varphi_0, \ldots, \varphi_n)$ нейната детерминанта,

$$
D(\varphi_0,\ldots,\varphi_n) := \det \begin{bmatrix} (\varphi_0,\varphi_0) & (\varphi_1,\varphi_0) & \cdots & (\varphi_n,\varphi_0) \\ (\varphi_0,\varphi_1) & (\varphi_1,\varphi_1) & \cdots & (\varphi_n,\varphi_1) \\ \vdots & \vdots & \ddots & \vdots \\ (\varphi_0,\varphi_n) & (\varphi_1,\varphi_n) & \cdots & (\varphi_n,\varphi_n) \end{bmatrix} .
$$

Това е детерминантата на Грам. Тя е различна от нула, защото $\varphi_0, \ldots, \varphi_n$ бяха избрани линейно независими. Следователно системата (4) има единствено решение $a_0, \ldots, a_n$. И така, намирането на елемента на най-добро приближение в едно хилбертово пространство H се свежда към решаване на линейната система (4).

Работата по решаване на системата (4) може да се облекчи значително, ако сме избрали подходящ базис $\varphi_0, \ldots, \varphi_n$. Известно е например, че във всяко линейно пространство съществува ортогонален базис. Да предположим сега, че $\varphi_0, \ldots, \varphi_n$ са ортогонална система, т.е. $(\varphi_i, \varphi_j) = 0$ за $i \neq j$. Тогава (4) добива вида

$$
a_k(\varphi_k,\varphi_k) = (f,\varphi_k)\ , \qquad k = 0,\ldots,n,
$$

и оттук получаваме

$$
(5)\qquad a_k = \frac{(f,\varphi_k)}{(\varphi_k,\varphi_k)}\ , \qquad k=0,\ldots,n\ .
$$

И така, доказахме следното твърдение.

Теорема 2. *Нека $\varphi_0, \ldots, \varphi_n$ е ортогонална система. Тогава елементът p на най-добро приближение на $f \in H$ с елементи от Ω_n се дава с формулата*

$$
p = \sum_{k=0}^{n} \frac{(f,\varphi_k)}{(\varphi_k,\varphi_k)}\, \varphi_k\ .
$$

Да намерим сега израз за грешката $\varepsilon_n(f) = \|f-p\|$. Имаме

$$
\varepsilon_n^2(f) = (f-p, f-p) = (f,f) - (p,f) \qquad (\text{защото } f-p \perp p)\ .
$$

Оттук, като представим p във вида $p = a_0\varphi_0 + \ldots + a_n\varphi_n$, получаваме връзката

$$a_0(\varphi_o, f) + a_1(\varphi_1, f) + \ldots + a_n(\varphi_n, f) = (f, f) - \varepsilon_n^2(f) \ .$$

Добавяме към това уравнение системата (4) и ги разглеждаме като хомогенна система от $n+2$ уравнения по отношение на $(a_0, a_1, \ldots, a_n, 1)$. Тъй като тя има ненулево решение, нейната детерменанта е нула, т.е.

$$\det \begin{bmatrix} (\varphi_0, \varphi_0) & \ldots & (\varphi_n, \varphi_0) & (f, \varphi_0) \\ \vdots & \ddots & \vdots & \vdots \\ (\varphi_0, \varphi_n) & \ldots & (\varphi_n, \varphi_n) & (f, \varphi_n) \\ (\varphi_0, f) & \ldots & (\varphi_n, f) & (f, f) - \varepsilon_n^2(f) \end{bmatrix} = 0 \ .$$

От това равенство определяме $\varepsilon_n^2(f)$:

(6) $$\varepsilon_n^2(f) = \frac{D(\varphi_0, \ldots, \varphi_n, f)}{D(\varphi_0, \ldots, \varphi_n)} \ .$$

И така, доказахме равенството

$$\min_{\{a_k\}_0^n} \left\| f - \sum_{k=0}^{n} a_k\varphi_k \right\|^2 = \frac{D(\varphi_0, \ldots, \varphi_n, f)}{D(\varphi_0, \ldots, \varphi_n)} \ .$$

Тази формула е в сила за произволен базис $\varphi_0, \ldots, \varphi_n$. Ако $\varphi_0, \ldots, \varphi_n$ е *ортонормирана система*, т.е. ако $(\varphi_i, \varphi_j) = 0$ за $i \neq j$ и $(\varphi_i, \varphi_i) = 1$ за $i = 0, \ldots n$, получаваме директно

$$\begin{aligned} \varepsilon_n^2(f) &= (f - p, f - p) = (f, f) - (p, f) \\ &= (f, f) - \sum_{k=0}^{n} a_k(\varphi_k, f) \\ &= \|f\|^2 - \sum_{k=0}^{n} a_k^2 \quad \text{(защото, съгласно (5), } a_k = (\varphi_k, f)) \ . \end{aligned}$$

Тъй като $\varepsilon_n^2(f) > 0$ за $f \notin \Omega_n$, то оттук следва известното *неравенство на Бесел:*

$$\left(\sum_{k=0}^{n} a_k^2 \right)^{1/2} \leq \|f\| \ .$$

Бележка. От (6), между другото, следва по индукция, като се вземе пред вид, че $D(g_1) = (g_1, g_1) > 0$ за всяко $g_1 \neq 0$, че детерминантата на Грам $D(g_1, \ldots, g_n)$ е строго положителна, ако елементите $g_1, \ldots, g_n$ са линейно независими.

Частни случаи

I. Средноквадратични приближения.

Нека $[a,b]$ е даден интервал (краен или безкраен). Нека $\mu(x)$ е интегруема теглова функция в $[a,b]$. Да означим с $L_2[a,b]$ пространството от всички функции, определени в $[a,b]$, за които

$$\int_a^b \mu(x) f^2(x)\, dx < \infty \ .$$

Ясно е, че $L_2[a,b]$ е едно линейно пространство. Да въведем в него скаларно произведение по следния начин

$$(f,g) := \int_a^b \mu(x) f(x) g(x)\, dx \ .$$

Лесно се вижда, че това определение наистина удовлетворява изискванията за скаларно произведение. Така $L_2[a,b]$ става хилбертово пространство. Нормата

$$\|f\| := \left\{ \int_a^b \mu(x) f^2(x)\, dx \right\}^{1/2}$$

се нарича *средноквадратична норма.* Тя поражда *средноквадратичното разстояние*

$$\rho(f,g) := \left\{ \int_a^b \mu(x) \left[f(x) - g(x)\right]^2 \, dx \right\}^{1/2} \ .$$

Нека $\varphi_0(x), \ldots \varphi_n(x)$ са произволни линейно независими функции от пространството $L_2[a,b]$. В частност, $\{\varphi_i\}$ могат да бъдат например алгебричните полиноми $1, x, x^2, \ldots, x^n$. Тогава в $L_2[a,b]$ можем да разглеждаме задачата за *средноквадратично приближение* на дадена функция $f \in L_2[a,b]$ с *обобщени полиноми* $a_0\varphi_0(x) + a_1\varphi_1(x) + \ldots + a_n\varphi_n(x)$.

Съгласно общата теория на приближения в хилбертови пространства, в сила е следната

Теорема 3. *За всяка функция f от $L_2[a,b]$ съществува единствен полином*

$$p(x) = \sum_{k=0}^{n} a_k^* \varphi_k(x),$$

за който

$$\int_a^b \mu(x) \left[f(x) - p(x)\right]^2 \, dx = \min_{\{a_k\}} \int_a^b \mu(x) \left[f(x) - \sum_{k=0}^{n} a_k \varphi_k(x) \right]^2 dx \ .$$

При това, ако $\varphi_0, \ldots, \varphi_n$ е ортонормирана система, то

$$p(x) = \sum_{k=0}^{n} \int_a^b \mu(t) f(t) \varphi_k(t)\, dt \ . \ \varphi_k(x) \ . \tag{7}$$

II. Метод на най-малките квадрати.

На практика често се сблъскваме със следната задача.

От теоретични съображения е известно, че изследваната от нас функция f е от определен вид, зависеща от n параметри $a_1, \ldots, a_n$ (например, $\sum_{k=1}^{n} a_k x^{k-1}$, $\prod_{k=1}^{n} \sin a_k x$, $\sum_{k=1}^{n} e^{a_k x}$ и т.н.). Можем да изчислим стойността на f с определена точност в краен брой точки. При това, намирането на стойността на f в дадена точка, може да бъде свързано с провеждане на скъпо струващ експеримент. Целта е да възстановим приближено параметрите $a_1, \ldots, a_n$ с възможно най-голяма точност въз основа на информацията

$$f(x_1), f(x_2), \ldots, f(x_m) \quad m > n \; .$$

Най-често тези числа са приближения на истинската стойност на функцията f.

Например, нека знаем, че зависимостта $y = f(x)$, която изследваме е линейна, т.е.

$$f(x) = Ax + B \; ,$$

при някакви A и B. Разполагаме с експериментално получени стойности на $f(x)$: $f_i = f(x_i)$, $i = 1, \ldots, m$. Те са представени по-долу на Рис. 7.

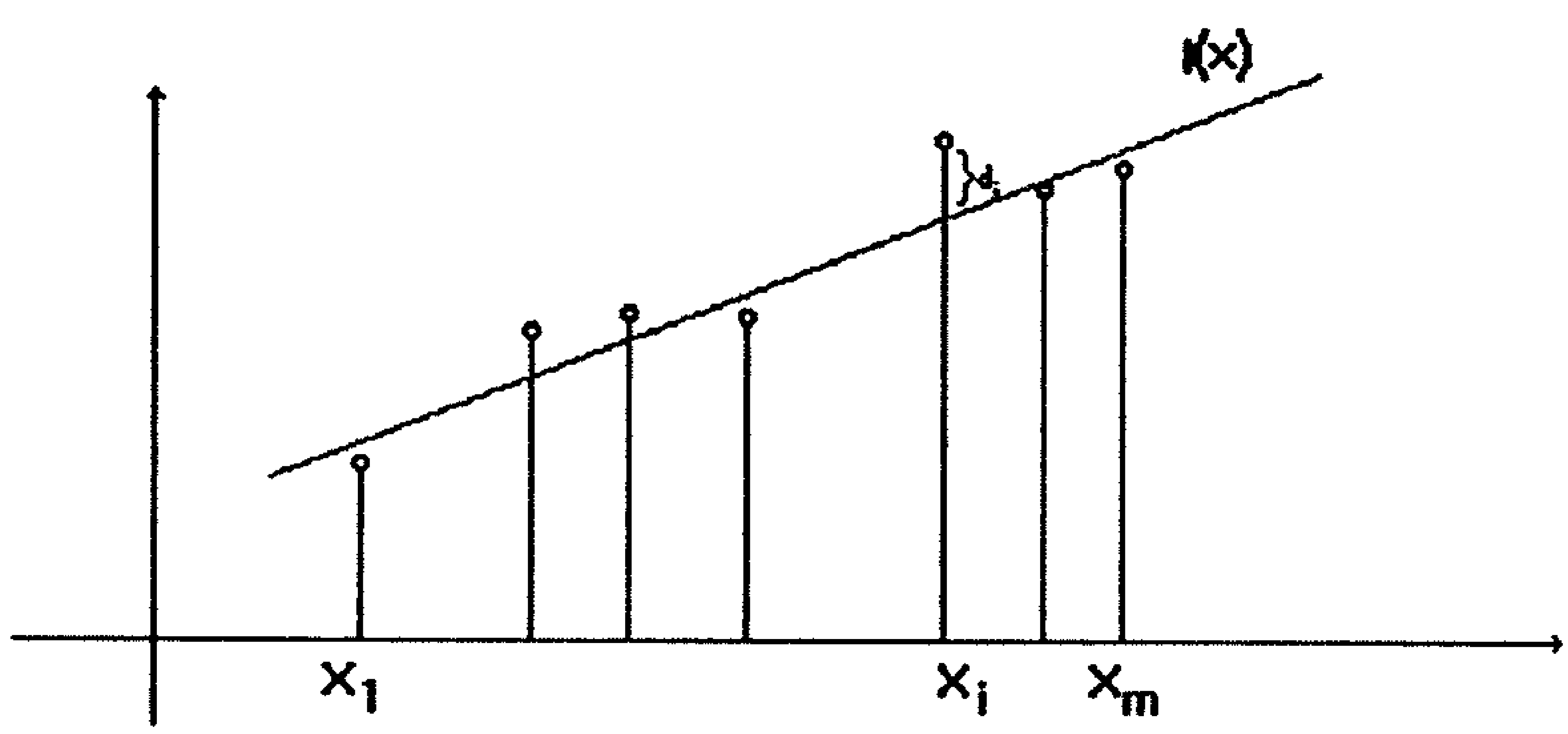

Рис. 7

Поради неточност на измерването или несъвършенство на експеримента, точките (x_i, f_i), $i = 1, \ldots, n$, явно не лежат на една права. Знаем, че функцията $f(x)$ е линейна. Тогава, коя права да вземем за представител на получените данни? Има много кандидати за такъв представител. Например, бихме могли да вземем произволни две точки $(x_i, f_i), (x_j, f_j)$ от таблицата и за приближение на f да вземем правата l през тези две точки. Това е един случаен избор. Да се опитаме да подходим по-теоретично. Търсим функция от вида

$$l(x) = Ax + B \ .$$

Да означим с d_i отклонението на експериментално получената стойност f_i в точката x_i от предлаганата стойност чрез l, т.е.

$$d_i := f_i - (Ax_i + B) \ , \qquad i = 1, \ldots, m \ .$$

Съществуват няколко разумни подхода за избора на параметрите A и B на l.

1) Избираме A и B така, че да минимизират величината

$$\max_{1 \le i \le m} |d_i| \ .$$

По този начин се стараем да направим максималното разстояние между f и l в точките $x_1, \ldots, x_m$ възможно най-малко. Такъв критерий е приемлив, но неговата реализация е трудна, тъй като води до решаване на една нелинейна задача (функцията $\max\limits_i |d_i|$ е нелинейна относно A и B).

2) Избираме A и B така, че да минимизират величината

$$\sum_{i=1}^{m} |d_i| \ .$$

Възраженията срещу критерия 1), остават и в този случай. Те са се възприемали твърде сериозно в миналото, когато хората не са разполагали със средства за бързо смятане. Затова може би изборът е паднал на следващия критерий, който води до линейна система за определяне на параметрите.

3) Избираме A и B така, че да минимизират израза

$$S(A, B) := \sum_{i=1}^{n} d_i^2 \ .$$

В този случай

$$S(A, B) = \sum_{i=1}^{m} [f_i - (Ax_i + B)]^2$$

и необходимите условия за минимум (които са и достатъчни) водят до системата

$$\frac{\partial S}{\partial A} = 0 \quad \Rightarrow \quad \sum_{i=1}^{m} [f_i - (Ax_i + B)] x_i = 0,$$

$$\frac{\partial S}{\partial B} = 0 \quad \Rightarrow \quad \sum_{i=1}^{m} [f_i - (Ax_i + B)] = 0.$$

Този подход за определяне на неизвестни параметри на функцията по таблица от данни се нарича *метод на най-малките квадрати*. Да го представим в една по-обща форма. Нека $\{F(x, a_1, \ldots, a_n)\}$ е фамилия от функции, която се описва от параметрите $a_i \in I_i$, $i = 1, \ldots, n$. Нека $f_1, \ldots, f_m$ са стойности на конкретна функция от тази фамилия.

Определение. Ще казваме, че $F(x, a_1, \ldots, a_n)$ е приближение на данните $f_1, \ldots, f_m$ по метода на най-малките квадрати, ако $a_1, \ldots, a_n$ минимизират израза

$$\sum_{i=1}^{m} \mu_i [F(x_i, a_1, \ldots, a_n) - f_i]^2 ,$$

където $\{\mu_i\}_1^m$ са отнапред зададени положителни числа (наречени *тегла*).

Да разгледаме една конкретна ситуация – приближаване на функции с алгебрични полиноми от степен n в множеството от точки $x_1 < \ldots < x_m$, $(m > n)$. И така, искаме да намерим приближение

$$p(x) = a_0 + a_1 x + \ldots + a_n x^n$$

на f по метода на най-малките квадрати, въз основа на стойностите $f_i = f(x_i)$, $i = 1, \ldots, m$. Нека $\{\mu_i\}$ са дадени тегла. Тогава, съгласно казаното по-горе, $a_0, a_1, \ldots, a_n$ се определят така, че да минимизират израза

$$\Phi(a_0, \ldots, a_n) := \sum_{i=1}^{m} \mu_i \left[f_i - \sum_{k=0}^{n} a_k x_i^k \right]^2 .$$

Вижда се, че $\Phi^{1/2}(a_0, \ldots, a_n)$ е всъщност разстоянието между f и p в хилбертовото пространство H_Δ от функции, определени в $x_1, \ldots, x_m$, снабдено със скаларното произведение

$$(f, g) := \sum_{i=1}^{m} \mu_i f(x_i) g(x_i) .$$

Наистина, това скаларно произведение поражда нормата

$$\|f\| := \left\{ \sum_{i=1}^{m} \mu_i f^2(x_i) \right\}^{1/2} .$$

която, от своя страна, поражда разстоянието

$$\rho(f, g) = \left\{ \sum_{i=1}^{m} \mu_i [f(x_i) - g(x_i)]^2 \right\}^{\frac{1}{2}} .$$

В тези термини, нашата функция $\{\Phi(a_0, \ldots, a_n)\}^{1/2}$ е точно равна на разстоянието между f и p. Следователно методът на най-малките квадрати води до задача за най-добро приближение с алгебрични полиноми в хилбертовото пространство H_Δ. От общата теория следва, че решението $a_0, \dots, a_n$ се определя от линейната система (4), която в този случай приема вида

$$\sum_{i=1}^{m} \mu_i [a_0 x_i^k + a_1 x_i^{k+1} + \cdots + a_n x_i^{k+n}] = \sum_{i=1}^{m} \mu_i f(x_i)\, x_i^k$$

$$(k = 0, \dots, n).$$

За да избегнем решаването на тази система можем да изберем предварително подходящ базис в пространството π_n от алгебрични полиноми. Например, ако търсехме полинома p във вида

$$p(x) = b_0 P_0(x) + \cdots + b_n P_n(x),$$

където полиномите $\{P_k(x)\}$ образуват ортогонална система в множеството от точки $x_1, \dots x_m$ с тегла $\{\mu_i\}$, то горната система щеше да се редуцира до една диагонална система

$$b_k \sum_{i=1}^{m} \mu_i P_k^2(x_i) = \sum_{i=1}^{m} \mu_i P_k(x_i) f(x_i),$$

откъдето коефициентите b_k се определят веднага.

Глава 2

ЧИСЛЕНО ДИФЕРЕНЦИРАНЕ И ИНТЕГРИРАНЕ

Приближеното пресмятане на производна или определен интеграл от дадена функция е важен елемент в числените методи. Ние ще се запознаем тук с някои класически методи, които се основават на идеята, че една приближена формула е добра, когато тя е точна за алгебричните полиноми от възможно най-широк клас. Формулите, които ще изведем, се получават като вместо функцията се диференцира или интегрира съответния интерполационен полином.

16. ЧИСЛЕНО ДИФЕРЕНЦИРАНЕ

Тук ще се занимаем с въпроса за численото диференциране (т.е. с приближеното пресмятане на производната $f'(x)$. Най-напред да обърнем внимание, че диференцирането е един неустойчив процес в смисъл, че малки изменения на функцията f могат да доведат до големи изменения на нейната производна. Това налага да се подхожда много внимателно при численото диференциране, като се анализира детайлно всеки конкретен случай. Ние ще се спрем още веднъж по-подробно на този въпрос по-нататък.

Нека $f(x)$ е определена в $[a, b]$ и $x_0, \ldots, x_n$ са различни точки от $[a, b]$. Да предположим, че $f(x)$ има непрекъснати производни от достатъчно висок ред. Съгласно формулата на Нютон,

$$f(x) = L_n(f; x) + f[x_0, \ldots, x_n, x]\omega(x), \tag{1}$$

където

$$\omega(x) = (x - x_0) \ldots (x - x_n),$$

а $L_n(f; x)$ е интерполационният полином за f с възли $x_0, \ldots, x_n$. Най-напред ще покажем, че функцията $g(x) = f[x_0, \ldots, x_n, x]$ е диференцируема в точката x. Наистина, съгласно определението на производна,

$$\begin{aligned} g'(x) &= \lim_{h \to 0} \frac{g(x + h) - g(x)}{h} \\ &= \lim_{h \to 0} \frac{f[x_0, \ldots, x_n, x + h] - f[x_0, \ldots, x_n, x]}{x + h - x} \\ &= \lim_{h \to 0} f[x_0, \ldots, x_n, x + h, x] \\ &= f[x_0, \ldots, x_n, x, x], \end{aligned}$$

защото, както видяхме по-рано (виж Теорема 6.5) разделената разлика е непрекъсната функция на своите аргументи, ако f е достатъчно гладка. И така,

$$\frac{d}{dx}f[x_0,\dots,x_n,x]=f[x_0,\dots,x_n,x,x].$$

Тогава от (1) получаваме

$$f'(x)=L_n'(f;x)+f[x_0,\dots,x_n,x,x]\omega(x)+f[x_0,\dots,x_n,x]\omega'(x).$$

Следователно грешката $E(f)$ при приближението $f'(x)\approx L_n'(f;x)$ се дава с израза

$$E(f)=f[x_0,\dots,x_n,x,x]\omega(x)+f[x_0,\dots,x_n,x]\omega'(x).$$

Като използваме връзката

$$f[y_0,\dots,y_k]=\frac{f^{(k)}(\xi)}{k!}$$

можем да запишем $E(f)$ по следния начин:

$$E(f)=\frac{f^{(n+2)}(\xi)}{(n+2)!}\,\omega(x)+\frac{f^{(n+1)}(\eta)}{(n+1)!}\,\omega'(x), \tag{2}$$

където ξ и η са някакви точки от интервала (a,b). В общия случай ние много рядко знаем $f^{(n+1)}$ и $f^{(n+2)}$ и почти никога не знаем нищо повече за ξ и η, освен това, че са от (a,b). Затова на практика се използва доста грубата оценка

$$|E(f)|\le\frac{M_{n+2}}{(n+2)!}|\omega(x)|+\frac{M_{n+1}}{(n+1)!}|\omega'(x)|,$$

където M_k е горната граница на $|f^{(k)}(t)|$ в $[a,b]$.

В някои случаи изразът за грешката (2) може да се опрости значително, например, когато точката x съвпада с някой от възлите $x_0,\dots,x_n$, или пък когато $\omega'(x)=0$. В първия случай, при $x=x_k$, имаме $\omega(x_k)=0$ и

$$\omega'(x_k)=\prod_{i=0,i\ne k}^{n}(x_k-x_i).$$

Тогава (2) добива вида

$$E(f)=\frac{f^{(n+1)}(\eta)}{(n+1)!}\prod_{i=0,i\ne k}^{n}(x_k-x_i) \tag{3}$$

за някакво $\eta\in(a,b)$.

Аналогично, ако $\omega'(x)=0$, то (2) добива вида

$$E(f)=\frac{f^{(n+2)}(\xi)}{(n+2)!}\,\omega(x). \tag{4}$$

Ще имаме $\omega'(x) = 0$ например, ако възлите са четен брой и са симетрично разположени около точката x.

Сега да разгледаме някои прости частни случаи.

Нека $n = 1$. Да изберем за възли точките $x_0 = a$ и $x_1 = a+h$. Да намерим приближен израз за $f'(x)$ при $x = a$. Имаме

$$f'(a) \approx L_1'(f;a),$$

където $L_1(f;t) = f(a) + f[a, a+h](t-a)$. Следователно

$$f'(a) \approx f[a, a+h] = \frac{f(a+h) - f(a)}{h} \ . \tag{5}$$

В този случай точката a е възел, затова прилагаме формулата (3) за оценка на грешката. Получаваме

$$E(f) = -\frac{f''(\eta)}{2}h \ . \tag{6}$$

Формулата (5) има прост геометричен смисъл. Производната $f'(a)$, която е равна на ъгловия коефициент на допирателната към $f(x)$ в точката с абсциса a, се заменя с ъгловия коефициент на секущата през точките с абсциси a и $a+h$ (виж Рис. 8).

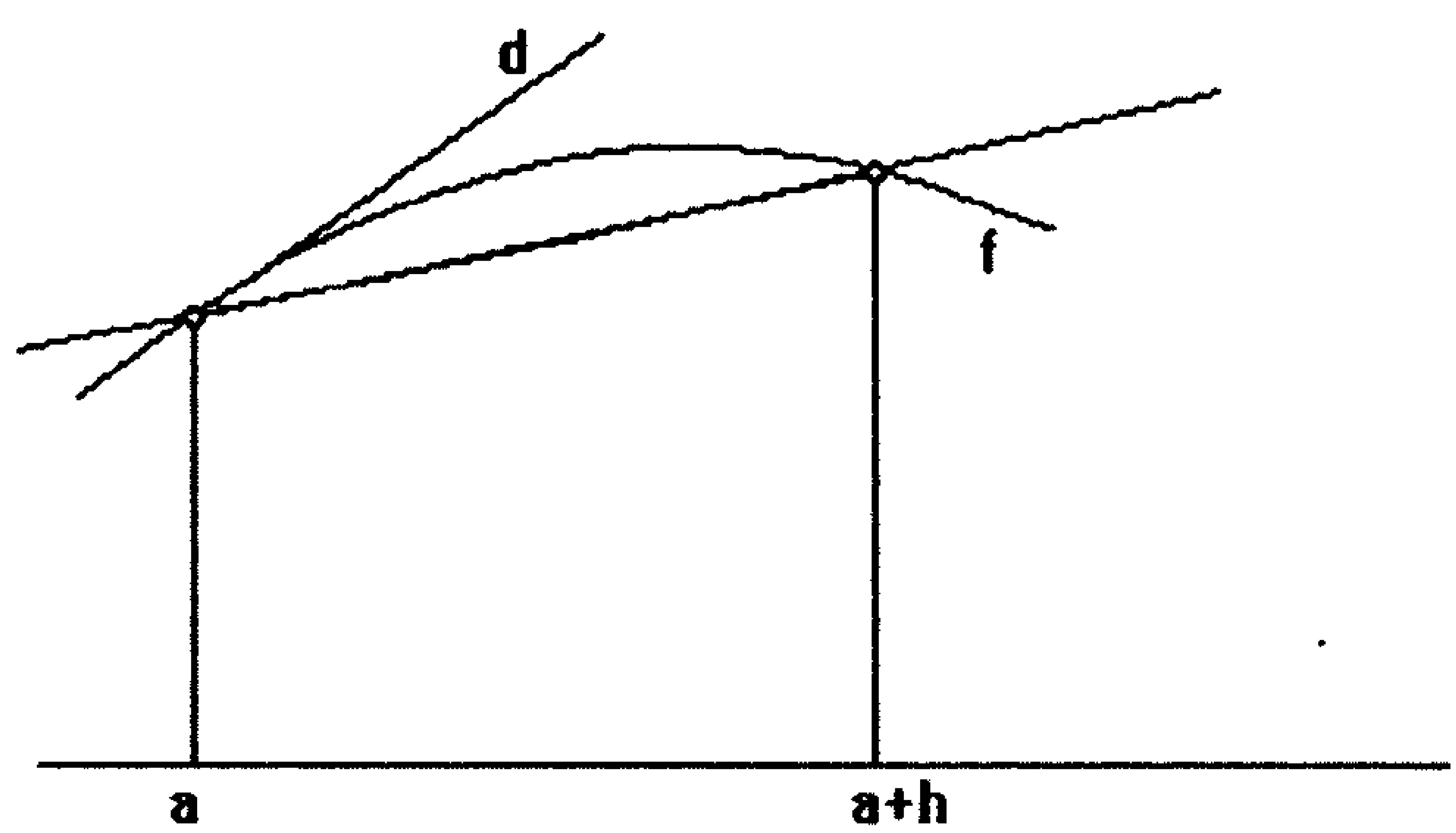

Рис. 8

Нека отново $n = 1$ и възлите x_0 и x_1 са симетрично разположени около точката a, в която търсим производната. Да положим $x_0 = a-h$, $x_1 = a+h$. Очевидно

$$L_1(f;t) = f(a-h) + f[a-h, a+h](t-a+h).$$

Следователно $f'(a) \approx L_1'(f;a) = f[a-h, a+h]$. Получихме формулата

$$f'(a) \approx \frac{f(a+h) - f(a-h)}{2h} \ . \tag{7}$$

Геометричният смисъл на (7) е изяснен с Рис. 9 – ъгловият коефициент на допирателната към $f(x)$ в точката a се приближава с ъгловия коефициент на секущата през точките $a-h$ и $a+h$.

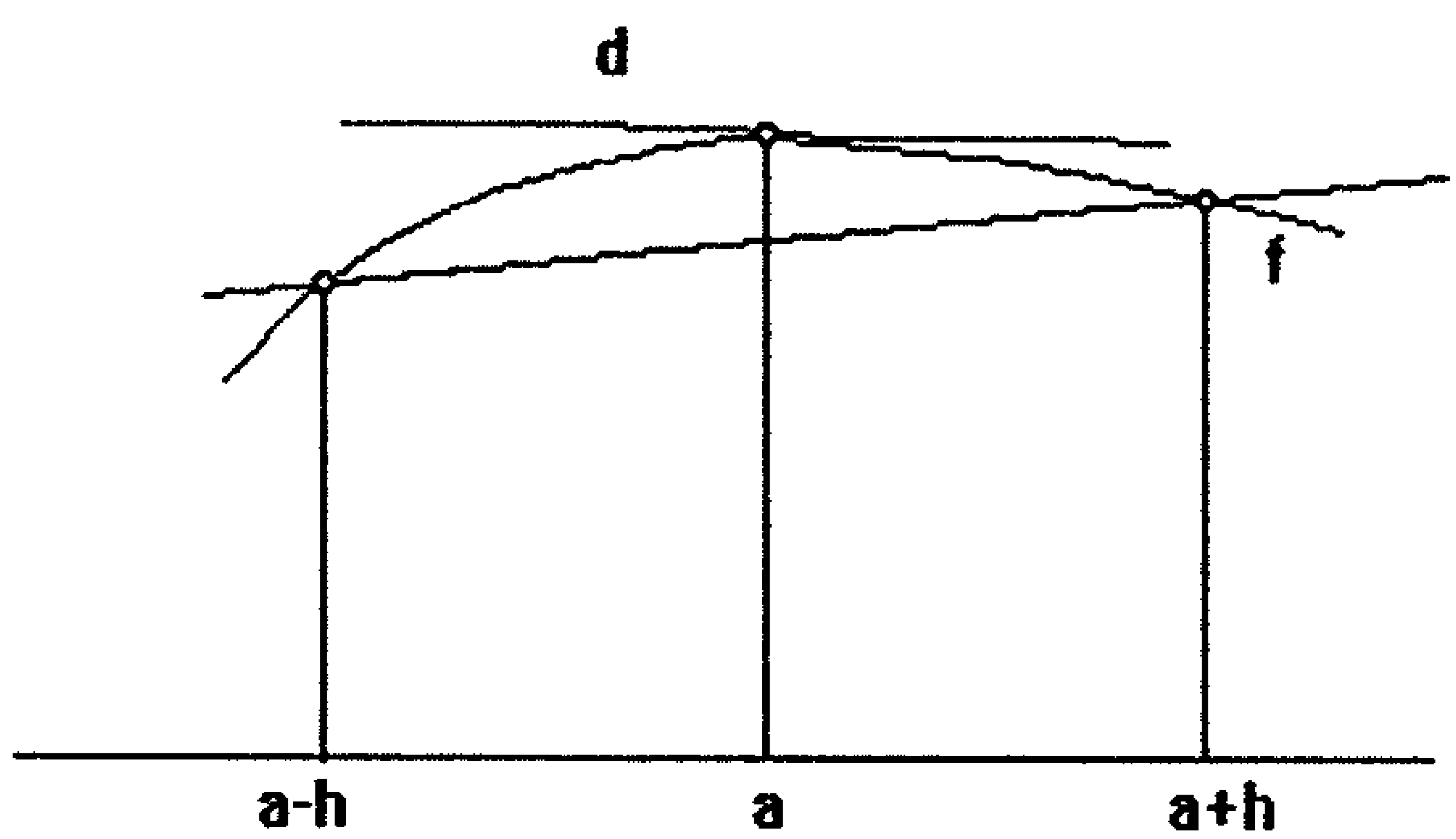

Рис. 9

Грешката $E(f)$ намираме, като използваме (4),

$$E(f) = -\frac{f'''(\xi)}{6}h^2 \ . \tag{8}$$

Веднага прави впечатление, че грешката (8) е много по-малка при малки h от грешката (6), докато съответните формули (5) и (7) са "еднакво сложни" – и двете използват по две стойности на функцията $f(x)$. За характеризация на порядъка на грешката (както и на други величини в числения анализ) се използват символите O ("о" голямо) и o ("о" малко). Казваме, че $\varphi(h)$ е $O(\psi(h))$ при $h \to 0$, ако съществува константа K такава, че $\frac{\varphi(h)}{\psi(h)} \leq K$ при $h \to 0$. Казваме, че $\varphi(h)$ е $o(\psi(h))$ при $h \to 0$, ако $\frac{\varphi(h)}{\psi(h)} \to 0$ при $h \to 0$. И така, съгласно дадените определения, формулата (5) има грешка от порядъка $O(h)$, докато грешката на (7) е $O(h^2)$. Както ще се убедим по-нататък, формулата (7) се използва много често при теоретични изследвания, особено при обосновка на числените методи за решаване на диференциални уравнения. За съжаление тя може да бъде приложена само за приближаване на производната във вътрешните точки $x_1, \dots x_n$ на дадена таблица от стойности $f(x_0), \dots, f(x_n)$ на функцията $f(x)$. За крайните точки x_0 и x_n може да се използва (5). Но тя има грешка $O(h)$. Би било добре да има формула за приближено смятане на $f'(x_0)$ и $f'(x_n)$ с порядък на грешката $O(h^2)$. Именно такава формула ще изведем сега. За целта ще привлечем още един възел, за да увеличим точността на приближение.

Нека $n = 2$. Избираме възли $x_0 = a$, $x_1 = a + h$, $x_2 = a + 2h$. Ще приближаваме производната на $f(x)$ в точката $x = a$. В този случай имаме

$$L_2(f;x) = f(a) + f[a, a+h](x-a) + f[a, a+h, a+2h](x-a)(x-a-h) \ .$$

Оттук получаваме

$$\begin{aligned} L_2'(f;a) &= f[a,a+h] + f[a,a+h,a+2h](-h) \\ &= \frac{f(a)}{-h} + \frac{f(a+h)}{h} - h\left\{\frac{f(a)}{2h^2} + \frac{f(a+h)}{-h^2} + \frac{f(a+2h)}{2h^2}\right\} \\ &= \frac{-3f(a)+4f(a+h)-f(a+2h)}{2h}\,. \end{aligned}$$

Следователно

$$f'(a) \approx \frac{-3f(a)+4f(a+h)-f(a+2h)}{2h}\,. \tag{9}$$

Тъй като точката a е възел, то за оценка на грешката $E(f)$ ще приложим формулата (3). Получаваме

$$E(f) = \frac{f'''(\eta)}{3}h^2\,. \tag{10}$$

Грешката има порядък $O(h^2)$.

Ако построим, по описания вече начин, формула за приближение на $f'(a)$ при възли $x_0 = a - h$, $x_1 = a$, $x_2 = a + h$, то ще видим, че получената формула съвпада с (7), т.е. коефициентът пред $f(a)$ в приближения израз ще бъде равен на нула. Това разкрива причината за по-голямата точност на (7) в сравнение с (5) – тази формула е построена всъщност по три (а не по две) стойности на изследваната функция $f(x)$.

Интерполационният полином на Лагранж $L_n(f;x)$ с възли $x_0,\dots,x_n$ се използва също за приближено пресмятане на производни от по-висок ред. Просто $f^{(k)}(x)$ се заменя $L_n^{(k)}(f;x)$. За оценка на грешката обаче се налага да се диференцира функцията $f[x_0,\dots,x_k,x]\omega(x)$ k пъти, след което се получават изрази от вида (2) с $k+1$ члена. Ние няма да извеждаме тук тези изрази. Накрая ще обърнем внимание, че връзката между производната и разделената разлика ни дава още следната формула за числено диференциране: За $x \in [x_0,x_n]$,

$$f^{(n)}(x) \approx f[x_0,\dots,x_n]n!\,. \tag{11}$$

Както се вижда, формулата (5) се получава от (11) при $n = 1$ и $x = x_0$.

От изведените изрази за грешката при численото диференциране се вижда, че грешката намалява, когато стъпката h намалява. На пръв поглед изглежда, че човек може да получи производната $f'(a)$ с каквато си иска точност, стига да може да се изчисли $f(x)$ в близки до a точки x. На практика се оказва, че това не е така. Забелязва се, че при практическо използване на коя да е от изведените формули, при намаляване на h грешката отначало намалява, а след това започва да расте. Причината е в "неустойчивостта" на формулите за числено диференциране, за която споменахме в началото на тази лекция. Тук ще обясним по-подробно това явление.

Нека приближаваме $f'(a)$ по формулата (7),

$$f'(a) \approx \frac{f(a+h) - f(a-h)}{2h} .$$

Да приемем, че електронната сметачна машина, с която работим, представя числата с точност 10^{-8}. Тогава вместо с точните стойности $f(a+h)$ и $f(a-h)$ ние работим в машината с числата

$$\begin{aligned} \tilde{f}(a+h) &= f(a+h) + \varepsilon_1, \\ \tilde{f}(a-h) &= f(a-h) + \varepsilon_2, \end{aligned}$$

където

$$(12) \qquad |\varepsilon_i| \le 10^{-8}, \quad i = 1, 2.$$

Тогава за приближената стойност на $f'(a)$ получаваме числото

$$\frac{\tilde{f}(a+h) - \tilde{f}(a-h)}{2h} = \frac{f(a+h) - f(a-h)}{2h} + \frac{\varepsilon_1 - \varepsilon_2}{2h}.$$

Съгласно (8),

$$\frac{f(a+h) - f(a-h)}{2h} = f'(a) + E,$$

където

$$(13) \qquad |E| \le Mh^2 ,$$

с някаква константа M. Следователно изразът $\frac{\tilde{f}(a+h) - \tilde{f}(a-h)}{2h}$ приближава $f'(a)$ с грешка $E + (\varepsilon_1 - \varepsilon_2)/2h$. Тази грешка има порядък $\varphi(h) = Mh^2 + \frac{2.10^{-8}}{2h}$, поради (12) и (13). Тъй като $\varphi'(h) = 2Mh - \frac{10^{-8}}{h^2}$, то $\varphi(h)$ ще достига своя минимум при $h = h_0$, където h_0 е нулата на $\varphi'(h)$,

$$h_0 = \sqrt[3]{\frac{1}{2.10^8 M}} = \frac{1}{10^3} \sqrt[3]{\frac{5}{M}} .$$

Следователно $\varphi(h)$ намалява, когато h намалява до h_0, а след това започва да расте при по-нататъшно намаляване на стъпката h. Затова при числено приложение се налага за всеки конкретен случай да се установи тази критична стойност h_0 на стъпката и да се използват само такива h, за които $h > h_0$.

Да изведем формула за приближение на $f''(a)$ въз основа на стойностите $f(a-h)$, $f(a)$, $f(a+h)$ при предположение, че f има непрекъсната четвърта производна в $[a-h, a+h]$.

Първи начин. Да означим с $L_2(f;x)$ интерполационния полином на Лагранж за функцията $f(x)$ с възли $a-h$, a, $a+h$. Съгласно формулата на Нютон

$$L_2(f;x) = f(a-h) + f[a-h,a](x-a+h) + f[a-h,a,a+h](x-a+h)(x-a)$$

и

$$f(x) = L_2(f;x) + f[a-h,a,a+h,x](x-a+h)(x-a)(x-a-h)$$

Оттук можем да получим едно приближение за $f''(a)$ по следния начин:

$$\begin{aligned} f''(a) \approx L_2''(f;a) &= 2f[a-h,a,a+h] \\ &= \frac{f(a-h)-2f(a)+f(a+h)}{h^2}. \end{aligned}$$

При това приближение допускаме грешка $E(f)$,

$$\begin{aligned} E(f) &= f''(a) - L_2''(f;a) \\ &= \{f[a-h,a,a+h,x](x-a+h)(x-a)(x-a-h)\}''|_{x=a} \\ &= 2f[a-h,a,a+h,a,a]\omega'(a) + f[a-h,a,a+h,a]\omega''(a) \\ &= -\frac{f^{IV}(\xi)}{12}h^2 \quad (\text{ защото } \omega''(a)=0). \end{aligned}$$

Втори начин. Тук предлагаме още един начин, за да се запознаем с "метода на неопределените коефициенти" за построяване на формули за приближение на линейни функционали. Ще демонстрираме този метод на разгледания по-горе пример.

Развиваме в ред на Тейлър около точката a стойностите $f(a-h)$, $f(a)$, $f(a+h)$. Получаваме

$$\begin{aligned} f(a-h) &= f(a) - f'(a)h + \frac{f''(a)}{2!}h^2 - \frac{f'''(a)}{3!}h^3 + \frac{f^{IV}(\xi_1)}{4!}h^4 \\ f(a) &= f(a) \qquad\qquad (14) \\ f(a+h) &= f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \frac{f'''(a)}{3!}h^3 + \frac{f^{IV}(\xi_2)}{4!}h^4, \end{aligned}$$

където ξ_1 и ξ_2 са някакви точки от интервалите $(a-h,a)$ и $(a,a+h)$, съответно. Целта ни е да намерим коефициенти α, β и γ такива, че изразът

$$\alpha f(a-h) + \beta f(a) + \gamma f(a+h)$$

да бъде равен на $f''(a) + O(h^k)$, където грешката $O(h^k)$ да бъде възможно най-малка, т.е. показателят k да бъде по възможност най-голям. Да сумираме равенствата (14) умножени с α, β и γ, съответно. Получаваме

$$\begin{aligned}
\alpha f(a-h) + \beta f(a) + \gamma f(a+h) &= (\alpha+\beta+\gamma) f(a) + (-\alpha+\gamma) f'(a) h \\
&+ (\alpha+\gamma)\frac{f''(a)}{2} h^2 + (-\alpha+\gamma)\frac{f'''(a)}{6} h^3 \\
&+ \left[\alpha f^{IV}(\xi_1) + \gamma f^{IV}(\xi_2)\right] \frac{h^4}{24} \,.
\end{aligned}$$

Целта ни е вдясно да получим това, което искаме да приближаваме (в този конкретен случай $f''(a)$) и след това да се опитаме чрез специален избор на параметрите α, β и γ да анулираме коефициентите пред най-ниските степени на h, пред h^0, h^1, h^2 и т.н., доколкото можем. Значи в нашия случай трябва да изберем α, β и γ от условията

$$\begin{aligned}
(\alpha+\gamma)\frac{h^2}{2} &= 1 \\
(\alpha+\beta+\gamma) &= 0 \\
-\alpha+\gamma &= 0 \,.
\end{aligned}$$

Това са три линейни уравнения с три неизвестни. Решаваме системата и получаваме

$$\alpha = \gamma = \frac{1}{h^2}, \quad \beta = -\frac{2}{h^2} \,.$$

Веднага се вижда, че при този избор на параметрите α, β и γ се анулира и коефициентът пред $h^3 : (-\alpha+\gamma)\frac{f'''(a)}{6}$.

Следователно

$$f''(a) = \frac{1}{h^2} f(a-h) - \frac{2}{h^2} f(a) + \frac{1}{h^2} f(a+h) + E(f),$$

където

$$E(f) = -\frac{f^{IV}(\xi_1) + f^{IV}(\xi_2)}{2} \frac{h^2}{12} \,.$$

Тъй като $f^{IV}(t)$ е непрекъсната функция и числото $\frac{f^{IV}(\xi_1)+f^{IV}(\xi_2)}{2}$ се намира между точната долна и точната горна граница на $f^{IV}(t)$, то съществува точка $\xi \in (a-h, a+h)$ такава, че $\frac{f^{IV}(\xi_1)+f^{IV}(\xi_2)}{2} = f^{IV}(\xi)$. Следователно

$$E(f) = -\frac{f^{IV}(\xi)}{12} h^2 \,.$$

17. ИНТЕРПОЛАЦИОННИ КВАДРАТУРНИ ФОРМУЛИ

Определеният интеграл е основно математическо понятие. Редица величини в естествените науки, инженерното дело, икономиката и други области на приложение на математиката се изразяват чрез определени интеграли. Затова много често на практика възниква необходимостта от числено пресмятане на определени интеграли. Известно е обаче от курса по анализ, че стойността на един определен интеграл $I(f) := \int_a^b f(x)\,dx$ може да бъде пресметната точно в много редки случаи – когато подинтегралната функция е достатъчно проста. В общия случай числото $I(f)$, което се определя като граница на числова редица, е недостъпно за математика, въоръжен с молив, лист и курса по анализ. Съществуват обаче редица числени методи, които позволяват определеният интеграл да бъде пресметнат с дадена точност. Тук ще разгледаме някои от тези методи.

Просто правило за приближено пресмятане на интеграли може да се получи като заменим подинтегралната функция $f(x)$ с интерполационния полином на Лагранж. Да предположим, че са известни стойностите на $f(x)$ в точките $x_0, \dots, x_n$. Съгласно формулата на Нютон

$$f(x) = L_n(f;x) + f[x_0, \dots, x_n, x]\omega(x) \ , \tag{1}$$

където $L_n(f;x)$ е интерполационният полином на Лагранж,

$$L_n(f;x) = \sum_{k=0}^{n} f(x_k) l_k(x),$$

а $\omega(x) = (x - x_0)\dots(x - x_n)$. Като интегрираме (1) почленно от a до b получаваме формулата

$$I(f) \approx \sum_{k=0}^{n} c_k f(x_k) \ , \tag{2}$$

където

$$c_k = I(l_k) = \int_a^b \prod_{i=0, i\neq k}^{n} \frac{x - x_i}{x_k - x_i}\,dx, \quad k = 0, \dots, n \ . \tag{3}$$

Грешката на това приближение е

$$R(f) := I(f) - I(L_n(f)) = \int_a^b f[x_0, \dots, x_n, x]\omega(x)\,dx \ . \tag{4}$$

Формула за приближение, в която определеният интеграл се приближава с линейна комбинация на стойностите на подинтегралната функция или нейни производни в краен брой точки, се нарича *квадратурна формула*. (2) е една квадратурна формула. Точките $x_0, \ldots, x_n$ са *възли*, а числата $c_0, \ldots, c_n$ – *коефициенти* на квадратурната формула.

Определение. Една квадратурна формула от вида (2) се нарича *интерполационна*, ако нейните коефициенти c_k се получават по формулата (3).

С други думи, формула с $n+1$ възела се нарича интерполационна квадратурна формула (или формула от интерполационен тип), ако тя се получава чрез интегриране на интерполационния полином на Лагранж със същите възли.

Ще казваме,че квадратурната формула (2) е точна за функцията f, ако $R(f) = 0$.

Теорема 1. *Ако квадратурната формула (2) е интерполационна, то тя е точна за всеки полином от класа π_n. Обратно, ако една формула от вида (2) е точна за всички полиноми от класа π_n, то тя е интерполационна.*

Доказателство. Нека (2) е интерполационна квадратурна формула и $f \in \pi_n$. Тогава $f(x) \equiv L_n(f;x)$ и следователно $R(f) = 0$, т.е. формулата е точна за f.

Да допуснем, че (2) е точна за всички полиноми f от π_n. Тогава тя ще бъде точна и за полиномите $l_i(x)$. Следователно

$$I(l_i) = \sum_{k=0}^{n} c_k l_i(x_k) = c_i, \quad i = 0, \ldots n,$$

защото $l_i(x_k) = \delta_{ik}$. Получихме (3). Теоремата е доказана.

Както вече видяхме, грешката на интерполационната квадратурна формула се дава с израза (4). Тази формула не е удобна за приложение, тъй като грешката се изразява отново чрез интеграл, който дори е по-сложен от предишния. Въз основа на (4) обаче се правят оценки, които се използват на практика. Ще разгледаме два случая, в които изразът за грешката може да се запише в по-прост вид.

Нека полиномът $\omega(x)$ не си сменя знака в (a, b). Да предположим още, че $f(x)$ има непрекъсната $(n+1)$-ва производна в $[a, b]$. Тогава $f[x_0, \ldots, x_n, x]$ е непрекъсната функция на x в $[a, b]$ и съгласно теоремата за средните стойности, съществува точка $t \in (a, b)$ такава, че

$$R(f) = f[x_0, \ldots, x_n, t] \int_a^b \omega(x)\, dx.$$

По-нататък, от връзката между разделената разлика и производната следва, че съществува точка $\xi \in [a, b]$, за която

$$R(f) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \int_a^b \omega(x)\, dx \;. \tag{5}$$

Ако $\omega(x)$ си сменя знака само един път в $[a, b]$ и $\int_a^b \omega(x)\,dx = 0$, то изразът (4) също може да се опрости. В този случай използваме рекурентната връзка

$$f[x_0, \dots, x_n, x_{n+1}, x] = \frac{f[x_0, \dots, x_n, x] - f[x_0, \dots, x_n, x_{n+1}]}{x - x_{n+1}},$$

за да получим

$$f[x_0, \dots, x_n, x] = f[x_0, \dots, x_n, , x_{n+1}, x](x - x_{n+1}) + f[x_0, \dots, x_{n+1}]$$

за всяка точка x_{n+1} от $[a, b]$. Следователно

$$\begin{aligned} R(f) &= \int_a^b f[x_0, \dots, x_n, x_{n+1}, x](x - x_{n+1})\omega(x)\,dx + \\ &+ f[x_0, \dots, x_{n+1}] \int_a^b \omega(x)\,dx \\ &= \int_a^b f[x_0, \dots, x_n, x_{n+1}, x](x - x_{n+1})\omega(x)\,dx\ . \end{aligned}$$

В последното равенство използвахме условието, че $\int_a^b \omega(x)\,dx = 0$. Да предположим сега, че x_{n+1} е точката, в която $\omega(x)$ си сменя знака. Тогава функцията $(x - x_{n+1})\,\omega(x)$ ще има постоянен знак в $[a, b]$. По-нататък, при предположение, че f има непрекъсната производна, заключаваме въз основа на теоремата за средните стойности, че съществува точка $\xi \in (a, b)$, за която

$$R(f) = \frac{f^{(n+2)}(\xi)}{(n+2)!} \int_a^b (x - x_{n+1})\,\omega(x)\,dx. \tag{6}$$

Да отбележим, че в този случай грешката се изразява чрез $(n+2)$-та производна. Следователно $R(f) = 0$ за всяко $f \in \pi_{n+1}$, т.е. квадратурната формула е точна за всички полиноми дори от $(n+1)$-ва степен.

Използването на интерполационния полином за приближено пресмятане на интеграли е било предложено от Нютон. Английският инженер Коутс е пресметнал коефициентите на интерполационните квадратурни формули в $[0, 1]$ в случая на равноотдалечени възли $x_k = k/n,\ k = 0, \dots, n$, за $n = 1, 2, 3, \dots, 15$ и публикувал таблиците с тези коефициенти. Затова интерполационните квадратурни формули с равноотдаличени възли се наричат *квадратурни формули на Нютон-Коутс.*

Сега ще изведем в явен вид някои *елементарни квадратурни формули.* Нека $n = 0$. Тогава

$$L_0(f; x) = f(x_0)$$

и следователно

$$I(f) \approx I(L_0) = f(x_0)(b - a).$$

Специално при $x_0 = \frac{a+b}{2}$ получаваме

(7) $$\int_a^b f(x)\,dx \approx f\left(\frac{a+b}{2}\right)(b-a).$$

В този случай функцията $\omega(x) := x - \frac{a+b}{2}$ си сменя знака само в точката $x = (a+b)/2$ и $\int_a^b \omega(x)\,dx = 0$. Следователно, при $x_0 = x_1 = (a+b)/2$, полиномът $(x-x_0)(x-x_1) = \left(x - \frac{a+b}{2}\right)^2$ ще има постоянен знак в (a, b). Тогава, съгласно (6),

(8) $$E(f) = \frac{f''(\xi)}{2!}\int_a^b (x-x_0)^2\,dx = f''(\xi)\frac{(b-a)^3}{24}.$$

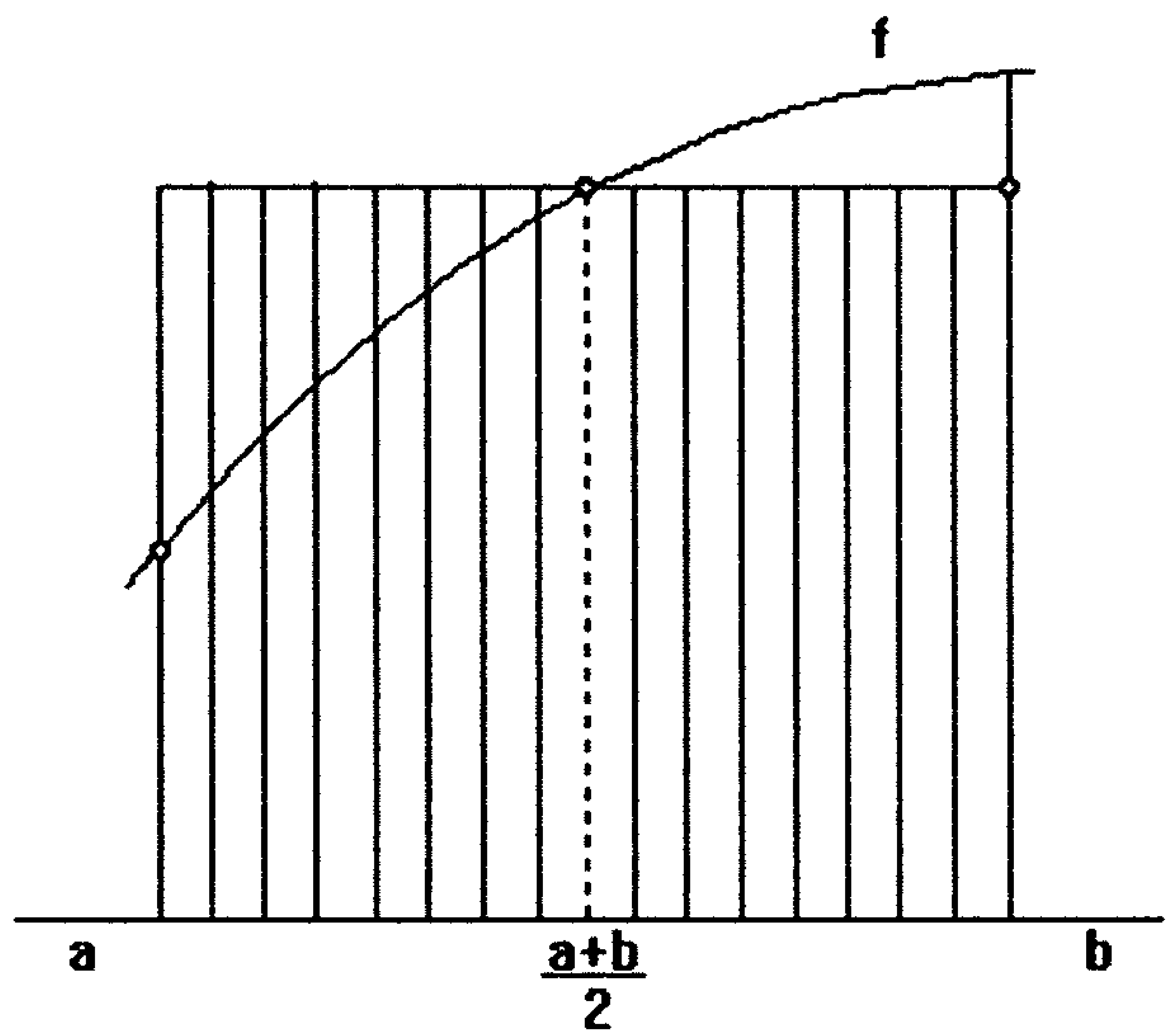

Рис. 10

Формулата (7) е известна като *квадратурна формула на правоъгълниците*. Тя има прост геометричен смисъл (виж Рис. 10). Интегралът $I(f)$, който е равен на лицето на фигурата, определена от графиката на f се приближава с лицето на правоъгълника с основа $[a, b]$ и височина $f\left(\frac{a+b}{2}\right)$. Оттук идва и наименованието на тази формула.

Нека сега $n = 1$. Да изберем $x_0 = a$, $x_1 = b$. Тогава

$$\begin{aligned} L_1(f; x) &= f(a) + f[a,b](x-a), \\ f(x) &= L_1(f; x) + f[a,b,x](x-a)(x-b). \end{aligned}$$

Заменяме $I(f)$ с $I(L_1)$ и получаваме квадратурната формула

(9) $$\int_a^b f(x)\,dx \approx \frac{b-a}{2}[f(a)+f(b)]\ .$$

За определянето на грешката ще се възползваме от (5), тъй като в този случай полиномът $\omega(x) = (x-a)(x-b)$ има постоянен знак в (a,b). Имаме

$$R(f) = \frac{f''(\xi)}{2}\int_a^b (x-a)(x-b)\,dx.$$

Пресмятаме интеграла и получаваме

(10) $$R(f) = -\frac{f''(\xi)}{12}(b-a)^3\ .$$

Формулата (9) се нарича *квадратурна формула на трапеците*. Нейният геометричен смисъл е изяснен на Рис. 11 по-долу.

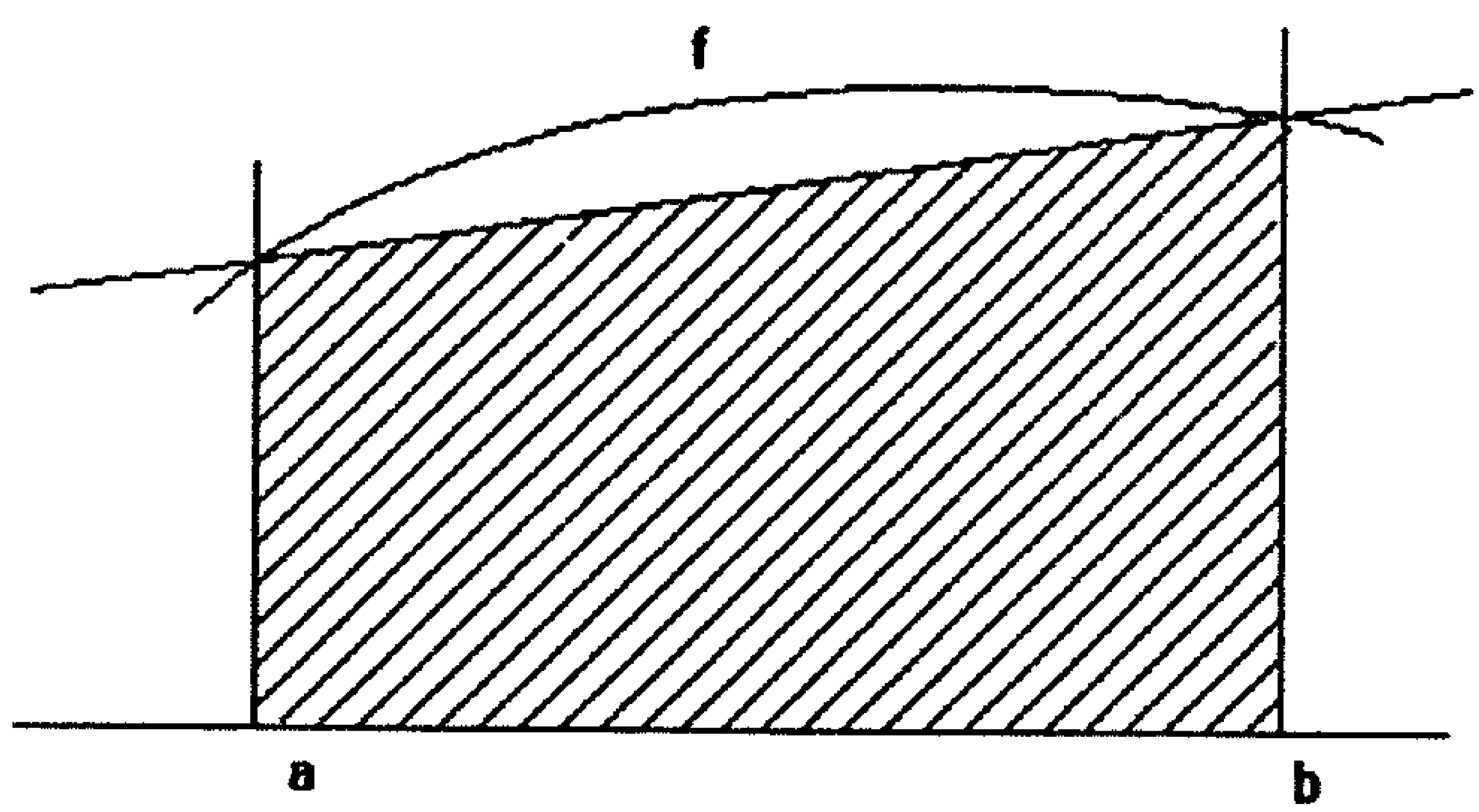

Рис. 11

Сега да разгледаме една интерполационна квадратурна формула с три равноотдалечени възела.

Нека $n = 2$. Имаме

$$f(x) = L_2(f;x) + f[x_0,x_1,x_2,x](x-x_0)(x-x_1)(x-x_2).$$

Оттук получаваме формулата

(11) $$I(f) \approx I(L_2).$$

Да изберем $x_0 = a$, $x_1 = (a+b)/2$, $x_2 = b$. В този случай функцията $\omega(x) = (x-a)\left(x-\frac{a+b}{2}\right)(x-b)$ си сменя знака в (a,b) само в точката $x_3 := (a+b)/2$.

Освен това $\int_a^b \omega(x)\,dx = 0$. Следователно остатъчният член се представя както в (6). Имаме

$$R(f) = \frac{f^{IV}(\xi)}{4!}\int_a^b (x-a)\left(x-\frac{a+b}{2}\right)^2(x-b)\,dx \ .$$

Пресмятаме интеграла и получаваме окончателно

$$R(f) = -\frac{f^{IV}(\xi)}{2880}(b-a)^5 \ . \tag{12}$$

За да получим явния вид на квадратурната формула (11) бихме могли да запишем $L_2(f;x)$ по формулата на Нютон и да пресметнем $I(L_2)$. Ние ще покажем тук един по-прост начин. Да означим за удобство с $p(x)$ интерполационния полином $L_2(f;x)$. По формулата на правоъгълниците (7) и на трапеците (9) имаме, съответно,

$$\begin{aligned} I(p) &= p\left(\frac{a+b}{2}\right)(b-a) + \frac{p''(\xi_1)}{24}(b-a)^3, \\ I(p) &= \frac{b-a}{2}[p(a)+p(b)] - \frac{p''(\xi_2)}{12}(b-a)^3, \end{aligned}$$

където ξ_1 и ξ_2 са някакви точки от (a,b). Но $p \in \pi_2$. Следователно $p''(t)$ е константа за всяко t. Оттук $p''(\xi_1) = p''(\xi_2)$. Тогава да умножим второто уравнение с $\frac{1}{2}$ и да го прибавим към първото. Получаваме

$$I(p) + \frac{1}{2}I(p) = p\left(\frac{a+b}{2}\right)(b-a) + \frac{b-a}{4}[p(a)+p(b)].$$

Тъй като полиномът $p(x)$ интерполира $f(x)$ в точките a, $\frac{a+b}{2}$ и b, то от горното равенство следва

$$I(p) = \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] \ . \tag{13}$$

Получихме формулата

$$\int_a^b f(x)\,dx \approx \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] \ . \tag{14}$$

Това е знаменитата *квадратурна формула на Симпсън*. От израза за грешката и (12) се вижда, че тя е точна за всички полиноми от степен по-малка или равна на 3.

Изведените тук квадратурни формули (на правоъгълниците, на трапеците, на Симпсън) се наричат *елементарни квадратурни формули*. В този си вид те рядко се използват на практика, защото грешката при тяхното приложение е голяма, особено ако интервалът на интегриране $[a, b]$ е голям. Това се вижда от изразите (8), (10) и (12) за грешките. На практика обикновено се постъпва по следния начин: Интервалът $[a, b]$ се разделя на m равни части с помощта на точките $x_0, \ldots, x_m$. След това във всеки подинтервал $[x_{i-1}, x_i]$ се прилага някоя от елементарните квадратурни формули за пресмятане на интеграла $\int_{x_{i-1}}^{x_i} f(x)\,dx$ и получените изрази се сумират. В резултат се получава така наречената *съставна квадратурна формула*. Ние ще изведем тук и съставните формули в явен вид, защото те имат голямо приложение.

Съставна квадратурна формула на правоъгълниците. Нека $x_i = a + ih$, $i = 0, \ldots m$, $h = (b-a)/m$. По формулата на правоъгълниците имаме

$$\int_{x_{i-1}}^{x_i} f(x)\,dx = f\left(\frac{x_{i-1}+x_i}{2}\right)(x_i - x_{i-1}) + \frac{f''(\xi_i)}{24}h^3,$$

където ξ_i е някаква точка от интервала (x_{i-1}, x_i). Сумираме горните равенства за $i = 1, \ldots, m$ и получаваме квадратурната формула

$$\int_a^b f(x)\,dx \approx \frac{b-a}{m}\sum_{i=1}^{m} f\left(\frac{x_{i-1}+x_i}{2}\right)$$

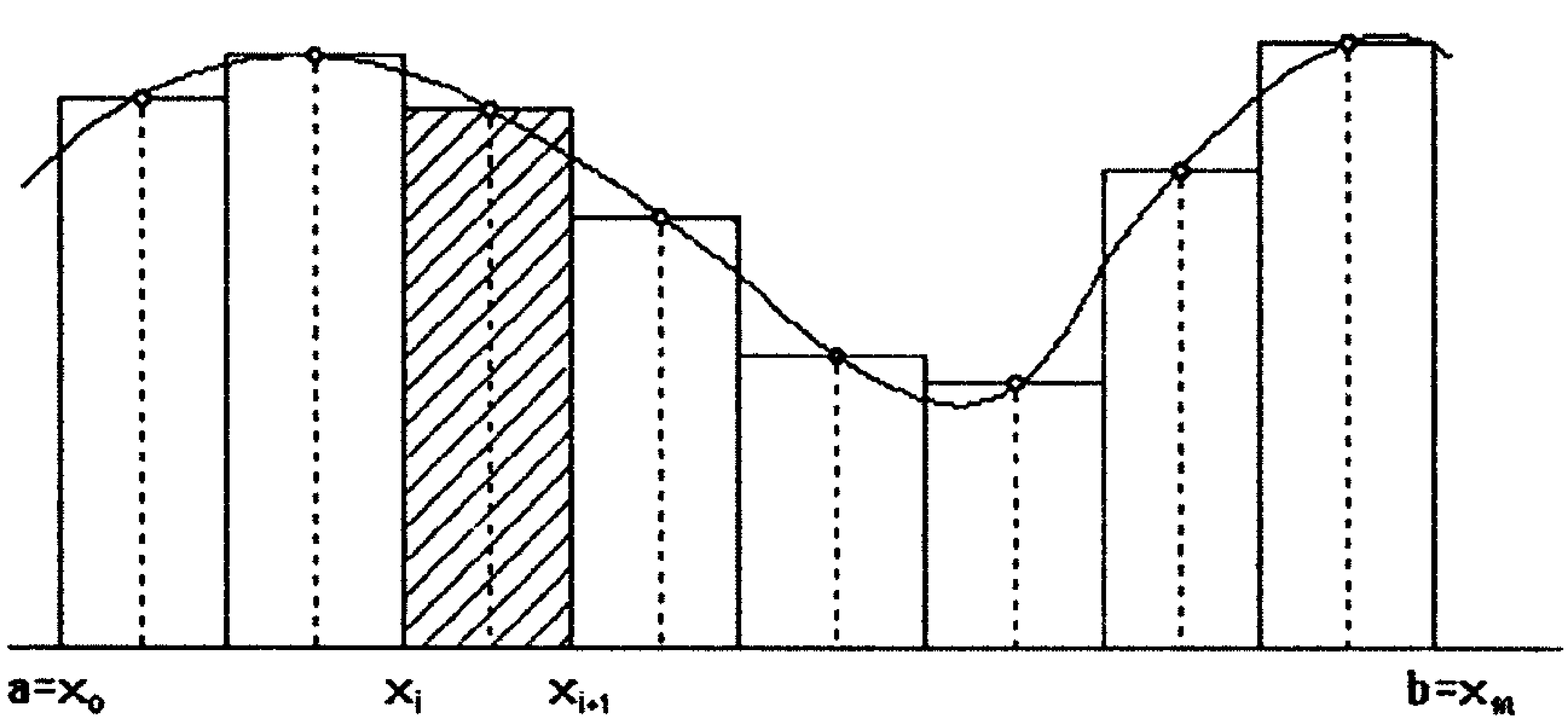

Рис. 12

с грешка

$$R(f) = \frac{(b-a)^3}{24m^2}\frac{1}{m}\sum_{i=1}^{m} f''(\xi_i)\,.$$

Но числото $\frac{1}{m}[f''(\xi_1) + \cdots + f''(\xi_m)]$ е средно аритметично на m стойности на $f''(x)$ в $[a, b]$. Следователно то се намира между точната долна и точната горна граница на $f''(x)$ в $[a, b]$. Оттук следва, че съществува точка $\xi \in [a, b]$ такава, че

$$\frac{1}{m}[f''(\xi_1) + \cdots + f''(\xi_m)] = f''(\xi).$$

Тогава за грешката на съставната квадратурна формула на правоъгълниците получаваме

$$R(f) = \frac{(b-a)^3}{24m^2} f''(\xi).$$

Сега вече се вижда, че с помощта на съставната формула можем да пресметнем интеграла $I(f)$ с каквато точност искаме, стига да изберем m достатъчно голямо число.

Съставна квадратурна формула на трапеците. Напълно аналогично, като използваме формулата на трапеците (9), получаваме

$$\begin{aligned} \int_a^b f(x)\,dx &\approx \frac{b-a}{2m}[f_0 + 2f_1 + \cdots + 2f_{m-1} + f_m], \\ R(f) &= -\frac{(b-a)^3}{12m^2} f''(\xi), \quad \xi \in [a, b]. \end{aligned}$$

Тук за краткост сме писали f_i вместо $f(x_i)$.

Съставна квадратурна формула на Симпсън. В този случай разделяме $[a, b]$ на четен брой подинтервали $[x_{i-1}, x_i]$, $i = 1, \ldots, 2m$ и след това прилагаме формулата на Симпсън за двойния подинтервал $[x_{i-1}, x_{i+1}]$, $i = 1, 3, 5, \ldots, 2m-1$. Получаваме

$$\begin{aligned} \int_a^b f(x)\,dx &\approx \frac{b-a}{6m}\Big\{f_0 + f_{2m} + 2[f_2 + f_4 + \ldots + f_{2m-2}] \\ &+ 4[f_1 + f_3 + \ldots + f_{2m-1}]\Big\}, \\ R(f) &= -\frac{(b-a)^5}{2880m^4} f^{IV}(\xi), \quad \xi \in [a, b]. \end{aligned}$$

18. КВАДРАТУРНА ФОРМУЛА НА ГАУС

Да разгледаме квадратурната формула от общия вид

$$\int_a^b \mu(x)f(x)\,dx \approx \sum_{k=1}^{n} A_k f(x_k), \tag{1}$$

където $\mu(x)$ е дадено тегло, определено в $[a,b]$, $a \le x_1 < \cdots < x_n \le b$, а $\{A_k\}_1^n$ са реални числа. Вече видяхме, че при всеки избор на възлите $\{x_k\}_1^n$ можем да намерим коефициенти $\{A_k\}_1^n$ такива, че получената квадратурна формула (1) да бъде точна за всички алгебрични полиноми от степен $n-1$. За целта, достатъчно е да построим интерполационната квадратурна формула от този вид

$$\begin{aligned}\int_a^b \mu(x)f(x)\,dx &\approx \int_a^b \mu(x)L_{n-1}(f;x)\,dx\\ &= \sum_{k=1}^{n}\left\{\int_a^b \mu(x)\prod_{i=1,i\neq k}^{n}\frac{x-x_i}{x_k-x_i}\,dx\right\}f(x_k).\end{aligned}$$

Дали не съществуват някакви специални възли $\{x_k^*\}_1^n$, при които съответната интерполационна квадратурна формула да бъде точна за полиномите и от степен по-висока от $n-1$? Ние вече срещнахме такива примери: формулата на Симпсън е с три възела, а е точна за всички полиноми не само от втора, но и от трета степен. Тук се сблъскваме с една нова, важна характеристика на квадратурните формули.

Определение. Казваме, че една квадратурна формула има *алгебрическа степен на точност* (накратко АСТ) m, ако тя е точна за всички алгебрични полиноми от степен $\le m$ и съществува полином от степен $m+1$, за който тя не е точна.

Каква най-висока АСТ може да има една квадратурна формула с n възела? При кои възли се достига най-висока АСТ? Това са въпросите, на които ще се спрем тук.

Лесно се вижда, че максималната алгебрическа степен на точност на формулата (1) е по-голяма или равна на $n-1$. Наистина, при всеки избор на точките $\{x_k\}_1^n$ можем да построим съответната интерполационна квадратурна формула с възли в $\{x_k\}_1^n$, която по определение е точна за всички полиноми от π_{n-1}, т.е. има АСТ равна поне на $n-1$. Сега ще покажем с един контрапример, че не съществува квадратурна формула от вида (1) с АСТ по-висока от $2n-1$. Наистина, ако има такава формула, то тя трябва да е точна и за полинома

$$\omega^2(x) = (x-x_1)^2\dots(x-x_n)^2,$$

който е от степен $2n$. Но

$$\int_a^b \mu(x)\omega^2(x)\,dx > 0,$$

докато

$$\sum_{k=1}^{n} A_k\omega^2(x_k) = 0.$$

Следователно квадратурната формула (1) не е точна за полинома $f(x) = \omega^2(x)$. И така, най-високата АСТ на формулата (1) е по-малка или равна на $2n-1$.

Свободните параметри в квадратурната формула (1) са $2n$ на брой – това са възлите $\{x_k\}$ и коефициентите $\{A_k\}$. Следователно имаме известно основание да очакваме, че съществува такъв избор на параметрите $\{x_k\}$ и $\{A_k\}$, при който да се удовлетворят $2n$ уравнения, изразяващи точността на формулата за базисните функции $1, x, x^2, \ldots, x^{2n-1}$, които са също $2n$ на брой. По-долу ще покажем, че наистина съществуват възли $\{x_k\}_1^n$ и коефициенти $\{A_k\}_1^n$, при които съответната квадратурна формула (1) има АСТ равна на $2n-1$. Тази формула е била построена за първи път от Гаус и затова се нарича *квадратурна формула на Гаус.*

Теорема 1. *При всяко естествено число n съществува единствена квадратурна формула от вида (1) с АСТ $= 2n-1$. Възлите $\{x_k\}_1^n$ на тази формула са нулите на полинома от степен n, ортогонален в $[a,b]$ с тегло $\mu(x)$ на всички алгебрични полиноми от степен $n-1$.*

Доказателство. Нека $\omega(x)$ е полиномът от степен n, с коефициент 1 пред x^n, който е ортогонален в $[a,b]$ с тегло $\mu(x)$ на всички полиноми от степен $n-1$. Видяхме (вж. Теорема 14.1), че съществува единствен такъв полином и той има точно n реални, различни нули в (a,b). Да означим тези нули с $x_1, \ldots, x_n$. Имаме $\omega(x) = (x-x_1)\ldots(x-x_n)$. Да построим интерполационната квадратурна формула от вида (1) с възли – нулите $\{x_k\}_1^n$ на $\omega(x)$. Ще покажем, че тази формула има АСТ $= 2n-1$, която е възможно най-високата. Наистина, нека f е произволен полином от степен $2n-1$. Да разделим $f(x)$ на $\omega(x)$. Получаваме

$$f(x) = \omega(x)q(x) + r(x), \tag{2}$$

където q и r са полиноми от степен по-малка или равна на $n-1$. Тогава

$$\begin{aligned}\int_a^b \mu(x)f(x)\,dx &= \int_a^b \mu(x)\omega(x)q(x)\,dx + \int_a^b \mu(x)r(x)\,dx\\ &= \int_a^b \mu(x)r(x)\,dx.\end{aligned}$$

Използвахме, че $\omega(x)$ е ортогонален на $q(x)$. Тъй като формулата (1) е интерполационна, тя е точна за $r(x)$. Следователно

$$\int_a^b \mu(x)r(x)\,dx = \sum_{k=1}^n A_k r(x_k).$$

Остава да забележим, че $r(x_k) = f(x_k)$, $k = 1,\dots,n$. Това следва от (2), като се вземе предвид, че $\omega(x_k) = 0$. Окончателно получаваме

$$\int_a^b \mu(x)f(x)\,dx = \sum_{k=1}^n A_k r(x_k) = \sum_{k=1}^n A_k f(x_k).$$

Квадратурната формула е точна за произволен полином f от класа π_{2n-1}. Следователно тя има АСТ равна на $2n-1$.

Сега да докажем обратното. Нека квадратурната формула (1) има АСТ $= 2n-1$. Ще покажем, че полиномът $\omega(x) = (x-x_1)\dots(x-x_n)$ е ортогонален на всеки полином от π_{n-1}. Наистина, нека Q е произволен полином от π_{n-1}. Тогава полиномът $f(x) = Q(x)\omega(x)$ е от степен $2n-1$ и квадратурната формула (1) ще бъде точно за него. Имаме

$$\int_a^b \mu(x)Q(x)\omega(x)\,dx = \sum_{k=1}^n A_k Q(x_k)\omega(x_k) = 0,$$

т.е. ω е ортогонален на Q. Твърдението е доказано.

Единствеността на квадратурната формула с най-висока АСТ следва от единствеността на полинома $(x-x_1)\dots(x-x_n)$, ортогонален на всички полиноми от π_{n-1}.

Коефициентите $\{A_k\}$ на квадратурната формула на Гаус са положителни числа. Това се вижда по следния начин: Полиномът $\omega_k(x) := \omega(x)/(x-x_k)$ е от степен $n-1$. Тогава полиномът $\varphi_k(x) := \omega_k^2(x)/\omega_k^2(x_k)$ е от степен $2n-2$, той е неотрицателен и $\varphi_k(x_k) = 1$. Тъй като φ_k се интегрира точно по формулата на Гаус, то

$$0 < \int_a^b \mu(x)\varphi_k(x)\,dx = \sum_{j=1}^n A_j\varphi_k(x_j) = A_k$$

и твърдението е доказано.

Формулата

$$A_k = \int_a^b \mu(x)\frac{\omega_k^2(x)}{\omega_k^2(x_k)}\,dx$$

може да се използва за пресмятане на коефициентите $\{A_k\}$ на квадратурната формула на Гаус. Ние ще дадем сега един друг начин за пресмятане на A_k, при който се избягва интегрирането.

Нека $P_0(x), P_1(x), \dots, P_n(x)$ е редицата от ортогонални полиноми в $[a,b]$ с тегло $\mu(x)$. Да предположим, че тази редица е ортонормирана, т.е.

$$\int_a^b \mu(x) P_k^2(x)\, dx = 1 \text{ за всяко } k.$$

Освен това ще предположим, че водещите коефициенти α_k на $P_k(x)$ са положителни. Нека $x_1, \dots, x_n$ са нулите на $P_n(x)$.

$$P_n(x) = \alpha_n (x - x_1) \dots (x - x_n) = \alpha_n x^n + \cdots, \ \alpha_n > 0$$

От доказаната теорема следва, че $x_1, \dots, x_n$ са възлите на квадратурната формула на Гаус. За да намерим коефициентите $\{A_k\}$, да разгледаме сумата

$$D[f] := \alpha_{n-1} \sum_{k=1}^{n} A_k P_{n-1}(x_k) f(x_k).$$

При $f \in \pi_{n-1}$ полиномът $f(x)P_{n-1}(x)$ е от степен $2n-2$ и се интегрира точно по формулата на Гаус. Следователно,

$$\begin{aligned} D[f] &= \alpha_{n-1} \int_a^b \mu(x) f(x) P_{n-1}(x)\, dx = 0 \quad \text{при} \quad f \in \pi_{n-2}, \\ D[x^{n-1}] &= \alpha_{n-1} \int_a^b \mu(x) x^{n-1} P_{n-1}(x)\, dx = \int_a^b \mu(x) P_{n-1}^2(x)\, dx = 1. \end{aligned}$$

Но тези две свойства характеризират напълно разделената разлика $f[x_1, \dots, x_n]$. Следователно

$$D[f] = f[x_1, \dots, x_n] = \alpha_n \sum_{k=1}^{n} \frac{f(x_k)}{P_n'(x_k)}.$$

Като приравним коефициентите пред $f(x_k)$ в горния израз и тези в $D[f]$, получаваме

$$A_k = \frac{\alpha_n}{\alpha_{n-1}} \frac{1}{P_n'(x_k) P_{n-1}(x_k)}, \quad k = 1, \dots, n.$$

Това са известни формули за пресмятане на коефициентите на квадратурната формула на Гаус.

По-долу ще дадем оценка на грешката на квадратурната формула на Гаус. Да предположим, че функцията $f(x)$ има непрекъсната $2n$-та производна в $[a,b]$. По формулата на Нютон,

$$f(x) = H_{2n-1}(x) + f[x_1, x_1, \dots, x_n, x_n, x](x - x_1)^2 \dots (x - x_n)^2,$$

където $H_{2n-1}(x)$ е полиномът от степен $2n-1$, който удовлетворява интерполационните условия

$$H_{2n-1}(x_k) = f(x_k), \quad H_{2n-1}'(x_k) = f'(x_k), \quad k = 1, \dots, n.$$

Тогава, като използваме, че квадратурната формула на Гаус е точна за H_{2n-1}, получаваме

$$\begin{aligned}
R_n(f) &:= \int_a^b \mu(x) f(x)\,dx - \sum_{k=1}^{n} A_k f(x_k) \\
&= \int_a^b \mu(x) f(x)\,dx - \sum_{k=1}^{n} A_k H_{2n-1}(x_k) \\
&= \int_a^b \mu(x) f(x)\,dx - \int_a^b \mu(x)\, H_{2n-1}(x)\,dx \\
&= \int_a^b \mu(x) f[x_1, x_1, \ldots, x_n, x_n, x]\,\omega^2(x)\,dx \\
&= \frac{f^{(2n)}(\xi)}{(2n)!} \int_a^b \mu(x)\,\omega^2(x)\,dx,
\end{aligned}$$

където ξ е някаква точка от $[a, b]$.

19. КВАДРАТУРНИ ФОРМУЛИ ОТ ГАУСОВ ТИП

Ще разглеждаме квадратурни формули от вида

$$(1) \qquad \int_a^b \mu(x)f(x)\,dx \approx \sum_{i=1}^{m} B_i f(t_i) + \sum_{k=1}^{n} A_k f(x_k) \ ,$$

където $a \le t_1 < \cdots < t_m \le b$, $a \le x_1 < \cdots < t_n \le b$ и никое от числата $\{t_i\}_1^m$ не съвпада с никое от $\{x_i\}_1^n$. При фиксирани $\{t_i\}_1^m$ ще се опитаме да определим останалите параметри $\{B_i\}_1^m$, $\{A_i\}_1^n$ и $\{x_i\}_1^n$ така, че квадратурната формула (1) да има възможно най-висока *алгебрическа степен на точност*, която понататък ще означаваме съкратено с АСТ(1). Общият брой на свободните параметри е $2n+m$. Следователно можем да очакваме, че те могат да бъдат избрани така, че квадратурната формула (1) да е точна за полиномите $1, x, \ldots, x^{2n+m-1}$, т.е. (1) да има алгебрическа степен на точност поне $2n+m-1$.

Следващата теорема дава характеризация на тези възли и коефициенти, при които $ACT(1) = 2n+m-1$.

Да въведем означенията

$$\begin{aligned} \sigma(x) &:= (x-t_1)\ldots(x-t_m), \\ \omega(x) &:= (x-x_1)\ldots(x-x_n). \end{aligned}$$

Теорема 1. *Квадратурната формула (1) е точна за всички полиноми от степен $\le 2n+m-1$ тогава и само тогава, когато тя е интерполационна и полиномът $\omega(x)$ е ортогонален в $[a,b]$ с тегло $\mu(x)\sigma(x)$ на всички полиноми от степен $n-1$.*

Доказателство. Ще използваме същата идея, както при доказателството на теоремата на Гаус. Ако $ACT(1) = 2n+m-1$, то (1) е очевидно от интерполационен тип. Да докажем ортогоналността на $\omega(x)$. Нека $Q(x)$ е произволен полином от π_{n-1}. Тогава полиномът $f(x) = \omega(x)\sigma(x)Q(x)$ е от π_{2n+m-1} и следователно

$$\int_a^b \mu(x)f(x)\,dx = \sum_{i=1}^{m} B_i f(t_i) + \sum_{k=1}^{n} A_k f(x_k) = 0 \ .$$

Това показва, че $\omega(x)$ е ортогонален на Q с тегло $\mu(x)\sigma(x)$ в $[a,b]$. Необходимостта е доказана.

Да допуснем сега, че ω е ортогонален на всеки полином от π_{n-1} с тегло $\mu(x)\sigma(x)$. Да построим интерполационната формула (1) с възли $x_1, \ldots, x_n$ – нулите на ω. Ще докажем, че (1) е точна за всяко $f \in \pi_{2n+m-1}$. Наистина, нека $f \in \pi_{2n+m-1}$. Тогава f може да се представи във вида

$$f(x) = \omega(x)\sigma(x)Q(x) + r(x) \ ,$$

с някакво $Q \in \pi_{n-1}$ и $r \in \pi_{n+m-1}$. Използвайки условието, че

$$\int_a^b \mu(x)\sigma(x)\omega(x)Q(x)\,dx = 0$$

и че (1) е точна за r, получаваме

$$\begin{aligned}\int_a^b \mu(x)f(x)\,dx &= \int_a^b \mu(x)\sigma(x)\omega(x)Q(x)\,dx + \int_a^b \mu(x)r(x)\,dx\\ &= \sum_{i=1}^m B_i r(t_i) + \sum_{k=1}^n A_k r(x_k)\\ &= \sum_{i=1}^m B_i f(t_i) + \sum_{k=1}^n A_k f(x_k),\end{aligned}$$

т.е. (1) е точна за f. Теоремата е доказана.

Да предположим сега, че $\sigma(x) \geq 0$ в $[a,b]$. Тогава $\mu(x)\sigma(x) \geq 0$ и следователно съществува единствен полином $\omega \in \pi_n$, който е ортогонален в $[a,b]$ с тегло $\mu(x)\sigma(x)$ на полиномите от π_{n-1}. Това означава, че съществува *единствена* квадратурна формула от вида (1) с $ACT = 2n+m-1$. Нещо повече, няма квадратурна формула от вида (1) , която има ACT по висока от $2n+m-1$. Това следва веднага от факта, че (1) не е точна например за полинома $\sigma(x)\omega^2(x)$. Следователно е в сила следното твърдение.

Следствие 2. *Ако $\sigma(x) \geq 0$ в $[a,b]$, съществува единствена квадратурна формула от вида (1) с най-висока алгебрическа степен на точност, равна на $2n+m-1$.*

Оттук нататък ще предполагаме, че $\sigma(x) \geq 0$.

Ще дадем едно представяне на остатъчния член

$$R_{n,m}(f) := \int_a^b \mu(x)f(x)\,dx - S(f)$$

($S(f)$ е изразът в дясната страна на (1)). Нека $f \in C^{2n+m}[a,b]$ и p е полиномът от π_{2n+m-1}, който интерполира f в точките $t_1,\ldots,t_m$ и $x_1,x_1,\ldots,x_n,x_n$. По формулата на Нютон

$$f(x) = p(x) + f[t_1,\ldots,t_m,x_1,x_1,\ldots,x_n,x_n,x]\sigma(x)\omega^2(x)\ .$$

Ако $ACT(1) = 2n+m-1$, то квадратурната формула (1) интегрира точно p. Имаме

$$\begin{aligned}\int_a^b \mu(x)f(x)\,dx &= \int_a^b \mu(x)p(x)\,dx + \int_a^b \mu(x)f[t_1,\ldots,x_n,x_n,x]\sigma(x)\omega^2(x)\,dx\\ &= S(p) + \frac{f^{(2n+m)}(\xi)}{(2n+m)!}\int_a^b \mu(x)\sigma(x)\omega^2(x)\,dx,\end{aligned}$$

където ξ е точка от (a, b). Тъй като $S(p) = S(f)$, то

(3)
$$R_{n,m}(f) = \frac{f^{(2n+m)}(\xi)}{(2n+m)!} \int_a^b \mu(x)\sigma(x)\omega^2(x)\,dx.$$

Ще разгледаме по-подробно два специални случая, когато фиксираните възли са в краищата на интервала:

1. $m = 2,\ t_1 = a,\ t_2 = b$,
2. $m = 1,\ t_1 = a$, (или $t_1 = b$).

За простота ще приемем, че $a = -1,\ b = 1$ и $\mu(x) \equiv 1$.

Квадратурна формула на Лобато. В случай 1 квадратурната формула (1) добива вида

(2)
$$\int_{-1}^{1} f(x)\,dx \approx B_1 f(-1) + B_2 f(1) + \sum_{k=1}^{n} A_k f(x_k).$$

Нейната максимална алгебрическа степен на точност е $2n+1$. Съгласно Теорема 1, възлите $x_1, \ldots, x_n$ на екстремалната квадратурна формула са нулите на полинома $\omega(x)$, ортогонален в $[-1, 1]$ с тегло $(x-1)(x+1)$ на всеки полином от π_{n-1}. Да намерим ω. За целта ще представим $(x^2-1)\omega(x)$ като линейна комбинация на полиномите на Льожандър. Имаме

$$(x^2-1)\omega(x) = c_0 L_0(x) + c_1 L_1(x) + \cdots + c_{n+2} L_{n+2}(x).$$

Умножаваме двете страни с $L_k(x)$ и интегрираме. Получаваме

$$0 = \int_{-1}^{1} (x^2-1)\omega(x) L_k(x)\,dx = c_k \int_{-1}^{1} L_k^2(x)\,dx \quad \text{при } k = 0, \ldots, n-1.$$

Следователно $c_0 = \cdots = c_{n-1} = 0$ и

$$(x^2-1)\omega(x) = c_n L_n(x) + c_{n+1} L_{n+1}(x) + c_{n+2} L_{n+2}(x).$$

Специално при $x = 1$ и $x = -1$ имаме

$$\begin{aligned} 0 &= c_n + c_{n+1} + c_{n+2} \\ 0 &= (-1)^n [c_n - c_{n+1} + c_{n+2}]. \end{aligned}$$

Оттук определяме $c_{n+1} = 0$, $c_{n+2} = -c_n$. Следователно

$$(x^2-1)\omega(x) = c_n [L_n(x) - L_{n+2}(x)].$$

И така, възлите $-1, 1, x_1, \ldots, x_n$ на *квадратурната формула на Лобато* са нулите на полинома $L_{n+2}(x) - L_n(x)$. При намерени възли, коефициентите се определят от условието, че формулата е интерполационна.

Струва си да се отбележи, че $x_1, \ldots, x_n$ са нулите на полинома $L'_{n+1}(x)$. Наистина, интегрирайки по части получаваме

$$\int_{-1}^{1} (1-x^2)L'_{n+1}(x)f(x)\,dx = -\int_{-1}^{1} L_{n+1}(x)\Big[(1-x^2)f(x)\Big]'\,dx.$$

Но последният интеграл е равен на нула за всяко $f \in \pi_{n-1}$, защото L_{n+1} е ортогонален на π_n. Следователно $L'_{n+1}(x)$ е ортогонален на всеки полином f от π_{n-1} с тегло $(1-x^2)$. Тогава

$$L'_{n+1}(x) = c(x-x_1)\ldots(x-x_n)$$

с някаква константа c.

Сега не е трудно да се намерят коефициентите B_1 и B_2 на квадратурната формула на Лобато. Тъй като формулата е точна за полинома $(1+x)L'_{n+1}(x)$, то

$$\int_{-1}^{1} (1+x)L'_{n+1}(x)\,dx = 2L'_{n+1}(1)\,B_2\ .$$

От друга страна, чрез интегриране по части,

$$\int_{-1}^{1} (1+x)L'_{n+1}(x)\,dx = (1+x)L_{n+1}(x)\Big|_{-1}^{1} = 2L_{n+1}(1) = 2.$$

Следователно $B_2 = 1/L'_{n+1}(1)$. Аналогично, получаваме
$B_1 = (-1)^n/L'_{n+1}(-1)$. Но лесно се проверява, че $L'_{n+1}(1) = \frac{(n+1)(n+2)}{2}$, откъдето заключаваме, че

$$B_1 = B_2 = \frac{2}{(n+1)(n+2)}.$$

Квадратурна формула на Радо. При $m = 1$ и $t_1 = -1$ квадратурната формула (1) приема вида

$$\int_{-1}^{1} f(x)\,dx \approx Bf(-1) + \sum_{k=1}^{n} A_k f(x_k). \tag{3}$$

Формулата с максимална алгебрическа степен на точност (равна на $2n$) от вида (3) е известна като *квадратурна формула на Радо*. Да намерим нейните възли. Съгласно Теорема 1, $x_1, \ldots, x_n$ са нулите на полинома $\omega(x)$, който е ортогонален с тегло $1+x$ на всички полиноми от π_{n-1} в $[-1, 1]$. Ще търсим отново $(1+x)\omega(x)$ като линейна комбинация на полиномите на Льожандър. Имаме

$$(1+x)\omega(x) = c_0L_0(x) + \cdots + c_nL_n(x) + c_{n+1}L_{n+1}(x).$$

Както и в предишния случай се вижда, че $c_0 = \cdots = c_{n-1} = 0$. Следователно

$$(1+x)\omega(x) = c_nL_n(x) + c_{n+1}L_{n+1}(x) .$$

При $x = -1$ получаваме

$$c_nL_n(-1) + c_{n+1}(-1)^{n+1} = (-1)^n(c_n - c_{n+1}) = 0,$$

т.е. $c_n = c_{n+1}$. И така,

$$(1+x)\omega(x) = c_n\left(L_n(x) + L_{n+1}(x)\right).$$

Оттук се вижда, че възлите $-1, x_1, \ldots, x_n$ на квадратурната формула на Радо са нулите на полинома $L_n(x) + L_{n+1}(x)$.

Задача. Покажете, че коефициентите в квадратурните формули на Лобато и Радо са положителни числа. Докажете, че $B = 2/(n+1)^2$.

Глава 3

ЧИСЛЕНО РЕШАВАНЕ НА УРАВНЕНИЯ

Точките x, за които $f(x) = 0$ се наричат *нули* на f или *корени* на уравнението $f(t) = 0$. Много задачи от практиката водят до необходимостта от числено пресмятане на корени на уравнения. Най-често това са *алгебрични уравнения*, т.е. уравнения от вида $p(t) = 0$, където p е алгебричен полином. Ако функцията f може да бъде достатъчно добре приближена с алгебричен полином p, то корените на уравнението $p(t) = 0$ ще бъдат добри приближения за корените на $f(t) = 0$. Оттук следва, че би било добре да разполагаме с числени методи за решаване на алгебрични уравнения. Тук ще се запознаем с такива методи. Преди това обаче, ще дадем някои класически резултати за оценки и брой на корените на едно алгебрично уравнение.

20. ОЦЕНКИ ЗА КОРЕНИТЕ

Ще започнем с едно правило на Коши за намиране на кръг, който съдържа всички нули на даден полином p с комплексни коефициенти. Знае се от алгебрата, че ако $p(z)$ е полином от степен n, то p има точно n нули в комплексната равнина. От голямо значение е да се намери една крайна област (например кръг), която съдържа всичките нули на p. Тогава бихме могли с други методи по-лесно да локализираме в тази крайна област всяка от нулите на p.

Теорема на Коши. *Нека $p(z) = z^n + a_1 z^{n-1} + \ldots + a_n$ е алгебричен полином с произволни комплексни коефициенти. Да предположим, че $a_n \neq 0$. Тогава всяка нула x на p удовлетворява неравенството*

$$|x| \leq R \,,$$

където R е единственият положителен корен на уравнението

$$t^n - |a_1|t^{n-1} - \ldots - |a_{n-1}|t - |a_n| = 0 \,. \tag{1}$$

С други думи, всички корени z на уравнението $p(z) = 0$ лежат в кръг с център в нулата и радиус R.

Доказателство. Най-напред ще покажем, че уравнението (1) наистина има само един положителен корен. (Това следва от теорема на Декарт, която ще докажем в следващата лекция, но ние предпочитаме да дадем тук директно доказателство.)

Вижда се, че при $t > 0$ (1) е еквивалентно с уравнението

$$t^n = |a_1|t^{n-1} + \ldots + |a_{n-1}|t + |a_n| \,,$$

а оттук, като разделим двете страни с t^n, стигаме до уравнението

$$1 = |a_1|\frac{1}{t} + \ldots + |a_{n-1}|\frac{1}{t^{n-1}} + |a_n|\frac{1}{t^n} .$$

Функцията в дясната страна е строго монотонно намаляваща в $(0, \infty)$, като приема прозволно големи стойности при t близко до нулата и клони към 0 при $t \to \infty$. Следователно графиката на тази функция ще пресече точно един път графиката на функцията $y = 1$ (в лявата страна), в някаква точка R от $(0, \infty)$ (Виж Рис. 13).

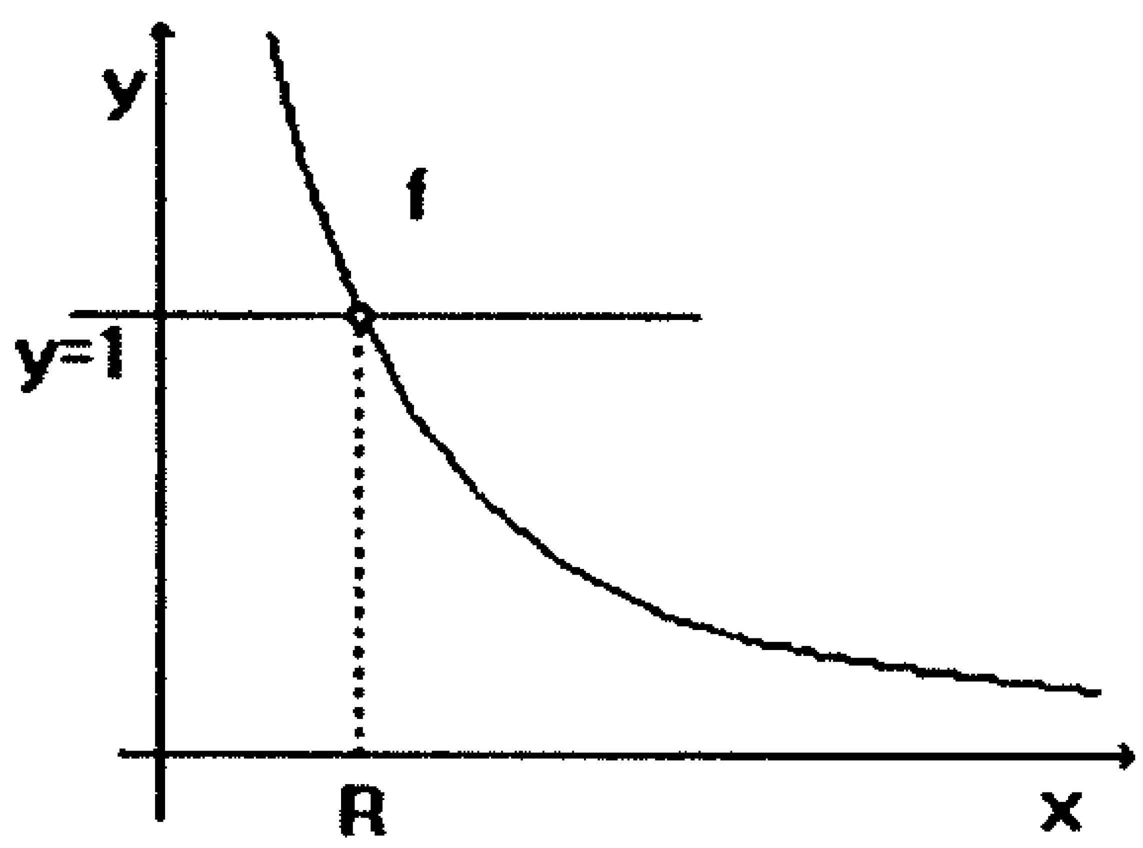

Рис. 13

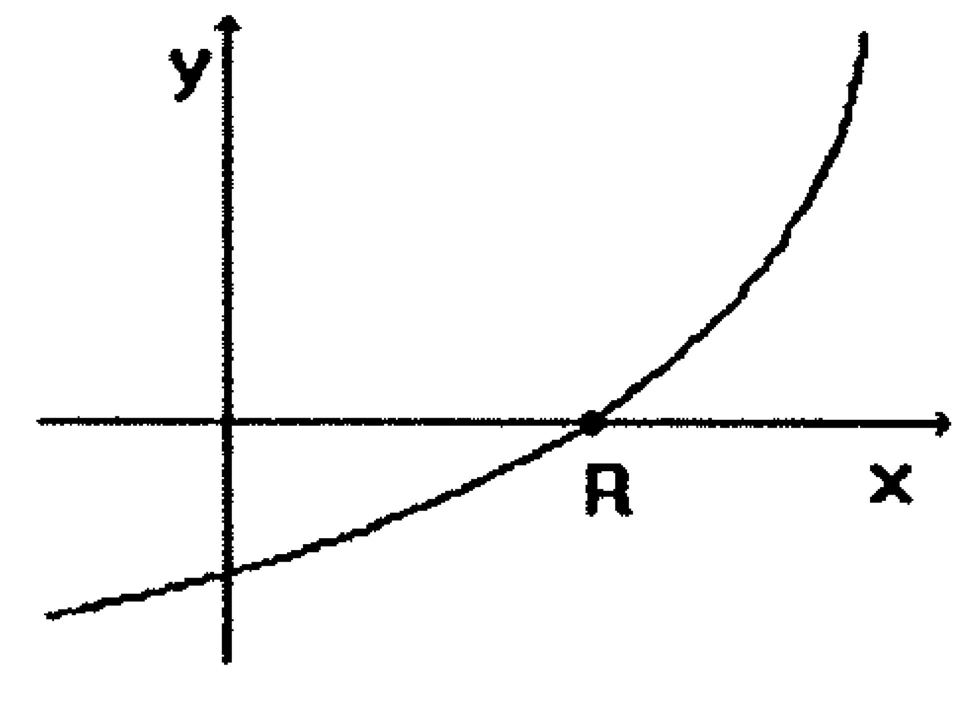

Рис. 14

Да означим с φ функцията

$$\varphi(t) := t^n - |a_1|t^{n-1} - \ldots - |a_{n-1}|t - |a_n| .$$

От казаното дотук следва, че φ има единствена положителна нула – точката R. На Рис. 14 е показана графиката на φ. Оттам се вижда, че ако $\varphi(x) \leq 0$ за някое $x > 0$, то $x \leq R$. Ние ще използваме този факт след малко.

Да предположим сега, че $p(z) = 0$ за някое z. Тогава

$$z^n = -a_1 z^{n-1} - \ldots - a_{n-1}z - a_n$$

и следователно

$$\begin{aligned} |z|^n &= |a_1 z^{n-1} + \ldots + a_{n-1}z + a_n| \\ &\leq |a_1|\,|z|^{n-1} + \ldots + |a_{n-1}|\,|z| + |a_n| . \end{aligned}$$

Последното неравенство показва, че $\varphi(|z|) \leq 0$ и съгласно направените по-горе забележки това влече $|z| \leq R$. Теоремата е доказана.

Сега ще дадем едно правило, предложено от Лагранж, за намиране на горна граница на положителните нули на едно алгебрично уравнение с реални коефициенти. Това правило може да бъде използвано, например за намиране на горна граница за корена R от предишната теорема. Изчисляването на добро приближение на R изисква други числени методи.

Правило на Лагранж. *Нека $f(x) = a_0x^n + \ldots + a_{n-1}x + a_n$ е произволен алгебричен полином с реални коефициенти и $a_0 \neq 0$. Да приемем за определеност, че $a_0 > 0$. Да означим с k индекса на първия отрицателен коефициент в редицата $a_0, a_1, \ldots, a_n$ и нека A е абсолютната стойност на най-големия по абсолютна стойност отрицателен коефициент в редицата $a_0, a_1, \ldots, a_n$. Тогава всеки положителен корен x на уравнението $f(t) = 0$ удовлетворява неравенството*

$$x < 1 + \sqrt[k]{\frac{A}{a_0}} \, .$$

Доказателство. Да предположим, че $x \geq 1 + \sqrt[k]{\frac{A}{a_0}} := \rho$. Тогава

$$\begin{aligned}
f(x) &= a_0x^n + \ldots + a_kx^{n-k} + \ldots + a_n \\
&\geq a_0x^n + a_kx^{n-k} + \ldots + a_n \qquad (\text{защото } a_i \geq 0 \text{ за } i = 0, \ldots, k-1) \\
&\geq a_0x^n - A(x^{n-k} + \ldots + x + 1) \ (\text{поради избора на } A) \\
&= a_0x^n - A\frac{x^{n-k+1} - 1}{x - 1} \\
&> a_0x^n - A\frac{x^{n-k+1}}{x - 1} \qquad (\text{защото } \frac{1}{x-1} > 0) \\
&= \frac{x^{n-k+1}}{x - 1}(a_0x^{k-1}(x-1) - A) \\
&> \frac{x^{n-k+1}}{x - 1}(a_0(x-1)^k - A) \qquad (\text{защото } x > x - 1)\, .
\end{aligned}$$

Тъй като, съгласно направеното в началото предположение,

$$x \geq 1 + \sqrt[k]{\frac{A}{a_0}} \, ,$$

то $(x-1)^k \geq A/a_0$ и следователно $a_0(x-1)^k - A \geq 0$. И така, показахме, че при $x \geq \rho$ имаме $f(x) > 0$, т.е. $f(x)$ не се анулира при $x \geq \rho$. Следователно всички положителни корени на уравнението $f(t) = 0$ са по-малки от ρ. Твърдението е доказано.

Има и други прости правила за намиране на горна граница на положителните корени на едно алгебрично уравнение. Тук ние няма да се спираме на тях.

Да отбележим, че разполагайки с метод за намиране на горна граница на положителните корени, ние бихме могли да намерим долна граница на отрицателните корени, т.е. да локализираме в краен интервал $[m, M]$ всички реални корени на едно алгебрично уравнение. Това става, като втората задача (за отрицателните корени) се сведе чрез смяна на променливата $x = -t$ към първата задача (за граница на положителните корени). Наистина, нека $-x_1 < \ldots < -x_j < 0$ са отрицателните корени на $f(x) = 0$. Да въведем полинома

$$g(t) := f(-t) .$$

Ясно е, че $0 < x_j < \ldots < x_1$ ще бъдат положителните корени на уравнението $g(t) = 0$. По някой от известните методи ние можем да намерим горна граница ρ на тези положителни корени,

$$x_j < \ldots < x_1 < \rho .$$

Тогава, очевидно $-\rho < -x_1 < \ldots < -x_j < 0$ и следователно $-\rho$ ще бъде долна граница на отрицателните корени на f. Аналогично, като използваме смените

$$x = \frac{1}{t}, \quad x = -\frac{1}{t} ,$$

ние можем да намерим долна граница на положителните и съответно горна граница на отрицателните корени на всяко алгебрично уравнение $f(x) = 0$.

Сега ще минем към една по-трудна задача: Да се намери броят на реалните корени на едно алгебрично уравнение, които лежат в даден интервал $[a, b]$. Почти във всички методи за оценка на този брой съществена роля играе следната лема, която всъщност важи не само за полиноми, но и за достатъчно гладки функции.

Лема 1. *Нека f има непрекъснати производни до k-тата включително в околност $\mathcal{U}$ на точката c. Нека*

$$f(c) = f'(c) = \ldots = f^{(k-1)}(c) = 0 \quad \text{и} \quad f^{(k)}(c) \neq 0 .$$

Тогава за всяко достатъчно малко $\varepsilon > 0$ имаме

$$\begin{aligned} f(c+\varepsilon)f'(c+\varepsilon) &> 0, \\ f(c-\varepsilon)f'(c-\varepsilon) &< 0. \end{aligned}$$

Доказателство. Лемата твърди, че преди всеки корен c на уравнението $f(t) = 0$ функцията f и нейната производна имат противни знаци, а след корена, те имат еднакви знаци. Доказателството се основава на формулата на Тейлър. За всяко достатъчно малко h (по-точно такова, че $c+h,\ c-h \in \mathcal{U}$) имаме

$$f(c+h) = f(c) + \frac{f'(c)}{1!}h + \frac{f''}{2!}h^2 + \ldots + \frac{f^{(k-1)}(c)}{(k-1)!}h^{k-1} + \frac{f^k(c+\theta h)}{k!}h^k ,$$

където θ е някакво число от интервала $(0,1)$.

Аналогично,

$$f'(c+h) = f'(c) + \frac{f''(c)}{1!}h + \ldots + \frac{f^{(k-1)}(c)}{(k-2)!}h^{k-2} + \frac{f^{(k)}(c+\theta_1 h)}{(k-1)!}h^{k-1} \ ,$$

където $\theta_1 \in (0,1)$. Тъй като по условие $f^{(j)}(c) = 0$ за $j = 0, \ldots, k-1$, то

$$\frac{f(c+h)}{f'(c+h)} = \frac{f^{(k)}(c+\theta h)}{f^{(k)}(c+\theta_1 h)} \frac{h}{k} \ .$$

Но $f^{(k)}(t) \neq 0$. Тъй като $f^{(k)}(t)$ е непрекъсната функция, то съществува околност $\mathcal{U}_1$ на c такава, че $f^{(k)}(t) \neq 0$ за всяко $t \in \mathcal{U}_1$. Нещо повече, $sign\ f^{(k)}(t) = sign\ f^{(k)}(c)$ за всяко $t \in \mathcal{U}_1$. В частност, при достатъчно малки h ще имаме

$$sign\ f^{(k)}(c+\theta h) = sign\ f^{(k)}(c+\theta_1 h) \ .$$

Следователно

$$sign\ \frac{f(c+h)}{f'(c+h)} = sign\ h \ .$$

Оттук при $h = \varepsilon$ и $h = -\varepsilon$ получаваме твърдението на лемата.

Следващата теорема, която принадлежи на Щурм, дава точния брой на корените на едно алгебрично уравнение в даден интервал $[a,b]$. Преди да я формулираме, ще въведем някои стандартни означения. Нека $\alpha_0, \alpha_1, \ldots, \alpha_n$ е редица от реални числа. С $S^-(\alpha_0, \ldots, \alpha_n)$ ще означаваме броя на *силните смени* на знаците в редицата $\alpha_0, \alpha_1, \ldots, \alpha_n$. С други думи, това е броят на двойките от вида $(+,-)$ или $(-,+)$ в редицата, която се получава от $\alpha_0, \alpha_1, \ldots, \alpha_n$ като заместим всяко положително α_i с "+", всяко отрицателно с "–" и изпуснем нулевите членове на редицата. Например,

$$S^-(-5,6,4,0,-1,2) = 3 \ .$$

С $S^+(\alpha_0, \ldots, \alpha_n)$ ще означаваме броя на *слабите смени* на знаците в редицата $\alpha_0, \alpha_1, \ldots, \alpha_n$. Това е максималният брой смени, който се получава като заместим нулите в редицата $\alpha_0, \alpha_1, \ldots, \alpha_n$ с $+1$ или -1. Например,

$$S^+(-2,0,-1,4) = 3 \ .$$

Нека $f(x)$ е произволен алгебричен полином от степен точно n, т.е. $f(x) = a_0x^n + \ldots + a_n$ и $a_0 \neq 0$. Да приложим алгоритъма на Евклид за намиране на най-големия общ множител на $f(x)$ и $f'(x)$. Имаме

$$\begin{aligned}
f(x) &= f'(x)Q_0(x) - R_1(x) \\
f'(x) &= R_1(x)Q_1(x) - R_2(x) \\
R_1(x) &= R_2(x)Q_2(x) - R_3(x) \\
\dots\dots\dots & . \dots\dots\dots\dots\dots\dots\dots\dots\dots \\
R_{i-1}(x) &= R_i(x)Q_i(x) - R_{i+1}(x) \\
\dots\dots\dots & . \dots\dots\dots\dots\dots\dots\dots\dots\dots \\
R_{k-2}(x) &= R_{k-1}(x)Q_{k-1}(x) - R_k(x).
\end{aligned}$$

Тук за удобство сме записали остатъците от делението със знак ". Вижда се, че степените на полиномите $R_i(x)$, $i = 1, 2, \dots, k$, строго намаляват. Операцията деление се повтаря докато получим общия множител $R_k(x)$ (в който случай следващият остатък $R_{k+1}(x)$ е равен на нула за всяко x) или стигнем до остатък $R_k(x) = const \neq 0$. И така, алгоритъмът на Евклид ще произведе редицата $f(x), f'(x), R_1(x), \dots, R_k(x)$. Да изброим някои свойства на тази редица при предположение, че $f(x)$ няма кратни нули в $[a, b]$.

1) Ако $f(c) = 0$, то $f(c - \varepsilon)$ и $f'(c - \varepsilon)$ имат противни знаци, а $f(c + \varepsilon)$ и $f'(c + \varepsilon)$ имат еднакви знаци при всяко достатъчно малко $\varepsilon > 0$.

Това свойство следва от Лема 1.

2) Ако $R_i(c) = 0$ за някое i $(i = 0, 1, \dots, k-1)$, то $R_{i-1}(c) \neq 0$, $R_{i+1}(c) \neq 0$ и $R_{i-1}(c)$, $R_{i+1}(c)$ имат противни знаци. (Тук $R_{-1}(x) := f(x)$, $R_0(x) := f'(x)$).

Твърдението следва веднага от връзката $R_{i-1}(x) = R_i(x)Q_i(x) - R_{i+1}(x)$ при $x = c$. Тогава получаваме $R_{i-1}(c) = -R_{i+1}(c)$. Ако допуснем, че едното от тези две числа е нула, то поради рекурентната връзка ще получим, че $R_{i-1}(c) = R_{i-2}(c) = \dots = R_1(c) = f'(c) = f(c) = 0$. Следователно c е кратна нула на f, противоречие.

3) $R_k(x) \neq 0$ в $[a, b]$.

Това следва от условието, че $f(x)$ няма кратни нули в $[a, b]$. Наистина, по определение $R_k(x) = const \neq 0$ или $R_k(x)$ е общият множител на f и f'. Ако този общ множител се анулира в някаква точка c от $[a, b]$, то и $f(x)$, и $f'(x)$, щяха да се делят на $x - c$, което значи, че f има кратна нула в точката c от $[a, b]$. Но това противоречи на условието.

Редицата $f(x), f'(x), R_1(x), \dots, R_k(x)$ се нарича *редица на Щурм*.

Да означим, за краткост, с $S^-(x)$ броя на силните смени в редицата на

Щурм, т.е. $S^-(x) := S^-(f(x), f'(x), R_1(x), \dots, R_k(x))$.

Теорема на Щурм. *Нека $f(x)$ е произволен алгебричен полином от степен n, който няма кратни нули в $[a,b]$. Тогава броят на нулите на f в $[a,b]$ е точно равен на $S^-(a) - S^-(b)$.*

Доказателство. Ще проследим как се променя броят $S^-(x)$ на смените на знаците в редицата на Щурм когато x се движи от a към b. Тъй като всички функции в тази редица са алгебрични полиноми и следователно непрекъснати, то промяна в броя $S^-(x)$ може да настъпи само ако x премине през нула на някоя от функциите $f(x), f'(x), R_1(x), \dots, R_{k-1}(x)$. Да предположим, че $c \in [a,b]$ и $f(c) = 0$. Тогава, при достатъчно малки $\varepsilon > 0$, $f(c-\varepsilon)$ и $f'(c-\varepsilon)$ имат противни знаци, а $f(c+\varepsilon)$ и $f'(c+\varepsilon)$ имат еднакви знаци. Следователно между $f(x)$ и $f'(x)$ е имало смяна на знака малко преди c и тази смяна изчезва след c. С други думи, броят $S^-(x)$ намалява с 1 когато x преминава през нула на f.

Да видим сега какво става когато x премине през нула на $R_i(x)$ за някои $i = 0, \dots, k-1$. И така, нека $R_i(c) = 0$. Тогава, по свойство 2) на редицата на Щурм, $R_{i-1}(c) \neq 0$, $R_{i+1}(c) \neq 0$ и $R_{i-1}(c)R_{i+1}(c) < 0$. Но тогава $R_{i-1}(x)R_{i+1}(x) < 0$ за всяко x от достатъчно малка околност на c и следователно,

$$S^-(R_{i-1}(x), R_i(x), R_{i+1}(x)) = 1$$

за всяко x от тази околност. Това показва, че при преминаване на x през нула на някоя междинна функция от редицата на Щурм броят на смените $S^-(x)$ не се променя. И така, доказахме, че $S^-(x)$ намалява с 1 само при преминаване през нула на $f(x)$. Следователно броят на загубените смени когато x пробягва интервала $[a,b]$ е точно равен на броя на смените на f в $[a,b]$. Доказателството е завършено.

Да обърнем внимание, че доказателството на теоремата на Щурм се основава само върху свойствата 1) – 3) на редицата

$$f(x), R_0(x), R_1(x), \dots, R_k(x)\ .$$

Следователно ние бихме получили същия резултат за броя на нулите на f, ако разглеждахме призволна друга редица

$$f(x), P_0(x), P_1(x), \dots, P_k(x)\ , \qquad (2)$$

която удовлетворява условията 1) – 3). Такава редица се нарича *редица на Щурм*. И така, ако (2) е редица на Щурм в $[a,b]$, то броят на нулите на f в $[a,b]$ е точно равен на $S^-(a) - S^-(b)$.

Като приложим тази забележка ще покажем, че за произволен полином $f(x)$ (независимо дали има или няма кратни нули в $[a,b]$), числото $S^-(a) - S^-(b)$ е точно равно на броя на различните точки от $[a,b]$, в които

$f(x)$ се анулира. Наистина, ако $f(x)$ няма кратни нули, то това е всъщност твърдението от теоремата на Щурм. Нека f има кратни нули в $[a, b]$. Тогава f и f' имат общ множител $R_k(x)$, различен от константа. Този множител дели също така $R_1(x), \ldots, R_k(x)$. Тогава функциите

$$\frac{f(x)}{R_k(x)}, \ \frac{f'(x)}{R_k(x)}, \ \ldots, \ \frac{R_{k-1}(x)}{R_k(x)}, \ \frac{R_k(x)}{R_k(x)} \tag{3}$$

са определени в $[a, b]$ и удовлетворяват условията 1) – 3). Следователно (3) е редица на Щурм за функцията $f(x) := \frac{f(x)}{R_k(x)}$ и по теоремата на Щурм,

$$S := S^-\left(\frac{f(a)}{R_k(a)}, \ldots, 1\right) - S^-\left(\frac{f(b)}{R_k(b)}, \ldots, 1\right)$$

е броят на простите нули на $\frac{f(x)}{R_k(x)}$ в $[a, b]$, т.е. броят на нулите, в които $f(x)$ се анулира в $[a, b]$. Но тъй като броят на смените на знаците в редицата (3) е равен на броя на смените в редицата $f(x)$, $f'(x)$, $R_1(x), \ldots$, $R_k(x)$, то $S = S^-(a) - S^-(b)$. Следователно теоремата на Щурм, приложена за произволен полином f, дава броя на нулите на f в $[a, b]$ без да отчита кратностите им.

Едно от неудобствата на теоремата на Щурм е построяването на редицата, тъй като е свързано с делене на алгебрични полиноми. Сега ще приведем един друг резултат, който борави с по-лесно построими редици, но затова пък той не дава точния брой на нулите, а само една горна граница.

Да означим със $Z(f; (a, b))$ броя на нулите на f в (a, b), броейки кратностите.

Теорема на Бюдан-Фурие. *Нека $f(x)$ е алгебричен полином от степен точно n. Тогава*

$$\begin{aligned} Z(f; (a,b)) &= S^-\left(f(a), f'(a), f''(a), \ldots, f^{(n)}(a)\right) \\ &- S^+\left(f(b), f'(b), f''(b), \ldots, f^{(n)}(b)\right) \end{aligned}$$

или този брой е с четно число по-малък.

Доказателство. Да проследим какво става с изменението на броя на смените $S^-(f(x), f'(x), \ldots f^{(n)}(x)) =: V(x)$, когато x се движи от a към b. Ясно е, че изменение във $V(x)$ може да настъпи само когато x премине през нула на някоя от функциите $f(x)$, $f'(x), \ldots, f^{(n-1)}(x)$ ($f^{(n)}(x)$ е константа). Нека c е k-кратна нула на $f(x)$, т.е.

$$f(c) = f'(c) = \ldots = f^{(k-1)}(c) = 0, \ f^{(k)}(c) \neq 0 \ .$$

За определеност да приемем, че $f^{(k)}(c) > 0$. Тъй като $f^{(k)}(t)$ е непрекъсната функция, то $f^{(k)}(t) > 0$ за всяко t от някаква околност $\mathcal{U}$ на c. Тогава по

Лема 1, за достатъчно малки $\varepsilon > 0$ ще имаме

$$f^{(k-1)}(c-\varepsilon) < 0, \quad f^{(k-1)}(c+\varepsilon) > 0 .$$

Аналогично,

$$f^{(k-2)}(c-\varepsilon) > 0, \quad f^{(k-2)}(c+\varepsilon) > 0$$

и т.н. Следователно

$$\begin{aligned} S^-\left(f(c-\varepsilon), f'(c-\varepsilon), \ldots, f^{(k)}(c-\varepsilon)\right) &= k, \\ S^-\left(f(c+\varepsilon), f'(c+\varepsilon), \ldots, f^{(k)}(c+\varepsilon)\right) &= 0 . \end{aligned}$$

И така, *ако x премине през нула на f, броят $V(x)$ намалява точно с кратността на тази нула.*

Да предположим сега, че c е $k-$кратна нула на някоя от производните, но не е нула на f. Нека

$$f^{(i-1)}(c) \neq 0, \; f^{(i)}(c) = f^{(i+1)}(c) = \ldots = f^{(i+k-1)}(c) = 0, \; f^{(i+k)}(c) \neq 0$$

за някое $1 < i \leq n-k$. Въз основа на Лема 1, при достатъчно малки $\varepsilon > 0$,

$$\begin{aligned} R_1 &:= S^-\left(f^{(i-1)}(c-\varepsilon), f^{(i)}(c-\varepsilon), \ldots, f^{(i+k)}(c-\varepsilon)\right) \\ &= \begin{cases} k + S^-\left(f^{(i-1)}(c-\varepsilon), f^{(i)}(c-\varepsilon)\right), & \text{при четно } k, \\ k+1 - S^-\left(f^{(i-1)}(c-\varepsilon), f^{(i)}(c-\varepsilon)\right), & \text{при нечетно } k, \end{cases} \\ R_2 &:= S^-\left(f^{(i-1)}(c+\varepsilon), f^{(i)}(c+\varepsilon), \ldots, f^{(i+k)}(c+\varepsilon)\right) \\ &= S^-\left(f^{(i-1)}(c+\varepsilon), f^{(i+k)}(c+\varepsilon)\right) . \end{aligned}$$

Тъй като $f^{(i-1)}$ и $f^{(i+k)}$ са непрекъснати функции, то $f^{(i-1)}(t)$ и $f^{(i+k)}(t)$ са различни от нула в околност $\mathcal{U}$ на c. Следователно

$$S^-\left(f^{(i-1)}(t), f^{(i+k)}(t)\right) = \delta = const$$

за всяко t от $\mathcal{U}$, като $\delta = 1$ или $\delta = 0$, (т.е. между тези производни има или няма смяна на знака). Нека $\delta = 1$ и k е четно. Тогава

$$R_1 = k+1, \quad R_2 = 1 . \quad \text{Следователно } V(c-\varepsilon) - V(c+\varepsilon) = k ,$$

което е четно. Аналогично, при $\delta = 1$ и k – нечетно имаме:

$$R_1 = k, \quad R_2 = 1, \quad V(c-\varepsilon) - V(c+\varepsilon) = k-1 \quad (\text{четно}) .$$

При $\delta = 0$, $k -$ четно :

$$R_1 = k,\ R_2 = 0,\ V(c-\varepsilon) - V(c+\varepsilon) = k \qquad \text{(четно)} .$$

При $\delta = 0$, $k -$ нечетно :

$$R_1 = k+1,\ R_2 = 0,\ V(c-\varepsilon) - V(c+\varepsilon) = k+1 \ \text{(четно)} .$$

И така, *когато x премине през нула на $f^{(i)}$, $V(x)$ намалява винаги с четно число.*

Следователно при всяко достатъчно малко $\varepsilon > 0$,

$$Z(f;(a+\varepsilon, b-\varepsilon)) = V(a+\varepsilon) - V(b-\varepsilon)$$

или броят Z е с четно число по-малък. Но

$$\begin{aligned} \lim_{\varepsilon\to 0} V(a+\varepsilon) &= S^-\left(f(a), f'(a), \ldots, f^{(n)}(a)\right) \\ \lim_{\varepsilon\to 0} V(b-\varepsilon) &= S^+\left(f(b), f'(b), \ldots, f^{(n)}(b)\right) . \end{aligned}$$

Теоремата е доказана.

Тъй като

$$S^-\left(f(x), f'(x), \ldots, f^{(n)}(x)\right) \le S^+\left(f(x), f'(x), \ldots, f^{(n)}(x)\right) ,$$

то очевидно

$$Z(f;(a,b)) \le S^-\left(f(a), f'(a), \ldots, f^{(n)}(a)\right) - S^-\left(f(b), f'(b), \ldots, f^{(n)}(b)\right) ,$$

което е по-често срещаната (но по-слаба) форма на теоремата на Бюдан-Фурие.

Преди повече от 350 години, известният френски математик и философ Рене Декарт дава едно правило за оценка на броя на положителните нули на алгебричен полином чрез броя на смените на знаците в редицата от неговите коефициенти. Ние ще приведем по-долу това правило като следствие от теоремата на Бюдан-Фурие.

Правило на Декарт. *Нека*

$$f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n,\ a_0 \ne 0,\ a_n \ne 0 .$$

Тогава

$$Z(f;(0,\infty)) \le S^-(a_0, a_1, \ldots, a_n)$$

или Z е с четно число по-малко. С други думи, броят на положителните корени на уравнението $f(x) = 0$ е равен на броя на силните смени на знаците в редицата от неговите коефициенти или този брой е с четно число по-малък.

Доказателство. Тъй като $f(x)$ е алгебричен полином от фиксирана степен n, то $f, f', \dots, f^{(n-1)}$ имат краен брой положителни нули. Тогава можем да изберем число $M > 0$ такова, че никой от полиномите $f, f', \dots, f^{(n-1)}$ да не се анулира за $x \geq M$. Следователно

$$Z(f;(0,\infty)) = Z(f;(0,M)) \ . \tag{4}$$

Но по теоремата на Бюдан-Фурие,

$$Z(f;(0,M)) \leq S^-\left(f(0), f'(0), \dots, f^n(0)\right) - S^-\left(f(M), f'(M), \dots, f^n(M)\right) . \tag{5}$$

Тъй като $f(x) \neq 0$ за $M \leq x < \infty$, то

$$sign\ f(M) = sign\ \left(\lim_{x\to\infty} f(x)\right) = sign\ a_0 \ .$$

Последното равенство следва от факта, че знакът на $f(x)$, при много големи $x > 0$, се определя от знака на $a_0 x^n$, който е всъщност знака на a_0. Аналогично имаме

$$\begin{aligned} sign\ f^{(k)}(M) &= sign\ \left(\lim_{x\to\infty} f^{(k)}(x)\right) \\ &= sign\ n(n-1)\dots(n-k+1)a_0 x^{n-k} = sign\ a_0 \end{aligned}$$

за всяко $k = 1, 2, \dots, n-1$. При $k = n$, очевидно $sign\ f^{(n)}(x) = sign\ n!a_0 = sign\ a_0$. Следователно всички числа в редицата

$$f(M), f'(M), \dots, f^{(n)}(M)$$

имат един и същ знак – знака на водещия коефициент a_0. Тогава

$$S^-\left(f(M), f'(M), \dots, f^{(n)}(M)\right) = 0 \ . \tag{6}$$

По-нататък, като вземем предвид, че $f^{(k)}(0) = a_{n-k}\, k!$ за $k = 0, \dots, n$, получаваме

$$S^-\left(f(0), f'(0), \dots, f^{(n)}(0)\right) = S^-(a_n, a_{n-1}, \dots, a_0) = S^-(a_0, \dots, a_n) \ . \tag{7}$$

Тогава, от (4) – (7) следва

$$Z(f;(0,\infty)) \leq S^-(a_0, a_1, \dots, a_n) \ ,$$

което трябваше да докажем. По теоремата на Бюдан-Фурие, разликата в двете страни на (5), а оттук и в последното неравенство, е четно число. Правилото на Декарт е доказано.

21. МЕТОД НА СВИВАЩИТЕ ИЗОБРАЖЕНИЯ

Голяма част от методите за приближено пресмятане на корените на уравнения са *итерационни.* При тях се тръгва от някакво начално приближение x_0 и след това с извършването на определена числена процедура (*итерация*) се намира следващото приближение x_1. Въз основа на x_1 и x_0 се определя x_2 и т.н. Построява се една редица $x_0, x_1, x_2, \dots, x_n$, която клони към корена ξ на уравнението $f(x) = 0$. При достатъчно големи n числото x_n е приближение на корена ξ със зададена точност ε. Ние ще разгледаме тук един клас от итерационни методи, които се базират на така наречения *метод на свиващите изображения.*

Нека $f(x)$ е функция, определена в $[a, b]$. Ще изследваме уравнението $f(x) = 0$. За нас ще бъде удобно да запишем това уравнение във вида

$$x = \varphi(x) \ .$$

Това може да стане, например, като добавим x към двете страни на уравнението $f(x) = 0$ или направим друго, еквивалентно преобразование. Ако ξ е корен на уравнението $f(x) = 0$, то очевидно $\xi = \varphi(\xi)$. Да изберем точка x_0 от $[a, b]$ и да построим редицата

$$x_0, x_1, x_2, \dots, x_n, \dots$$

по правилото

$$x_n = \varphi(x_{n-1}), \quad n = 1, 2, \dots \tag{1}$$

Нашата цел е да построим редица $\{x_n\}$, която клони към корена ξ на уравнението $x = \varphi(x)$. Ясно е, че правилото (1) не поражда такава редица за произволна функция φ. Има обаче един клас от уравнения (т.е. от функции φ), при които простото итерационно правило (1) наистина дава редица $\{x_n\}$, която клони към корена ξ. Сега да видим, какви условия върху φ биха гарантирали такава сходимост.

Първо, трябва да можем да построим редицата $\{x_n\}$. За целта, всяка следваща точка от редицата трябва да принадлежи на дефиниционната област $[a, b]$ на φ. Това очевидно ще бъде изпълнено, ако

У.1. $\varphi(x) \in [a, b]$ за всяко $x \in [a, b]$.

Наистина, ако φ удовлетворява У.1 и изберем произволно *начално приближение* x_0 от $[a, b]$, то $x_1 = \varphi(x_0)$ ще принадлежи също на $[a, b]$. Оттук $x_2 = \varphi(x_1) \in [a, b]$ и т.н. Условието У.1 показва, че φ е едно изображение на интервала $[a, b]$ в себе си.

И така, доказахме

Лема 1. *Ако φ е изображение на $[a, b]$ в себе си, то при произволно начално приближение x_0 от $[a, b]$, всички останали точки от редицата $\{x_n\}$ принадлежат също на $[a, b]$.*

Ние търсим корена на уравнението $x = \varphi(x)$, т.е. търсим точка ξ от $[a, b]$, за която $\xi = \varphi(\xi)$. Точката ξ е *неподвижна точка* при изображението φ. Следващото просто условие върху φ гарантира поне една неподвижна точка.

У.2. φ е непрекъснато изображение на интервала $[a, b]$ в себе си.

Наистина, нека φ е непрекъсната функция, която удовлетворява У.1 (т.е. φ е непрекъснато изображение на $[a, b]$ в себе си). Ако $a = \varphi(a)$, то a е неподвижна точка. Аналогично, ако $b = \varphi(b)$, то b е неподвижна точка. Да допуснем, че $a \neq \varphi(a)$ и $b \neq \varphi(b)$. Тъй като φ е изображение на $[a, b]$ в себе си, то $\varphi(a) \in [a, b]$, $\varphi(b) \in [a, b]$ и следователно

$$a < \varphi(a), \quad \varphi(b) < b \ .$$

Да образуваме функцията $r(x) := x - \varphi(x)$. Тя е непрекъсната в $[a, b]$ и

$$r(a) = a - \varphi(a) < 0, \quad r(b) = b - \varphi(b) > 0$$

Следователно съществува точка ξ от $[a, b]$ такава, че $r(\xi) = 0$, т.е. $\xi = \varphi(\xi)$. Да формулираме ясно получения резултат.

Лема 2. *Ако φ е непрекъснато изображение на интервала $[a, b]$ в себе си, то φ има неподвижна точка в $[a, b]$.*

Това е твърде частен случай от известна теорема от топологията, според която всяко непрекъснато изображение на едно изпъкнало множество Ω от $\mathbb{R}^n$ в себе си има неподвижна точка.

Остава да видим какви условия върху φ ще гарантират сходимост на редицата $\{x_n\}$ към неподвижната точка ξ.

Казваме, че функцията g удовлетворява условието на Липшиц с константа q в $[a, b]$, ако

$$|g(x) - g(y)| \leq q \, |x - y| \quad \text{за всяко} \quad x, y \in [a, b] \ .$$

Теорема 3. *Нека φ е непрекъснато изображение на $[a, b]$ в себе си, което удовлетворява условието на Липшиц с константа $q < 1$. Тогава*

а) Уравнението $x = \varphi(x)$ има единствен корен ξ в $[a, b]$;

б) Редицата $\{x_n\}$ клони към ξ при $n \to \infty$.

Нещо повече,

$$|x_n - \xi| \leq (b - a)\, q^n \quad \text{за всяко } n \ . \tag{3}$$

Доказателство. По Лема 2, φ има поне една неподвижна точка. Да допуснем, че те са повече. Нека $\xi_1 = \varphi(\xi_1)$ и $\xi_2 = \varphi(\xi_2)$ за някои ξ_1, ξ_2 от $[a, b]$. Тогава при $\xi_1 \neq \xi_2$,

$$\begin{aligned}
|\xi_1 - \xi_2| &= |\varphi(\xi_1) - \varphi(\xi_2)| \\
&\leq q\,|\xi_1 - \xi_2| \quad \text{(по условието на Липшиц)} \\
&< |\xi_1 - \xi_2| \quad \text{(защото } q < 1) \ .
\end{aligned}$$

Стигнахме до абсурд. Следователно $\xi_1 = \xi_2$. Единствеността е доказана. Сега ще докажем оценката (3), от която очевидно следва б). Имаме

$$\begin{aligned}
|x_n - \xi| &= |\varphi(x_{n-1}) - \varphi(\xi)| \leq q\,|x_{n-1} - \xi| \\
&= q\,|\varphi(x_{n-2}) - \varphi(\xi)| \leq q^2\,|x_{n-2} - \xi| \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\leq q^n\,|x_0 - \xi| \ .
\end{aligned}$$

Тъй като $x_0 \in [a, b]$ и $\xi \in [a, b]$, то $|x_0 - \xi| < b - a$. Оценката (3), а с това и теоремата, е доказана.

Изображение φ, което удовлетворява условието на Липшиц с константа по-малка от 1, се нарича *свиващо изображение.* При него разстоянието между образите $\varphi(x)$ и $\varphi(y)$ е строго по-малко от разстоянието между праобразите x и y (т.е. φ "свива" разстоянията). От теоремата за крайните нараствания следва, че ако φ е диференцируема функция в $[a, b]$ и $|\varphi'(x)| \leq q < 1$ за всяко $x \in [a, b]$, то φ е свиващо изображение.

Наистина, по теоремата за крайните нараствания,

$$\varphi(x) - \varphi(y) = \varphi'(\eta)(x - y)$$

при някакво η между x и y. Тогава

$$|\varphi(x) - \varphi(y)| = |\varphi'(\eta)|\,|(x - y)| \leq q|x - y| \quad (q < 1),$$

т.е. φ е свиващо изображение.

Нека уравнението $x = \varphi(x)$ има корен ξ в $[a, b]$. Казваме, че *итерационният процес*, породен от функцията φ е сходящ в $[a, b]$, ако при всяко начално приближение x_0 от $[a, b]$, редицата $\{x_n\}$, построена по формулата $x_n = \varphi(x_{n-1})$, $n = 1, 2, \ldots$, е сходяща към корена ξ. Теорема 3 представя метода на свиващите изображения за построяване на сходящи итерационни процеси. Сега ще приведем една по-слаба форма на тази теорема, която често се използва.

Следствие 4. *Нека ξ е корен на уравнението $x = \varphi(x)$. Да предположим, че φ има непрекъсната производна в околност $\mathcal{U}$ на ξ и $|\varphi'(\xi)| < 1$. Тогава при достатъчно добро начално приближение x_0 итерационният процес, породен от φ, е сходящ. Нещо повече, съществуват константи $C > 0$ и $0 < q < 1$ такива, че*

$$|x_n - \xi| \le C\, q^n \quad \text{за всяко } n \; .$$

Доказателство. Тъй като $\varphi'(t)$ е непрекъсната функция в $\mathcal{U}$ и $|\varphi'(\xi)| < 1$, то съществуват $q < 1$ и $\varepsilon > 0$ такива, че

$$|\varphi'(t)| \le q \quad \text{за всяко } t \in [\xi - \varepsilon, \xi + \varepsilon] \; .$$

Освен това, при $t \in [\xi - \varepsilon, \xi + \varepsilon]$ имаме

$$|\varphi(t) - \xi| \le q\, |t - \xi| \le q\varepsilon < \varepsilon \; ,$$

т.е. $\varphi(t) \in [\xi - \varepsilon, \xi + \varepsilon]$. Следователно φ е свиващо изображение на интервала $[\xi - \varepsilon, \xi + \varepsilon]$ в себе си. Тогава всички твърдения на следствието следват от доказаната вече Теорема 3.

На Рис. 15 е дадена геометрична илюстрация на метода на свиващите изображения.

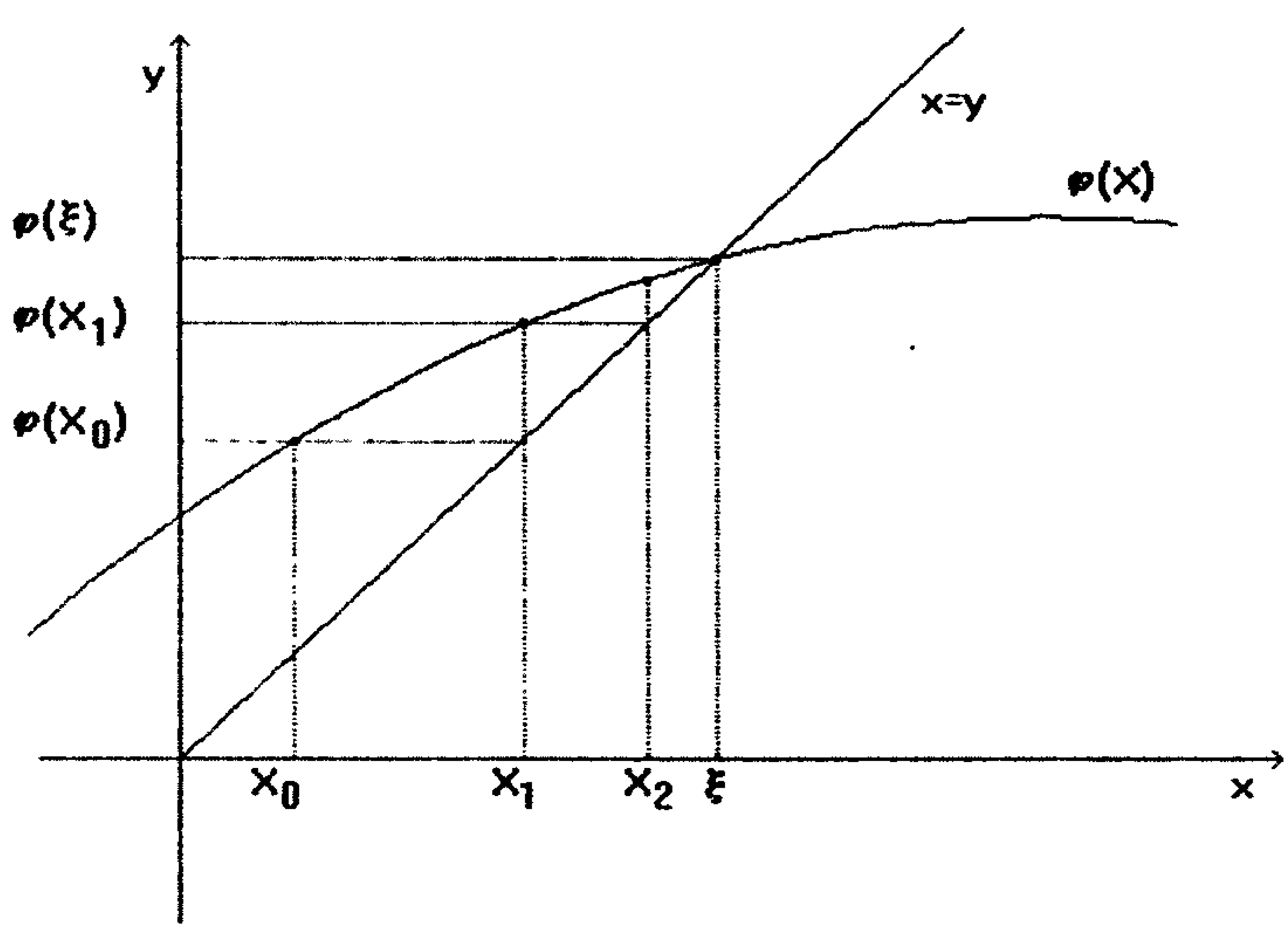

Рис. 15

Скоростта на сходимост в (3) се определя от общия член q^n на една геометрична прогресия. Затова е прието да се казва, че съответният итерационен процес е сходящ със *скорост на геометрична прогресия.* Това е доста бърза сходимост. Например, при $q = \frac{1}{2}$ и $n = 10$ получаваме точност от порядъка на 10^{-3}. Има обаче процеси, които са много по-бързо сходящи. За да характеризираме тяхната скорост, ще въведем понятието *ред на сходимост*.

Определение. Казваме, че итерационният процес $x_0, x_1, \ldots$ има *ред на сходимост* p, ако съществуват положителни константи C и $q < 1$ такива, че

$$|x_n - \xi| \leq Cq^{p^n} \quad \text{за всяко } n \ .$$

Следващата теорема ни дава един начин за определяне реда на сходимост на итерационния процес, породен от функцията φ.

Теорема 5. *Нека φ има непрекъснати производни до p-тата включително в околност на точката ξ. Нека*

$$\varphi(\xi) = \xi, \ \varphi'(\xi) = \ldots = \varphi^{(p-1)}(\xi) = 0, \quad \varphi^{(p)}(\xi) \neq 0 \ .$$

Тогава, при достатъчно добро начално приближение x_0, итерационният процес, породен от φ, има ред на сходимост p.

Доказателство. По формулата на Тейлър

$$\varphi(x) = \varphi(\xi) + \frac{\varphi'(\xi)}{1!}(x-\xi) + \ldots + \frac{\varphi^{(p-1)}(\xi)}{(p-1)!}(x-\xi)^{p-1} + \frac{\varphi^{(p)}(\xi + \theta(x-\xi))}{p!}(x-\xi)^p \ ,$$

където $|\theta| < 1$. Тъй като $\varphi^{(j)}(\xi) = 0$ за $j = 1, \ldots, p-1$, то

$$\varphi(x) - \varphi(\xi) = \frac{\varphi^{(p)}(\xi + \theta(x-\xi))}{p!}(x-\xi)^p \ .$$

Следователно, при всяко x от достатъчно малка околност $\mathcal{U}$ на ξ,

$$|\varphi(x) - \varphi(\xi)| \leq M\ |x - \xi|^p \ ,$$

където $M := \max_{t \in \mathcal{U}} \left|\varphi^{(p)}(t)\right| / p!$. Специално при $x = x_n$ имаме

$$\begin{aligned}
|x_{n+1} - \xi| &= |\varphi(x_n) - \varphi(\xi)| \leq M\ |x_n - \xi|^p \\
&= M\ |\varphi(x_{n-1}) - \varphi(\xi)|^p \leq M\ \{\ M\ |x_{n-1} - \xi|^p\ \}^p \\
&= M^{1+p}|x_{n-1} - \xi|^{p^2} \leq M^{1+p+p^2}|x_{n-2} - \xi|^{p^3} \leq \ldots \\
&\leq M^{1+p+\ldots+p^n}|x_0 - \xi|^{p^{n+1}} \\
&= M^{\frac{p^{n+1}-1}{p-1}}|x_0 - \xi|^{p^{n+1}} = M^{\frac{1}{1-p}}.\left\{M^{\frac{1}{p-1}}|x_0 - \xi|\right\}^{p^{n+1}}
\end{aligned}$$

Когато x_0 е достатъчно близко до ξ, $M^{1/(p-1)}|x_0 - \xi| < q < 1$ и следователно

$$|x_{n+1} - \xi| \leq c\, q^{p^{n+1}} \quad \text{за всяко } n,$$

където сме положили $C = M^{1/(1-p)}$. Доказателството е завършено.

По-долу ще разгледаме някои класически итерационни методи за приближено решаване на уравнения.

I. Метод на хордите.

Нека $[a, b]$ е даден краен интервал и $f(x)$ е два пъти диференцируема в него функция, която удовлетворява условията:

а) $f(a)\,f(b) < 0$,

б) $f'(x)\,f''(x) \neq 0$ за всяко x от $[a, b]$.

Не е трудно да се види, че тези условия осигуряват съществуването на единствен корен ξ на уравнението $f(x) = 0$ в $[a, b]$.

Наистина, първото условие гарантира съществуването на точка $\xi \in (a, b)$ такава, че $f(\xi) = 0$. От второто условие следва, че $f'(x)$ и $f''(x)$ не се анулират в $[a, b]$. Следователно $f'(x)$ и $f''(x)$ имат постоянен знак в $[a, b]$. Това показва, че $f(x)$ е строго монотонна функция, при това изпъкнала (ако $f''(x) > 0$) или вдлъбната (ако $f''(x) < 0$). Но една монотонна функция може да пресече оста x само в една точка. Единствеността на ξ е доказана.

Методът на хордите е итерационен процес, в който се построява редица от последователни приближения $x_0, x_1, \ldots$ на корена ξ на уравнението $f(x) = 0$ по следния начин:

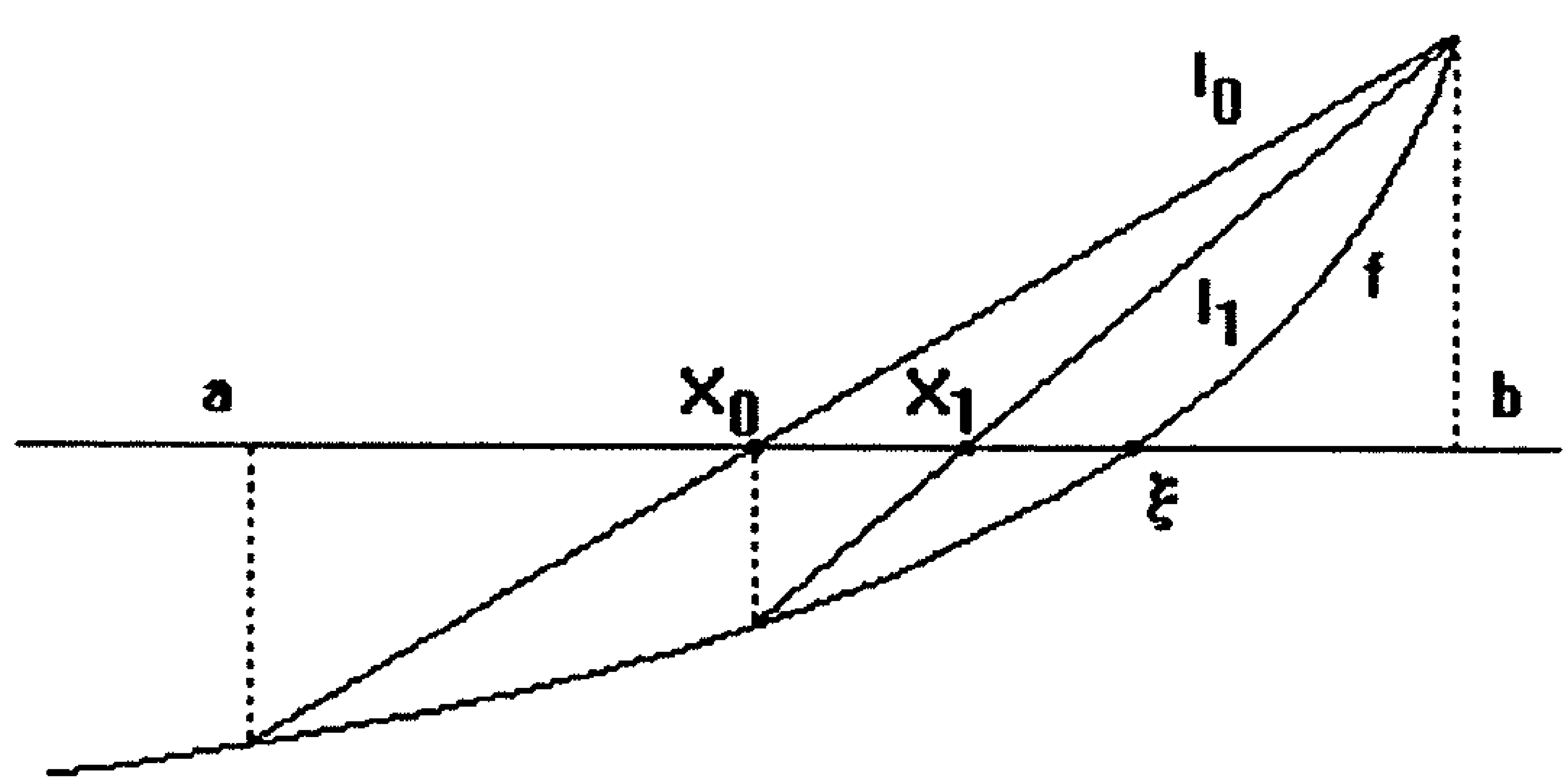

Рис. 16

Прекарва се права линия l_0, която минава през точките $(a, f(a))$ и $(b, f(b))$ (т.е. *хордата* към "дъгата " от графиката на функцията f в $[a, b]$, виж Рис. 16). Тя пресича оста x в някаква точка x_0. Това е началното приближение.

Естествено, x_0 лежи отляво на корена ξ, ако f е изпъкнала и отдясно на ξ, ако f е вдлъбната. На нашия пример $x_0 < \xi$. След това намираме следващото приближение x_1 като пресечна точка на оста x с хордата l_1, свързваща $(x_0, f(x_0))$ и $(b, f(b))$ (съответно $(x_0, f(x_0))$ и $(a, f(a))$, ако $f''(x) < 0$) и т.н. Изобщо x_{n+1} се получава като пресечна точка на оста x с хордата l_{n+1}, свързваща точките $(x_n, f(x_n))$ и $(b, f(b))$ (съответно $(a, f(a))$). Методът е илюстриран геометрично на Рис. 16.

Да намерим аналитичен израз за x_{n+1} чрез предишното приближение x_n. За определеност да смятаме, че $f''(x) > 0$ (както е на нашата Рис. 16). Правата l_{n+1} има уравнение

$$l_{n+1} = f(x_n)\,\frac{x-b}{x_n-b} + f(b)\,\frac{x-x_n}{b-x_n} = f(x_n) + f[x_n, b](x - x_n) .$$

Приближението x_{n+1} е корен на уравнението $l_{n+1}(x) = 0$. Следователно

$$x_{n+1} = x_n - \frac{f(x_n)}{f[x_n, b]} ,$$

или окончателно,

$$x_{n+1} = x_n - \frac{f(x_n)}{f(b) - f(x_n)}(b - x_n) . \tag{4}$$

Това е известната формула за намиране на последователните приближения на корена ξ по метода на хордите.

Сега ще покажем, че x_n наистина клони към ξ при $n \to \infty$. Използвайки изпъкналостта на f, може да се види, че $x_0, x_1, \ldots$ е монотонна и ограничена редица, следователно сходяща. Нека α е нейна граница. Тогава, като извършим граничен преход в (4), получаваме

$$\alpha = \alpha - \frac{f(\alpha)}{f(b) - f(\alpha)}(b - \alpha) , \quad \text{т.е.} \quad f(\alpha) = 0 .$$

Следователно $\alpha = \xi$ и сходимостта на x_n към ξ е доказана. Ние обаче ще използваме Следствие 4 от общата теория на метода на свиващите изображения, защото то ще ни даде и оценка за скоростта на сходимост. И така, от (4) е ясно, че методът на хордите е итерационен процес, породен от функцията

$$\varphi(x) = x - \frac{f(x)}{f(b) - f(x)}(b - x) .$$

Вижда се, че уравнението $x = \varphi(x)$ е еквивалентно с $f(x) = 0$. Понеже ще прилагаме Следствие 4 към φ, да намерим $\varphi'(\xi)$. Имаме

$$\varphi'(\xi) = 1 - f'(\xi)\left[\frac{b-\xi}{f(b) - f(\xi)}\right] - f(\xi)\left\{\frac{b-x}{f(b) - f(x)}\right\}'\Bigg|_{x=\xi} .$$

Тъй като $f(\xi) = 0$, то

$$\varphi'(\xi) = 1 - f'(\xi)\,\frac{b-\xi}{f(b)} = \frac{f(b) - f'(\xi)(b-\xi)}{f(b)} \ .$$

Като заместим $f(b)$ по формулата на Тейлър с

$$\begin{aligned} f(b) &= f(\xi) + f'(\xi)(b-\xi) + \frac{f''(\eta_1)}{2}(b-\xi)^2 \quad \text{(в числителя)}, \\ f(b) &= f(\xi) + f'(\eta_2)(b-\xi) \qquad\qquad\qquad\qquad \text{(в знаменателя)}, \end{aligned}$$

където η_1 и η_2 са някакви точки от (a, b), получаваме

$$\varphi'(\xi) = \frac{f''(\eta_1)(b-\xi)}{2f'(\eta_2)}.$$

Нека

$$M := \max_{t\in[a,b]} |f''(t)|, \qquad m := \min_{t\in[a,b]} |f'(t)| \ .$$

Тъй като по условие $f'(t) > 0$ в $[a, b]$, то $m > 0$. Тогава

$$|\varphi'(\xi)| \le \frac{M}{2m}|b-\xi|$$

и очевидно $|\varphi'(\xi)|$ може да стане по-малко от произволно, отнапред избрано $q < 1$, стига $b - \xi$ да е достатъчно малко, т.е. стига интервалът $[a, b]$ да е достатъчно малък. И така, ако сме отделили корена ξ в достатъчно малък интервал $[a, b]$, то

$$|\varphi'(\xi)| < q < 1 \ .$$

Оттук, по Следствие 4, итерационният процес породен от φ (т.е. методът на хордите) е сходящ със скорост на геометрична прогресия,

$$|x_n - \xi| \le const.\ q^n.$$

II. Метод на секущите.

Ще предполагаме, че f удовлетворява условията а), б) от предишния пример. При *метода на секущите* всяко следващо приближение x_{n+1} на корена ξ на уравнението $f(x) = 0$ се построява въз основа на предишните две приближения x_n и x_{n-1}. Избираме $x_0 = a$ или $x_0 = b$ така, че да бъде изпълнено условието $f(x_0)f''(x_0) > 0$. На Рис. 17 например $x_0 = b$. След това избираме точка x_1 такава, че $\xi < x_1 < x_0$. Ние не знаем ξ (това е коренът ξ, който търсим). Тогава как да разберем, че някаква точка x_1

удовлетворява горното условие? Това става чрез сравняване на знаците на $f(x_0)$ и $f(x_1)$. Ако $f(x_1) = 0$, то всъщност $x_1 = \xi$ и задачата е решена. Ако $f(x_1)f(x_0) > 0$, то x_0 и x_1 са от една и съща страна на ξ и нашето изискване е изпълнено. Ако $f(x_1)f(x_0) < 0$, то x_0 и x_1 са от различни страни на ξ и x_1 не удовлетворява наложеното изискване. В този случай изчисляването на $f(x_1)$ не е отишло напразно, защото сме локализирали корена ξ в интервала $[x_1, x_0]$, който е по-малък от първоначалния $[a, b]$. По-нататък можем да използваме именно този интервал, вместо $[a, b]$. След избора на x_0 и x_1, посторяваме следващото приближение x_2 като пресечна точка на *секущата* l_1, минаваща през точките $(x_0, f(x_0))$ и $(x_1, f(x_1))$ и оста x (т.е. нулата на $l_1(x)$). Следващата точка x_3 е нула на секущата l_2 през $(x_1, f(x_1)), (x_2, f(x_2))$ и т.н., x_{n+1} е нула на секущата l_n през $(x_n, f(x_n)), (x_{n-1}, f(x_{n-1}))$. Алгоритъмът за построяване на $\{x_n\}$ е показан на Рис. 17.

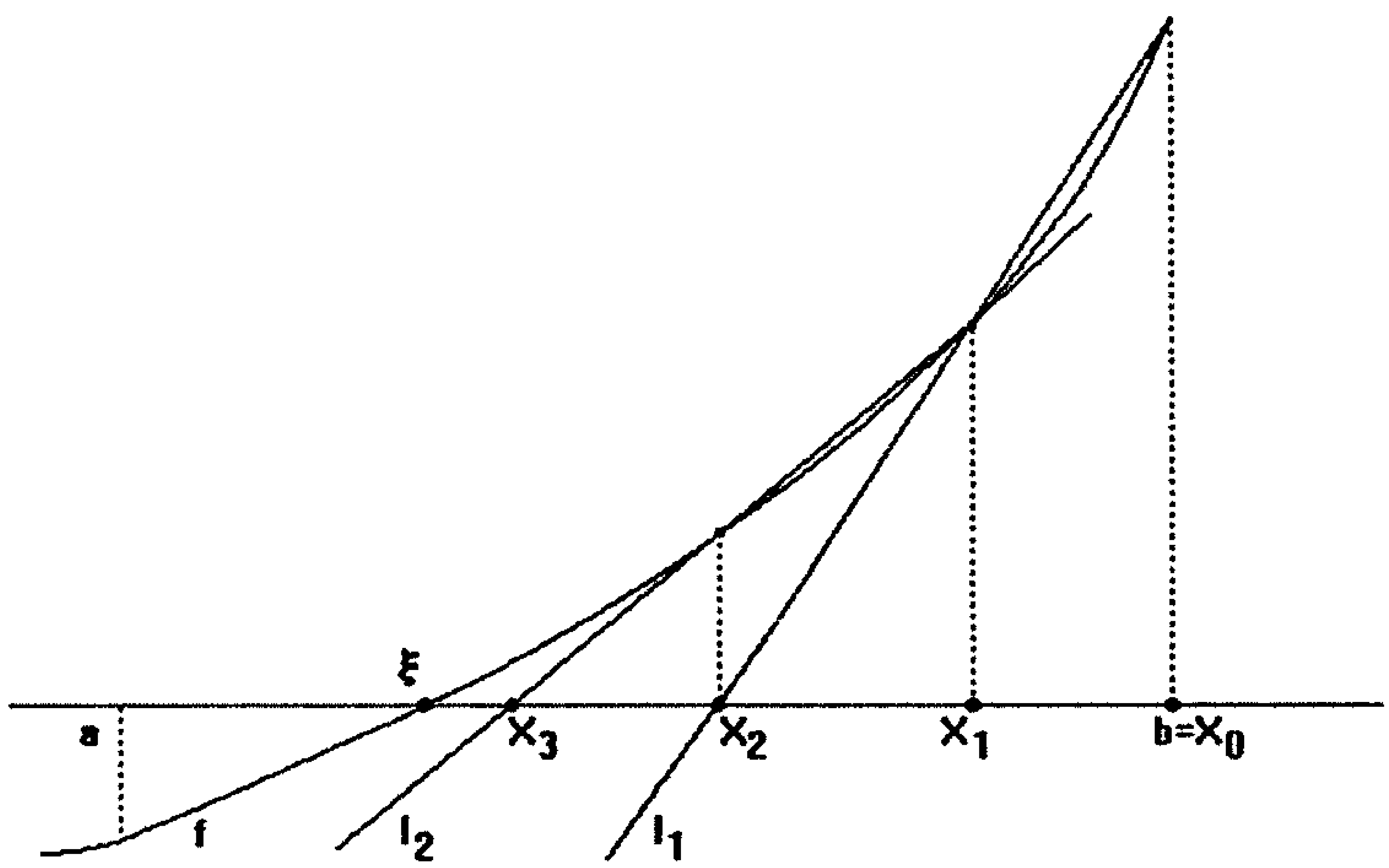

Рис. 17

Да намерим аналитичен израз за x_{n+1} чрез x_n и x_{n-1}. По формулата на Нютон

$$l_n(x) = f(x_n) + f[x_{n-1}, x_n](x - x_n)$$

и следователно x_{n+1} се определя от уравнението

$$f(x_n) + f[x_{n-1}, x_n](x_{n+1} - x_n) = 0 .$$

Оттук получаваме

$$x_{n+1} = x_n - \frac{f(x_n)}{f[x_{n-1}, x_n]} = x_n - \frac{f(x_n)}{f(x_{n-1}) - f(x_n)}(x_{n-1} - x_n) .$$

Сега ще покажем сходимостта на x_n към ξ при $n \to \infty$ и ще намерим реда на сходимост.

Както и в предишния пример, ще използваме означенията

$$M := \max_{t\in[a,b]} |f''(t)|, \qquad m := \min_{t\in[a,b]} |f'(t)| .$$

Теорема 6. *Нека* $\{x_n\}_0^\infty$ *е редицата от последователни приближения по метода на секущите. Да предположим, че началните приближения* x_0 *и* x_1 *удовлетворяват условието*

$$|x_0 - \xi| \le Cq^{r^0}, \qquad |x_1 - \xi| \le Cq^{r^1} ,$$

където $0 < q < 1$, C *е константа такава, че* $\frac{M}{2m}C < 1$ *и* $r = \frac{1+\sqrt{5}}{2}$. *Тогава*

$$|x_n - \xi| \le Cq^{r^n} \qquad \text{за всяко } n . \tag{5}$$

Доказателство. Ще приложим индукция по n. За $n = 0$ и $n = 1$ оценката (5) е вярна по условие. Да допуснем, че (5) е в сила за всяко естествено число $\le n$. Ще я докажем за $n+1$. За целта да представим $f(x)$ по формулата на Лагранж във вида

$$f(x) = l_n(x) + \frac{f''(\eta)}{2!}(x - x_{n-1})(x - x_n) \qquad (\eta \in [a,b])$$

и по формулата на Тейлър

$$f(x) = f(\xi) + f'(\eta_1)(x - \xi) \qquad\qquad (\eta_1 \in [a,b]) .$$

Като приравним двата израза при $x = x_{n+1}$ и вземем предвид, че $f(\xi) = 0$ и $l_n(x_{n+1}) = 0$, получаваме

$$|f'(\eta_1)|\,|x_{n+1} - \xi| = \left|\frac{f''(\eta)}{2}\right|\,|x_{n+1} - x_{n-1}|\,|x_{n+1} - x_n| .$$

Оттук

$$\begin{aligned} |x_{n+1} - \xi| &\le \frac{M}{2m}|x_{n+1} - x_{n-1}|\,|x_{n+1} - x_n| \\ &\le \frac{M}{2m}|x_{n-1} - \xi|\,|x_n - \xi| \quad (\text{защото } \xi < x_{n+1} < x_n < x_{n-1}) . \end{aligned}$$

Но съгласно индукционото предположение,

$$\begin{aligned} |x_{n-1} - \xi| &\le Cq^{r^{n-1}} , \\ |x_n - \xi| &\le Cq^{r^n} . \end{aligned}$$

Следователно

$$
\begin{aligned}
|x_{n+1}-\xi| &\le \frac{M}{2m} Cq^{r^{n-1}} Cq^{r^n} = \frac{MC}{2m} Cq^{r^{n-1}+r^n} \\
&< Cq^{r^{n-1}(1+r)} \qquad \left(\text{защото } \frac{MC}{2m} < 1 \text{ по условие}\right) .
\end{aligned}
$$

Но r е положителният корен на уравнението $r^2 - r - 1 = 0$. Следователно $r + 1 = r^2$ и тогава $r^{n-1}(1 + r) = r^{n+1}$. Горното неравенство добива вида

$$|x_{n+1} - \xi| \le Cq^{r^{n+1}} ,$$

което трябваше да докажем. С това доказателството е завършено.

Да обърнем внимание на факта, че $r = (1+\sqrt{5})/2 \approx 1,618$. Следователно методът на секущите е значително по-бързо сходящ от метода на хордите. При това формулата за пресмятане на x_{n+1} не е по-сложна от съответната формула при метода на хордите. И двата метода изискват изчисляването на само една нова стойност на f на всяка стъпка.

По-долу ще се запознаем с един друг метод, който е по-бързо сходящ и от метода на секущите.

III. Метод на Нютон (метод на допирателните).

И тук ще изискваме изпълнението на условията а), б). Избираме начално приближение $x_0 = a$ или $x_0 = b$ така, че да имаме $f(x_0)\, f''(x_0) > 0$. Следващото приближение x_1 се намира като пресечна точка на оста x с *допирателната* d_0 към правата $y = f(x)$ в точката x_0 (виж Рис. 18). След това намираме x_2 като нула на допирателната d_1 към f в x_1 и т.н., x_{n+1} е нулата на допирателната d_n към f в точката x_n.

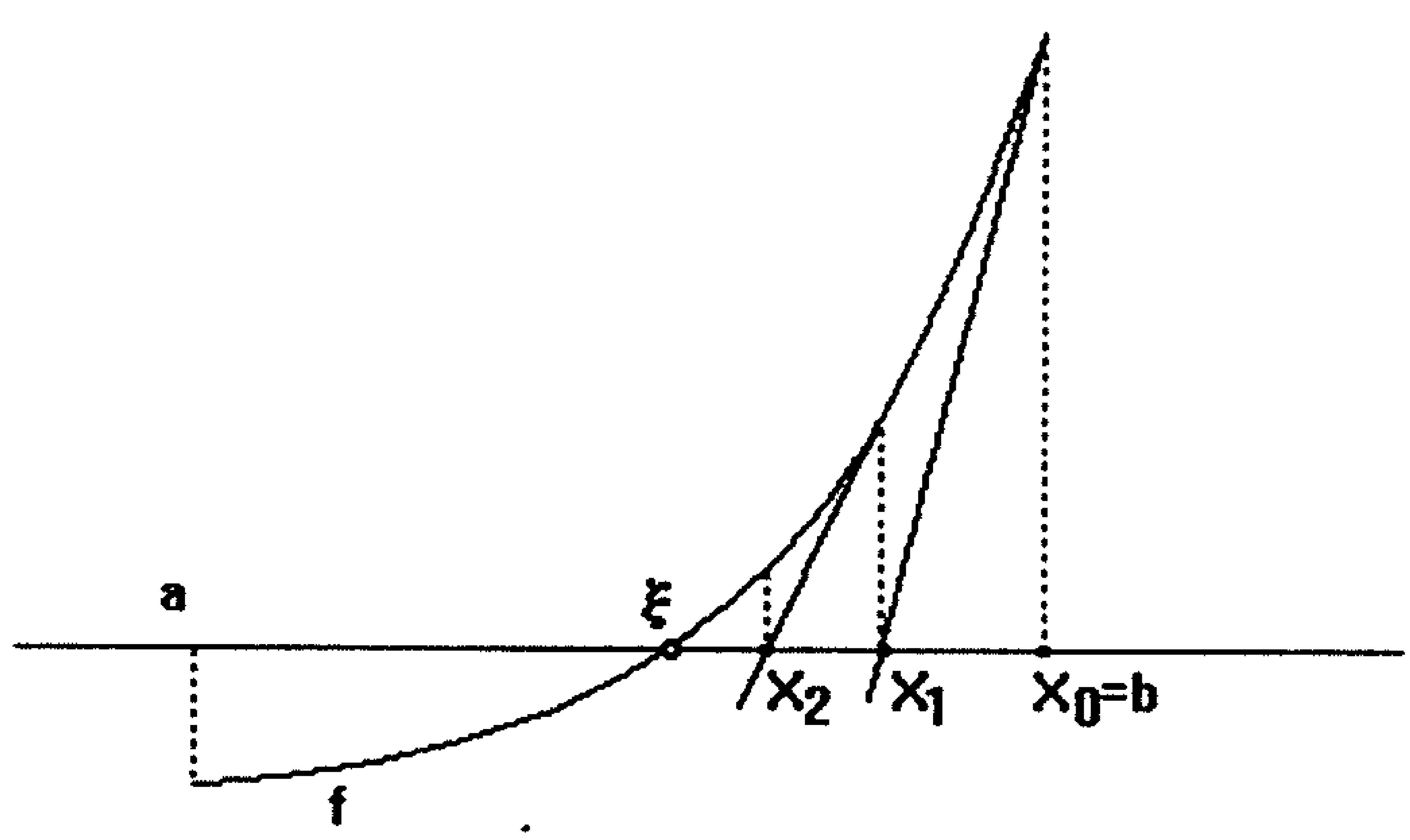

Рис. 18

Да намерим формулата за x_{n+1}. Имаме

$$l_n(x) = f(x_n) + f'(x_n)(x - x_n) \ .$$

Следователно x_{n+1} е решение на линейното уравнение

$$f(x_n) + f'(x_n)(x - x_n) = 0 \ .$$

Оттук получаваме

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \ .$$

Това е известната *формула на Нютон* за приближено пресмятане на корена на уравнението $f(x) = 0$.

За доказателство на сходимостта на метода ще използваме Теорема 5. Ясно е, че x_{n+1} се получава по формулата $x_{n+1} = \varphi(x_n)$ с

$$\varphi(x) = x - \frac{f(x)}{f'(x)} \ .$$

За $\varphi'(\xi)$ получаваме

$$\varphi'(\xi) = 1 - \frac{f'^2(\xi) - f(\xi)f''(\xi)}{f'^2(\xi)} = 0 \qquad (\text{защото } f(\xi) = 0) \ .$$

Може да се провери, че в общия случай $\varphi''(\xi) \neq 0$. Следователно, по Теорема 5, итерационният процес породен от φ (т.е. методът на Нютон) е сходящ и има ред на сходимост 2 при всяко достатъчно добро начално приближение x_0. С други думи, съществуват константи C и $q \in (0,1)$ такива, че

$$|x_n - \xi| \leq Cq^{2^n} \qquad \text{за всяко } n \ .$$

Това е доста бърза сходимост. За да я илюстрираме по-нагледно, да приемем, че $|\varphi''(t)| \leq 2$ в околност $\mathcal{U}$ на корена ξ. Нека $e_k := |x_k - \xi|$. Тогава при всяко x_0 от $\mathcal{U}$ за следващото приближение x_1, построено по метода на Нютон, ще имаме

$$\begin{aligned} e_1 &= |x_1 - \xi| = |\varphi(x_0) - \varphi(\xi)| \\ &= \left|\varphi'(\xi)(x_0 - \xi) + \frac{\varphi''(\eta)}{2}(x_0 - \xi)^2\right| \quad (\text{като развием } \varphi(x_0) \text{ по Тейлър}) \\ &= \frac{|\varphi''(\eta)|}{2}\, e_0^2 \qquad\qquad (\text{защото } \varphi'(\xi) = 0) \end{aligned}$$

и следователно $e_1 \leq e_0^2$. Аналогично, $e_2 \leq e_1^2$ и т.н. Ако например, x_0 приближава ξ с точност $0,01$, то x_1 ще приближава ξ с точност $e_1 = e_0^2 = 0,0001$, x_2

ще приближава ξ с точност $0,00000001$ и т.н. Вижда се, че броят на точните цифри след десетичната запетая ще се удвоява при всяка итерация.

Високата скорост на сходимост на метода на Нютон е съществено предимство, което го прави най-често използван метод за решаване на уравнения. Той, разбира се, има и недостатъци. Например методът изисква достатъчно добро начално приближение. Това значи, че е необходима предварителна работа за достатъчно добро локализиране на корена ξ преди да се приложи методът на Нютон за намирането му с голяма точност. Друг недостатък е изчисляването на първата производна на f на всяка стъпка. Ако f е дадена експериментално, т.е. стойността на f може да бъде намерена на всяка стъпка, но след отчитане на резултатите от даден експеримент, то изчисляването на производната на f може да предизвика затруднения.

Методът на Нютон е особено удобен за решаване на алгебрични уравнения. В този случай пресмятането на $f(x_n)$ и $f'(x_n)$ (необходими за изчисляването на x_{n+1}) може да се организира ефективно по следния начин. Нека

$$f(x) = a_0x^m + a_1x^{m-1} + \ldots + a_m \; .$$

Стойността на f в дадена точка z ще пресмятаме по известното правило на Хорнер

$$f(z) = (\ldots((a_0z + a_1)z + a_2)z + \ldots + a_{m-1})z + a_m$$

чрез процедурата:

$b_0 := a_0$
за $k = 1, \ldots, m$ извърши:
$b_k = b_{k-1}z + a_k$

и очевидно $f(z) = b_m$. Да забележим сега, че за всяко дадено z, съществува полином $g(x)$ от степен $m-1$ такъв, че

$$f(x) - f(z) = g(x)(x - z) \; . \tag{6}$$

От тази връзка следва, че $f'(z) = g(z)$. Оказва се, че коефициентите на

$$g(x) = b_0x^{m-1} + b_1x^{m-2} + \ldots + b_{m-1}$$

са точно равни на величините $\{b_k\}$ от процедурата на Хорнер за полинома f. Наистина, като сравним коефициентите пред x^{m-k} в двете страни на (6), получаваме

$$a_k = b_k - zb_{k-1}$$

или $b_k = b_{k-1}z + a_k$, което е точно връзката, по която се пресмятат величините $\{b_k\}$ в горната процедура на Хорнер. Следователно пресмятането на $f(z)$ и $f'(z) = g(z)$ може да се обедини в следната процедура:

$b_0 := a_0, \; c_0 := a_0$
за $k = 1, \ldots, m-1$ извърши:

$$b_k = b_{k-1}z + a_k$$
$$c_k = c_{k-1}z + b_k$$
$$b_m = b_{m-1}z + a_m.$$

След тези пресмятания, $b_m = f(z)$ и $c_{m-1} = g(z) = f'(z)$. Следователно при $z = x_n$ можем да пресметнем следващото приближение x_{n+1} по формулата

$$x_{n+1} = x_n - \frac{b_m}{c_{m-1}} .$$

Написването и тестването на компютърна програма за решаване на алгебрични уравнения по този прост алгоритъм е едно приятно занимание.

IV. Комбиниран метод.

Това е една модификация на метода на Нютон, при която пресмятането на приближението x_n се комбинира с пресмятането на друго приближение t_n, по метода на хордите, което се намира от другата страна на корена ξ. За построяване на редиците $\{t_n\}$ и $\{x_n\}$ се прилагат формулите (при предположение, че $f''(x) > 0$ в $[a, b]$)

$$1)\quad x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, \ldots ,$$

$$2)\quad t_{n+1} = t_n - \frac{f(t_n)}{f(x_n) - f(t_n)}(x_n - t_n) .$$

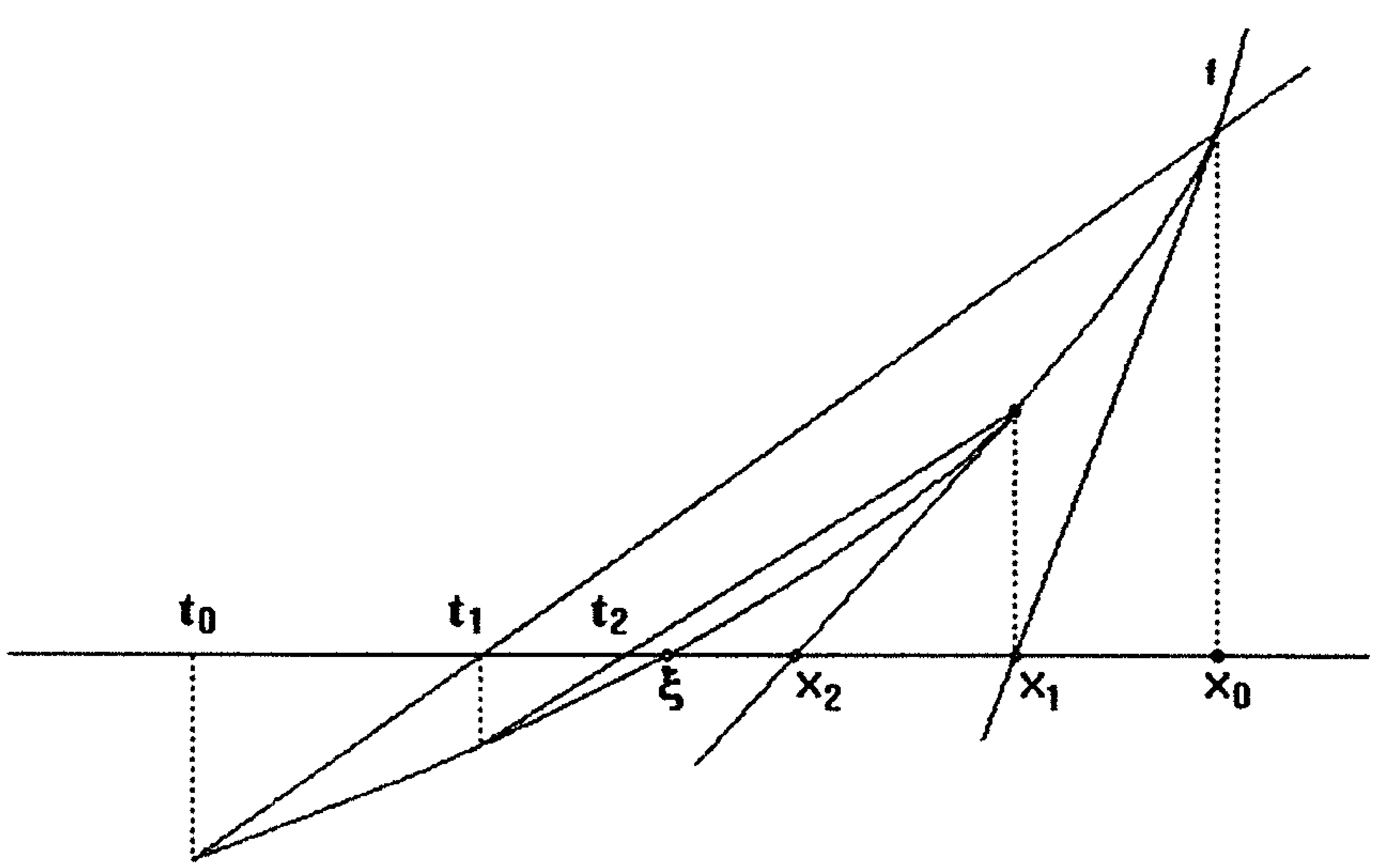

Рис. 19

В 1) е използван *стандартният метод на Нютон*, докато в 2) е приложен *методът на хордите* за интервала $[t_n, x_n]$. По построение, $t_n < \xi < x_n$ (виж Рис. 19). Обикновено за $n-$тото приближение на корена ξ се взима средата на интервала $[t_n, x_n]$. Следователно на всяка стъпка разполагаме с числена оценка на грешката

$$\left|\xi - \frac{t_n + x_n}{2}\right| \leq \frac{|x_n - t_n|}{2} \ .$$

Това е едно от предимствата на този метод. Може да се покаже, че той има ред на сходимост 2.

Г л а в а 4

РЕШАВАНЕ НА СИСТЕМИ ОТ УРАВНЕНИЯ

Ще разгледаме задачата за решаване на системи от n уравнения

$$f_i(x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, n,$$

с n неизвестни $x_1, \ldots, x_n$. Ще започнем с най-простия случай – когато $f_i(x_1, \ldots, x_n)$ са линейни функции по отношение на $x_1, \ldots, x_n$. Тогава горната система се записва във вида

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \ldots + a_{2n}x_n &= b_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + \ldots + a_{nn}x_n &= b_n \end{aligned} \tag{1}$$

където $\{a_{ij}\}$ и $\{b_i\}$ са дадени числа. (Тук ще предполагаме за простота че те са реални числа.) Ако означим с A матрицата $\{a_{ij}\}_{i=1}^{n}{}_{j=1}^{n}$ от коефициенти и с $\bar{b}$ вектора $(b_1, \ldots, b_n)$, то системата (1) се записва в матричен вид по следния начин

$$A\bar{x} = \bar{b}. \tag{2}$$

С $\det A$ ще означаваме детерминантата на A. Известно е от линейната алгебра, че линейната система (2) има единствено решение тогава и само тогава, когато $\det A \neq 0$. Дори решението на (2) може да се запише в явен вид по формулите на Крамер

$$x_k = \frac{\det A_k}{\det A}, \quad k = 1, \ldots, n,$$

където A_k е матрицата, получена от A като заместим k–тия стълб със стълба от свободните членове $\bar{b}$. На пръв поглед няма никакви проблеми в решаването на линейните системи. Това обаче не е така. Веднага се вижда, че формулите на Крамер са крайно неудобни за числено решаване на системи, защото изискват огромен брой операции за пресмятане на детерминантите. Има други, много по-добри числени методи за решаване на линейни системи. Тук ние ще се запознаем с някои от тях.

22. МЕТОД НА ГАУС

Така се нарича известният ни още от средното училище *метод на изключването.* При него първоначалната система

$$A\bar{x} = \bar{b}$$

се свежда до система от вида

$$R\bar{x} = \bar{c}\ ,$$

която има същото решение $\bar{x}$ и където R е горна триъгълна матрица, т.е.

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_{nn} \end{bmatrix} .$$

Преобразуването на таблицата $(A, \bar{b})$ в $(R, \bar{c})$ става стъпка по стъпка, чрез изваждане от дадени редове на матрицата други, умножени с число. Алгоритъмът е следният:

Ако $a_{11} \neq 0$, то от i–тия ред на таблицата

$$(A, \bar{b}) = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} & b_i \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & b_n \end{bmatrix}$$

се изважда първият, умножен с a_{i1}/a_{11}. Това се прави за $i = 2, 3, \dots, n$. Получава се таблица от вида

$$(A^{(1)}, \bar{b}^{(1)}) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \dots & a_{2n}^{(1)} & b_2^{(1)} \\ 0 & a_{32}^{(1)} & a_{33}^{(1)} & \dots & a_{3n}^{(1)} & b_3^{(1)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & a_{n2}^{(1)} & a_{n3}^{(1)} & \dots & a_{nn}^{(1)} & b_n^{(1)} \end{bmatrix} .$$

Това действие е еквивалентно на изключване на x_1 от второ, трето, ..., n–то уравнение на системата $A\bar{x} = \bar{b}$.

Ако $a_{22}^{(1)} \neq 0$, то от i–тия ред на $(A^{(1)}, \bar{b}^{(1)})$ изваждаме втория, умножен с $a_{i2}^{(1)}/a_{22}^{(1)}$, за $i = 3, 4, \ldots, n$. Получаваме нова таблица $(A^{(2)}, \bar{b}^{(2)})$ и т.н. докато стигнем до таблицата

$$(A^{(n-1)}, \bar{b}^{(n-1)}) = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \ldots & a_{2n}^{(1)} & b_2^{(1)} \\ 0 & 0 & a_{33}^{(2)} & \ldots & a_{3n}^{(2)} & b_3^{(2)} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & a_{nn}^{(n-1)} & b_n^{(n-1)} \end{bmatrix} .$$

Да запишем формулите, по които се намират елементите на $(A^{(k)}, \bar{b}^{(k)})$ от тези на $(A^{(k-1)}, \bar{b}^{(k-1)})$. Имаме

$$\begin{aligned} (a_{ik}^{(k)}, \ldots, a_{in}^{(k)}, b_i^{(k)}) &= (a_{ik}^{(k-1)}, \ldots, a_{in}^{(k-1)}, b_i^{(k-1)}) \\ &- \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}(a_{kk}^{(k-1)}, \ldots, a_{kn}^{(k-1)}, b_k^{(k-1)}) . \end{aligned}$$

Следователно

$$(1) \qquad a_{ij}^{(k)} = a_{ij}^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}} a_{kj}^{(k-1)}, \qquad j = k, k+1, \ldots, n .$$

Формулите (1) се прилагат последователно при $k = 1, 2, \ldots, n-1$.

Това е така нареченият *прав ход* на алгоритъма. След изпълнението му получаваме системата

$$(2) \qquad A^{(n-1)}\bar{x} = \bar{b}^{(n-1)} ,$$

която е еквивалентна на първоначалната $A\bar{x} = \bar{b}$ (т.е. има същото решение). При това, матрицата ѝ $A^{(n-1)}$ е горна триъгълна. Да означим за по-кратко $A^{(n-1)}$ с R и $\bar{b}^{(n-1)}$ с $\bar{c}$. Тогава подробният запис на (2) е

$$\begin{aligned} r_{11}x_1 + r_{12}x_2 + \ldots + r_{1n}x_n &= c_1 \\ r_{22}x_2 + \ldots + r_{2n}x_n &= c_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots \quad \ldots &\quad \ldots \\ r_{nn}x_n &= c_n \quad . \end{aligned}$$

и следователно тя може да бъде решена лесно, като неизвестните се определят в обратен ред, $x_n, x_{n-1}, \ldots, x_1$, по формулата

$$x_k = \frac{c_k - \sum\limits_{j=k+1}^{n} r_{kj}x_j}{r_{kk}}, \qquad k = n, n-1, \ldots, 2, 1 .$$

Това е *обратният ход* на алгоритъма.

Методът на Гаус може да бъде приложен, ако елементите a_{11}, $a_{22}^{(1)}$, …, $a_{nn}^{(n-1)}$ са различни от нула. Те се наричат *главни* (или *водещи*) елементи. Но дори и водещите елементи да са различни от нула, ако някой от тях е много малък по абсолютна стойност може да предизвика непренебрежими грешки от закръгляване на числата при работа с компютър, тъй като на всяка стъпка от алгоритъма се извършва деление на водещия елемент. За да се избегнат тези проблеми обикновено се използва една модификация на метода на Гаус, известна като *метод на Гаус с избор на главен елемент.* При нея, в началото на всяка стъпка, за водещ елемент се избира най-големият по абсолютна стойност елемент на съответната матрица. Например при първата стъпка се намира елемент a_{sl}, за който

$$|a_{sl}| = \max_{1\le i,j\le n} |a_{ij}| . \tag{3}$$

Този елемент a_{sl} ще играе ролята на водещ, т.е. от s–тото уравнение ще определяме x_l и ще го изключим от всички останали. На практика, след определянето на s и l, разменяме местата на първия и s-тия ред и на първия и l-тия стълб на таблицата $(A, \bar{b})$ и след това продължаваме както при стандартния метод – намираме $(A^{(1)}, \bar{b}^{(1)})$ по дадената по-горе формула. При втората стъпка търсим най-големия елемент от $\{a_{ij}^{(1)}\}_{i=2}^{n}{}_{j=2}^{n}$ и т.н.

Ясно е, че ако някой от избраните по този начин водещи елементи $a_{kk}^{(k-1)}$ е равен на нула, то $\det A = 0$. Следователно методът на Гаус с избор на главен елемент е приложим за всяка неизродена матрица A.

Понякога се прилага *частичен избор* на главния елемент. При него за водещ елемент се избира най-големият по абсолютна стойност в първия стълб на съответната таблица:

$$|a_{sl}| := \max\left\{ |a_{kk}^{(k-1)}|, |a_{k+1k}^{(k-1)}|, \ldots, |a_{nk}^{(k-1)}| \right\} .$$

Ясно е, че и в този случай, ако $a_{sl}^{(k-1)} = 0$, то $\det A = 0$.

Метод на Гаус-Жордан. Една друга модификация на метода на Гаус е известна като *метод на Гаус-Жордан.* При нея на k–тата стъпка x_k се изключва не само от $k{+}1, \ldots, n$–тото уравнение, но също така от $1, 2, \ldots, (k-1)$–во уравнение. По този начин първоначалната матрица A се трансформира в диагоналната матрица D и съответно, трансформираната система добива вида

$$d_{kk}x_k = c_k, \quad k = 1, \dots, n \ ,$$

която се решава веднага:

$$x_k = \frac{d_{kk}}{c_k}, \quad k = 1, \dots, n \ .$$

Методът на Гаус-Жордан се прилага при работа на машини с по-малка оперативна памет или при решаване на по-големи системи, защото при него е необходимо да се съхраняват по-малко данни.

Решаване на триединната задача по метода на Гаус. Решаването на една линейна система $A\bar{x} = \bar{b}$ е тясно свързано с решаването на други две задачи: намиране на обратната матрица на A и намиране на детерминантата на A. Съвкупността от тези три задачи се нарича *триединна задача*. Сега ще видим как много лесно бихме могли да решим останалите две задачи (намиране на A^{-1} и $\det A$), ако сме решили вече, по метода на Гаус, линейната система $A\bar{x} = \bar{b}$.

При метода на Гаус матрицата A се преобразува последователно в $A^{(1)}, A^{(2)}, \dots, A^{(n-1)} = R$, като от даден ред изваждаме друг ред, умножен с число. Но тази операция не променя детерминантата. Следователно $\det A = \det R$. Тъй като R е триъгълна матрица с елементи $a_{11}, a_{22}^{(1)}, \dots, a_{nn}^{(n-1)}$ по диагонала, то

$$\det A = \det R = a_{11}a_{22}^{(1)} \dots a_{nn}^{(n-1)} \ .$$

И така, при решение на системата $A\bar{x} = \bar{b}$ по Гаус получаваме като допълнителен резултат и детерминантата на A.

Сега ще видим как можем да намерим и елементите на A^{-1}. Нека

$$\begin{aligned} A^{-1} &= \{y_{ij}\}_{i=1}^{n}{}_{j=1}^{n} \\ \bar{y}_m &= (y_{1m}, y_{2m}, \dots, y_{nm})^T \qquad (m\text{-тия стълб на } A^{-1}) \\ \bar{e}_m &= (0, \dots, 1, \dots, 0)^T \qquad (1 \text{ на } m\text{-та позиция}) \ . \end{aligned}$$

Тъй като $A A^{-1} = E$, то ясно е, че

$$A\bar{y}_m = \bar{e}_m \ . \tag{4}$$

Следователно за да намерим елементите на $m-$тия стълб на A^{-1}, трябва да решим линейната система (4). Ако вече имаме решена системата $A\bar{x} = \bar{b}$ по Гаус, то имаме R (трансформацията на A). Следователно за решаването на (4) по Гаус не е необходимо да извършваме правия ход. Остава само да се намерят трансформациите на $\bar{e}_m$ и да се извърши обратния ход, което изисква значително по-малко работа. Това се прави за $m = 1, 2, \dots, n$ и се получават всички елементи на A^{-1}.

23. ТРИЪГЪЛНО РАЗЛАГАНЕ. МЕТОД НА ХОЛЕЦКИ.

Ако матрицата W е триъгълна, то съответната система

$$W\bar{x} = \bar{c}$$

се решава много лесно чрез последователно определяне на неизвестните $x_1, \dots, x_n$ от уравненията на системата. Това ни подсказва да потърсим разлагане на матрицата A на дадена линейна система като произведение на две триъгълни матрици, да речем, във вида

$$A = LR \,, \tag{1}$$

където L е *долна триъгълна* (т.е. $l_{ij} = 0$ за всяко $i < j$), а R е *горна триъгълна* (т.е. $r_{ij} = 0$ за всяко $j < i$). Наистина, ако A може да се представи по този начин, то решаването на системата

$$A\bar{x} = LR\bar{x} = \bar{b} \,,$$

се свежда към решаването на две по-прости системи

$$L\bar{y} = \bar{b} \qquad \text{и} \qquad R\bar{x} = \bar{y}.$$

Разбира се, това може да стане, ако диагоналните елементи на R и L са различни от нули. А това е очевидно изпълнено, ако $\det A \neq 0$, защото $\det A = \det R . \det L = r_{11} \dots r_{nn} l_{11} \dots l_{nn}$.

Всяка линейна система с ненулева детерминанта може да се сведе към еквивалентна на нея система с матрица от вида LR. Нещо повече, това може да се постигне, като се използва методът на Гаус с частичен избор на главния елемент. Ние ще представим тук доказателство само в един частен случай. Преди това ще припомним някои понятия от линейната алгебра.

Следващите преобразования на матрици се наричат *елементарни*:

1. Смяна на местата на два реда или два стълба.

2. Умножение на всички елементи на един ред (стълб) с някакво число $\neq 0$.

3. Прибавяне към елементите на един ред (стълб) съответните елементи на друг ред (стълб), умножени с някакво число.

Две матрици се наричат еквивалентни, ако се получават една от друга с помощта на краен брой елементарни преобразования.

Лесно се вижда, че всяко елементарно преобразование е равносилно на умножението на матрица с някаква неособена матрица. При това, ако преобразованието засяга редовете (стълбовете) на матрицата A, то множителят трябва да стои отляво (отдясно) на A. Освен това множителят се получава от единичната матрица E, като тя се подложи на същото преобразование.

Пример: Преобразованието "смяна на местата на втори и трети ред на матрицата A" е еквивалентно на матричното умножение $\tilde{A} = \tilde{E}A$,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \quad \tilde{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Нека $A\bar{x} = \bar{b}$ е произволна система с ненулева детерминанта. Да предположим отначало че всички водещи елементи $a_{11}, a_{22}^{(1)}, \ldots, a_{nn}^{(n-1)}$, получени по стандартния метод на Гаус (без избор на главен елемент), са различни от нула. Тогава чрез този метод на Гаус матрицата $(A, \bar{b})$ се преобразува в $(R, \bar{c})$, където $R = A^{(n-1)}$ е горна триъгълна с елементи $a_{11}, a_{22}^{(1)}, \ldots, a_{nn}^{(n-1)}$ по главния диагонал, а $\bar{c} = \bar{b}^{(n-1)}$. Както вече видяхме, преобразуването се извършва на $n-1$ стъпки

$$(A,\bar{b}) \to (A^{(1)}, \bar{b}^{(1)}) \to (A^{(2)}, \bar{b}^{(2)}) \to \ldots \to (A^{(n-1)}, \bar{b}^{(n-1)}) = (R, \bar{c}),$$

като матрицата $A^{(k)}$ се получава от $A^{(k-1)}$ с преобразованията: за $i = k+1, \ldots, n$, към i–тия ред на $A^{(k-1)}$ се прибавя k–тия ред на $A^{(k-1)}$, умножен с числото $-s_{ik} := -a_{ik}^{(k-1)}/a_{kk}^{(k-1)}$. Това са елементарни преобразования и те могат да се представят в матричен вид по следния начин:

$$A^{(k)} = S_{n,k} S_{n-1,k} \ldots S_{k+1,k} A^{(k-1)},$$

където $S_{i,k}$ е единичната матрица E, подложена на същото преобразование, т.е.

$$S_{ik} := \begin{array}{c} \\ \\ \end{array} \begin{array}{cccccc} & k & & i & & \\ & \downarrow & & \downarrow & & \end{array}$$

$$S_{ik} := \begin{bmatrix} 1 & \ldots & 0 & \ldots & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \ldots & 1 & \ldots & 0 & \ldots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \ldots & -s_{ik} & \ldots & 1 & \ldots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & \ldots & 0 & \ldots & 1 \end{bmatrix} \begin{array}{l} \\ \\ \leftarrow k \\ \\ \leftarrow i \\ \\ \\ \end{array}.$$

Да положим $S_k := S_{n,k} S_{n-1,k} \ldots S_{k+1,k}$. Тогава $A^{(k)} = S_k A^{(k-1)}$. Да отбележим, че S_k има вида

$$S_k = \begin{bmatrix} 1 & \ldots & 0 & \ldots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \ldots & 1 & \ldots & 0 \\ 0 & \ldots & -s_{k+1,k} & \ldots & 0 \\ 0 & \ldots & -s_{k+2,k} & \ldots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \ldots & -s_{n,k} & \ldots & 1 \end{bmatrix}.$$

Окончателно получаваме

$$R = A^{(n-1)} = S_{n-1}A^{(n-2)} = \ldots = S_{n-1}S_{n-2}\ldots S_1 A \ .$$

Да означим с L^{-1} матрицата $S_{n-1}S_{n-2}\ldots S_1$. Имаме

$$L^{-1} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ -s_{21} & 1 & \ldots & 0 \\ -s_{31} & -s_{32} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ -s_{n1} & -s_{n2} & \ldots & 1 \end{bmatrix} .$$

Лесно се проверява, че обратната матрица на L^{-1} е от вида

$$L = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ l_{21} & 1 & \ldots & 0 \\ l_{31} & l_{32} & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ l_{n1} & l_{n2} & \ldots & 1 \end{bmatrix} ,$$

т.е. тя е долна триъгълна. Тъй като $R = L^{-1}A$ по построение, то оттук,

$$LR = A \ .$$

Тук използвахме съществено предположението, че водещите елементи $a_{11}, a_{22}^{(1)}, \ldots, a_{nn}^{(n-1)}$ са различни от нула. Доказателството в общия случай е по-сложно и ние го пропускаме.

Сега ще разгледаме един клас от матрици, за които методът на Гаус без избор на главен елемент може да бъде приложен, т.е. за които водещите елементи $a_{11}, a_{22}^{(1)}, \ldots, a_{nn}^{(n-1)}$ са винаги различни от нула. От доказаното в първата част следва, че такива матрици могат да бъдат разложени по Гаус на произведение от две триъгълни.

Определение. Ще казваме, че реалната матрица A е *положително определена*, ако

а) тя е симетрична (т.е. $a_{ij} = a_{ji}$ за $\forall i, j$),

б) $(A\bar{x}, \bar{x}) \geq 0$ за всяко $\bar{x}$, като $(A\bar{x}, \bar{x}) = 0$ само при $\bar{x} = \bar{o}$.

Ако A е положително определена, то $\det A \neq 0$. Наистина, да допуснем противното. Тогава съществува ненулев вектор $\bar{x}$ такъв, че $A\bar{x} = \bar{o}$ (защото всяка хомогенна система с нулева детерминанта има ненулево решение). Оттук следва $(A\bar{x}, \bar{x}) = (\bar{o}, \bar{x}) = 0$, противоречие с положителната определеност.

Ако A е положително определена, то $\det A \neq 0$ и следователно A^{-1} съществува. Нещо повече, A^{-1} е също положително определена. Да докажем това. Нека $\bar{y} \neq \bar{o}$. Тогава $\bar{x} := A^{-1}\bar{y} \neq \bar{o}$. Имаме

$$(A^{-1}\bar{y}, \bar{y}) = (\bar{x}, \bar{y}) = (\bar{x}, A\bar{x}) = (A\bar{x}, \bar{x}) > 0,$$

което доказва, че A^{-1} е положително определена.

Да припомним още, че ако една матрица A е положително определена, то всичките ѝ главни минори са положителни. Това следва от известния *критерий на Силвестър*. Ние ще скицираме тук директно доказателство. Ще покажем най-напред, че всяка централна подматрица

$$A_k := \begin{bmatrix} a_{i_1 i_1} & \dots & a_{i_1 i_k} \\ \vdots & & \vdots \\ a_{i_k i_1} & \dots & a_{i_k i_k} \end{bmatrix}$$

е положително определена. Нека $\bar{y} = (y_{i_1}, \dots, y_{i_k})$ е произволен ненулев вектор от $\mathbb{R}^k$. Нека $\bar{x} = (x_1, \dots, x_n)$ е съответният му, допълнен с нули вектор от $\mathbb{R}^n$, т.е. $x_{i_j} = y_{i_j}$ за $j = 1, \dots, k$ и $x_i = 0$ при $i \neq i_1, \dots, i_k$. Очевидно

$$(A_k \bar{y}, \bar{y}) = (A\bar{x}, \bar{x}) .$$

Следователно $(A_k \bar{y}, \bar{y}) > 0$, ако $\bar{y} \neq \bar{o}$ и A е положително определена. Това показва, че A_k е също положително определена.

Остана да видим, че за всяка положително определена матрица A имаме $\det A > 0$. Оттук следва, че $\det A_k > 0$ за всеки главен минор $\det A_k$. Ще използваме индукция по n. За матрици 1×1 това е очевидно. Допускаме, че детерминантата на всяка положително определена матрица от ред $n-1$ е положителна. Нека A е произволна положително определена матрица от ред n. Както видяхме вече, A^{-1} е също положително определена. Нека α_{ij} са елементите на A^{-1}. Знаем от линейната алгебра, че

$$\alpha_{ij} = \frac{A_{ji}}{\det A} ,$$

където A_{ks} е адюнгираното количество на елемента a_{ks}. В частност

$$\alpha_{11} = \frac{A_{11}}{\det A} . \tag{2}$$

Но A_{11} е главният минор

$$\det \begin{bmatrix} a_{22} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{n2} & \dots & a_{nn} \end{bmatrix} ,$$

който е положителен, съгласно индукционното предположение. Освен това, пак по индукционното предположение, $\alpha_{11} > 0$. Тогава от (2) следва $\det A > 0$ и твърдението е доказано.

Теорема 1. *Ако A е положително определена матрица, то тя може да се разложи по единствен начин във вида*

$$A = LL^T$$

където L е долна триъгълна матрица.

Доказателството е конструктивно. То показва начина, по който това разложение може да бъде получено. Методът е предложен от Холецки и е свързан с неговото име. В някои книги той се нарича *метод на квадратния корен.* Да означим с $\{\alpha_{ij}\}$ елементите на L. Ще ги определяме от равенството

$$\begin{pmatrix} \alpha_{11} & 0 & \dots & 0 \\ \alpha_{21} & \alpha_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{pmatrix} \cdot \begin{pmatrix} \alpha_{11} & \alpha_{21} & \dots & \alpha_{n1} \\ 0 & \alpha_{22} & \dots & \alpha_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} .$$

Имаме $\alpha_{11}^2 = a_{11}$. Тъй като $a_{11} > 0$, то

$$\alpha_{11} = \sqrt{a_{11}}$$

и α_{11} е реално положително число. По-нататък,

$$a_{1j} = \alpha_{11}\alpha_{j1} \ , \qquad j = 2, \dots, n,$$

и оттук определяме всички останали елементи $\{\alpha_{j1}\}$ от първия стълб на L:

$$\alpha_{j1} = \frac{a_{1j}}{\alpha_{11}} \ , \qquad j = 2, \dots, n \ .$$

Аналогично определяме последователно елементите от втория, третия, ..., n−тия стълб на L. Да запишем общите формули. Имаме

$$a_{kk} = \alpha_{k1}^2 + \alpha_{k2}^2 + \dots + \alpha_{kk}^2 \ .$$

Оттук

(3) $$\alpha_{kk} = \left(a_{kk} - \alpha_{k1}^2 - \dots - \alpha_{kk-1}^2\right)^{1/2} \ .$$

От връзката

$$a_{kj} = \alpha_{k1}\alpha_{j1} + \alpha_{k2}\alpha_{j2} + \dots + \alpha_{kk}\alpha_{jk} \ , \qquad j = k+1, \dots, n \ ,$$

определяме

(4) $$\alpha_{jk} = \frac{a_{kj} - \sum\limits_{i=1}^{k-1} \alpha_{ki}\alpha_{ji}}{\alpha_{kk}} \ , \qquad j = k+1, \dots, n \ .$$

Разбира се този метод може да бъде приложен за всяка матрица A, ако величините, които коренуваме в (3), са различни от нула. Ще покажем, че при направените предположения за A те са дори положителни.

Да означим с A_k централната подматрица

$$A_k = \begin{bmatrix} a_{11} & \dots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \dots & a_{kk} \end{bmatrix} .$$

Тъй като

$$A_k = \begin{pmatrix} \alpha_{11} & \dots & 0 \\ \vdots & & \vdots \\ \alpha_{k1} & \dots & \alpha_{kk} \end{pmatrix} \cdot \begin{pmatrix} \alpha_{11} & \dots & \alpha_{k1} \\ \vdots & & \vdots \\ 0 & \dots & \alpha_{kk} \end{pmatrix} ,$$

то

$$\det A_k = \alpha_{11}^2 \dots \alpha_{kk}^2 . \tag{5}$$

Тъй като $\det A_k > 0$ (всички главни минори на A са положителни), то $\alpha_{11}^2 \dots \alpha_{kk}^2 > 0$ за всяко $k = 1, 2, \dots, n$. Но $\alpha_{11}^2 = a_{11} > 0$. Тогава

$$\alpha_{11}^2 \, \alpha_{22}^2 > 0 \quad \Rightarrow \quad \alpha_{22}^2 > 0 ,$$

$$\alpha_{11}^2 \, \alpha_{22}^2 \, \alpha_{33}^2 > 0 \quad \Rightarrow \quad \alpha_{33}^2 > 0 ,$$

$$\dots\dots\dots\dots\dots\dots \quad \dots \quad \dots\dots\dots$$

$$\alpha_{11}^2 \, \alpha_{22}^2 \dots \alpha_{nn}^2 > 0 \quad \Rightarrow \quad \alpha_{nn}^2 > 0 .$$

Да отбележим, че тук α_{kk}^2 е точно величината под корена в (3). Следователно α_{kk} е реално и положително число.

24. НОРМИ НА МАТРИЦИ. СХОДИМОСТ НА МАТРИЧЕН РЕД

Да разгледаме множеството $\mathcal{A}_n$ от всички реални матрици

$$A = \{a_{ij}\}_{i=1,\ j=1}^{n\quad n}$$

с размерност $n \times n$. Да въведем в $\mathcal{A}_n$ операциите *събиране* на матрици и *умножение* на матрица с число c по стандартния начин:

Ако $A = \{a_{ij}\}$ и $B = \{b_{ij}\}$, то

$$\begin{aligned} A + B &:= \{a_{ij} + b_{ij}\}\ , \\ cA &:= \{ca_{ij}\}\ . \end{aligned}$$

С O ще означаваме *нулевата* матрица (чиито елементи са 0), а с E – *единичната* матрица. След тези определения $\mathcal{A}_n$ става едно линейно пространство. Да отбележим, че в $\mathcal{A}_n$ може да се определи и операцията *умножение* на матрица с матрица:

$$AB := \{(\bar{a}_i, \bar{b}_j)\}, \quad \bar{a}_i = (a_{i1}, \ldots, a_{in}),\ \bar{b}_j = (b_{1j}, \ldots, b_{nj})^T\ .$$

Произведението AB е също елемент на $\mathcal{A}_n$. Сега ще въведем понятието "норма" в $\mathcal{A}_n$.

Матриците от $\mathcal{A}_n$ са таблици от n^2 числа. Те могат да се разглеждат като n^2–мерни вектори. (Операциите събиране и умножение с число са определени както при векторите.) Затова всички условия от известното ни определение за норма на вектор трябва да участват и в определението за норма на матрица. Освен това, към тези условия е добавено още едно, което засяга операцията умножение на матрица с матрица.

Определение. Казваме, че в $\mathcal{A}_n$ е определена *норма* $\|\cdot\|$, ако на всяка матрица $A \in \mathcal{A}_n$ е съпоставено число $\|A\|$ (т.е. нормата на A), което удовлетворява следните условия:

$$\begin{aligned} &1)\quad \|A\| \geq 0; \quad \|A\| = 0 \iff A = O\ . \\ &2)\quad \|\lambda A\| = |\lambda|\ \|A\| \text{ за всяко число } \lambda\ . \\ &3)\quad \|A + B\| \leq \|A\| + \|B\|\ . \\ &4)\quad \|AB\| \leq \|A\|.\|B\|\ . \end{aligned}$$

Дали има съответствия $A \to \|A\|$, които удовлетворяват горните условия? Оказва се, че всяка векторна норма определя съответна матрична норма. В сила е следното твърдение.

Теорема 1. *Нека $\|\cdot\|$ е норма на вектор в $\mathbb{R}^n$. Тогава съответствието $A \to \|A\|$, където*

$$\|A\| := \sup_{\|\bar{x}\|=1} \|A\bar{x}\| \tag{1}$$

определя норма в $\mathcal{A}_n$.

Доказателство. Ще проверим, че съответствието $A \to \|A\|$, определено в (1), удовлетворява всички условия от определението на норма.

1) Очевидно $\|A\| \geq 0$ като супремум на неотрицателни числа $\|A\bar{x}\|$. Освен това, $\|O\| = 0$. Остава само да се види, че от равенството $\|A\| = 0$ следва $A = O$. Наистина, да допуснем, че има матрица $A \in \mathcal{A}_n$, за която $\|A\| = 0$, но $A \neq O$. Тогава от равенството $AE = A$ следва, че поне един от векторите $A\bar{e}_1, A\bar{e}_2, \ldots, A\bar{e}_n$ (които са стълбове на A), е различен от $\bar{0}$. Тук с $\bar{e}_k$ сме означили k-тия единичен вектор, т.е.

$$\bar{e}_k := (0, \ldots, 1, \ldots, 0), \text{ където 1 е на } k-\text{та позиция} .$$

Нека например $A\bar{e}_k \neq \bar{0}$. Тогава и $A\bar{\xi} \neq \bar{0}$ при $\bar{\xi} := \bar{e}_k/\|\bar{e}_k\|$. Но $\|\bar{\xi}\| = 1$ и следователно

$$\|A\| \geq \|A\bar{\xi}\| > 0 .$$

което противоречи на допускането, че $\|A\| = 0$.

2) Използвайки съответното свойство на векторна норма, получаваме

$$\|\lambda A\| = \sup_{\|\bar{x}\|=1} \|\lambda A\bar{x}\| = \sup_{\|\bar{x}\|=1} \{|\lambda|\, \|A\bar{x}\|\} = |\lambda|\, \|A\| .$$

3) Тъй като векторната норма удовлетворява неравенството на триъгълника, то

$$\begin{aligned} \|A+B\| &= \sup_{\|\bar{x}\|=1} \|(A+B)\bar{x}\| = \sup_{\|\bar{x}\|=1} \|A\bar{x} + B\bar{x}\| \\ &\leq \sup_{\|\bar{x}\|=1} \{\|A\bar{x}\| + \|B\bar{x}\|\} \leq \sup_{\|\bar{x}\|=1} \|A\bar{x}\| + \sup_{\|\bar{x}\|=1} \|B\bar{x}\| \\ &= \|A\| + \|B\| . \end{aligned}$$

4) При доказателството на свойство 4) ще използваме следното важно неравенство

$$\|A\bar{x}\| \leq \|A\|.\|\bar{x}\| , \tag{2}$$

отнасящо се за величината (1). Когато $\|\bar{x}\| = 1$, то

$$\|A\| = \sup_{\|\bar{y}\|=1} \|A\bar{y}\| \geq \|A\bar{x}\|$$

и следователно (2) е в сила. Неравенството е очевидно вярно и при $\bar{x} = \bar{0}$. Нека сега $\|\bar{x}\| \neq 0$. Тогава векторът $\bar{x}/\|\bar{x}\|$ има норма 1 и съгласно току-що доказаното

$$\left\| A \frac{\bar{x}}{\|\bar{x}\|} \right\| \leq \|A\| ,$$

откъдето следва $\|A\bar{x}\| \leq \|A\|.\|\bar{x}\|$, т.е. (2).

Сега вече можем да докажем свойство 4). Имаме

$$\begin{aligned}\|AB\| &= \sup_{\|\bar{x}\|=1} \|AB\bar{x}\| = \sup_{\|\bar{x}\|=1} \|A.(B\bar{x})\| \\ &\leq \sup_{\|\bar{x}\|=1} \{\|A\|.\|B\bar{x}\|\} \qquad \text{(съгласно (2))} \\ &= \|A\|.\sup_{\|\bar{x}\|=1} \|B\bar{x}\| = \|A\|.\|B\| \;.\end{aligned}$$

Доказателството на теоремата е завършено.

Неравенството (2) изразява така нареченото условие за *съгласуваност* между векторната и матричната норма.

Определение. Ще казваме, че векторната норма $\|\bar{x}\|$ и матричната норма $\nu(A)$ са *съгласувани*, ако

$$\|A\bar{x}\| \leq \nu(A)\|\bar{x}\| \qquad \text{за всяко } \bar{x} \in \mathbb{R}^n \text{ и } A \in \mathcal{A}_n \;.$$

Може да има много матрични норми, които са съгласувани с дадена векторна норма. Най-малката от всички тези норми се нарича *подчинена* на векторната, т.е. $\nu(\cdot)$ е подчинена на $\|\cdot\|$, ако

$$\nu(A) \leq \mu(A)$$

за всяко $A \in \mathcal{A}_n$ и всяка друга, съгласувана с $\|\cdot\|$ матрична норма $\mu(A)$.

Вярно е следното

Твърдение: *Нека $\|\cdot\|$ е произволна дадена векторна норма. Тогава матричната норма, определена в $\mathcal{A}_n$ с равенството*

$$\|A\| := \sup_{\|\bar{x}\|=1} \|A\bar{x}\|$$

е подчинена на векторната.

Доказателството е елементарно. Нека $\nu(\cdot)$ е произволна друга норма, съгласувана с векторната $\|\cdot\|$. Тогава

$$\begin{aligned}\|A\| &= \sup_{\|\bar{x}\|=1} \|A\bar{x}\| = \|A\bar{x}_0\| \qquad \text{(за някакво } \bar{x}_0 \text{ с норма 1)} \\ &\leq \nu(A)\|\bar{x}_0\| \qquad \text{(поради съгласуваността на } \nu(\cdot)) \\ &= \nu(A) \qquad \text{(защото } \|\bar{x}_0\| = 1) \;.\end{aligned}$$

Вече имахме повод да споменем, че най-често ползваните векторни норми са:

$$\begin{aligned}
\|\bar{x}\|_\infty &:= \max_{1\le i\le n} |x_i| && \text{(равномерна или максимум норма)} \ , \\
\|\bar{x}\|_1 &:= \sum_{i=1}^{n} |x_i| && \text{(първа норма)} \ , \\
\|\bar{x}\|_2 &:= \left\{\sum_{i=1}^{n} x_i^2\right\}^{\frac{1}{2}} && \text{(евклидова норма)} \ .
\end{aligned}$$

Като използваме доказаното по-горе твърдение, можем да намерим подчинените им матрични норми, които ще означаваме със същите индекси (∞, 1 и 2). Ето едно упътване за тяхното намиране:

$$\begin{aligned}
\|A\|_\infty &:= \sup_{\|\bar{x}\|_\infty=1} \|A\bar{x}\|_\infty = \sup_{\|\bar{x}\|_\infty=1} \max_{1\le i\le n} \left|\sum_{j=1}^{n} a_{ij}x_j\right| \\
&= \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}| \ , \\
\|A\|_1 &:= \sup_{\|\bar{x}\|_1=1} \|A\bar{x}\|_1 = \sup_{\|\bar{x}\|_1=1} \sum_{i=1}^{n} \left|\sum_{j=1}^{n} a_{ij}x_j\right| \\
&= \sup_{\|\bar{x}\|_1=1} \sum_{j=1}^{n} \left\{\sum_{i=1}^{n} |a_{ij}|\right\} |x_j| = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}| \ , \\
\|A\|_2 &:= \sup_{\|\bar{x}\|_2=1} \|A\bar{x}\|_2 = \sup_{\|\bar{x}\|_2=1} \sqrt{(A\bar{x}, A\bar{x})} \\
&= \sup_{\|\bar{x}\|_2=1} \sqrt{(A^TA\bar{x}, \bar{x})} \ .
\end{aligned}$$

Тъй като матрицата A^TA е симетрична, всичките ѝ собствени стойности $\lambda_1, \dots, \lambda_n$ са положителни. Нека λ е най-голямата от тях и нека $\bar{e}$ е съответстващият ѝ собствен вектор, нормиран с условието $\|\bar{e}\|_2 = 1$. Тогава $(A^TA\bar{e}, \bar{e}) = (\lambda\bar{e}, \bar{e}) = \lambda\ \|\bar{e}\|_2^2 = \lambda$ и от полученото по-горе следва

$$\|A\|_2 \ge \sqrt{(A^TA\bar{e}, \bar{e})} = \lambda^{\frac{1}{2}} \ .$$

От друга страна, ако $\bar{e}_1, \dots, \bar{e}_n$ е ортонормираната система от собствени вектори на A^TA, съответстващи на $\lambda_1, \dots, \lambda_n$, то всеки вектор $\bar{x}$ от $\mathbb{R}^n$ се представя по единствен начин във вида

$$\bar{x} = c_1\bar{e}_1 + \ldots + c_n\bar{e}_n \ .$$

При това $\|\bar{x}\|^2 = c_1^2 + \ldots + c_n^2$. Следователно, ако $\|\bar{x}\|_2 = 1$, то

$$\begin{aligned}(A^TA\bar{x},\bar{x}) &= \left(A^TA\sum_{i=1}^n c_i\bar{e}_i, \sum_{i=1}^n c_i\bar{e}_i\right)\\ &= \left(\sum_{i=1}^n c_i\lambda_i\bar{e}_i, \sum_{i=1}^n c_i\bar{e}_i\right) = \sum_{i=1}^n \lambda_i c_i^2 \le \lambda\ \|\bar{x}\|_2^2 = \lambda\ .\end{aligned}$$

Това неравенство и полученото преди показват, че $\|A\|_2 = \lambda^{\frac{1}{2}}$.

След като сме въвели някаква норма в $\mathcal{A}_n$, тя поражда разстояние в $\mathcal{A}_n$ и следователно можем да говорим за сходимост на редици от матрици относно това разстояние. Ще казваме, че редицата $A_1, A_2, \ldots$ от матрици е сходяща към A, ако $\|A_k - A\| \to 0$ при $k \to \infty$, за която и да е норма $\|\cdot\|$ в $\mathcal{A}_n$. Тъй като матриците могат да бъдат разглеждани като вектори с размерност n^2 и матричната норма удовлетворява всичките 3 условия, предявени към векторната норма, то от еквивалентността на векторните норми в $\mathbb{R}^{n^2}$ следва, че всеки две матрични норми в $\mathcal{A}_n$ са еквивалентни. Следователно, ако една редица от матрици е сходяща относно една норма в $\mathcal{A}_n$, то тя е сходяща относно всяка друга норма и специално, относно максимум нормата. Оттук следва, че сходимост на редици от матрици може да се определи по следния, еквивалентен на дадения начин:

Определение. Казваме, че редицата от матрици $A_k = \left\{a_{ij}^{(k)}\right\}$, $k = 0, 1, \ldots$, е *сходяща* към $A = \{a_{ij}\}$, ако

$$a_{ij}^{(k)} \to a_{ij} \qquad \text{при} \quad k \to \infty\ .$$

С други думи, $A_k \to A$, ако имаме сходимост по елементи (координати).

След като знаем какво е сходяща матрична редица, по аналогия с числови редове, можем да определим и сходимост на матрични редове.

Определение. Казваме, че матричният ред

(3) $$a_0E + a_1A + a_2A^2 + a_3A^3 + \ldots$$

е *сходящ*, ако е сходяща редицата от парциалните му суми

$$S_m(A) := a_0E + a_1A + \ldots + a_mA^m, \qquad m = 0, 1, \ldots$$

Границата $S(A)$ на $S_m(A)$ при $m \to \infty$ се нарича *сума* на матричния ред (3). Тук $a_0, a_1, \ldots$ са реални числа, а $A^k := A.A\ldots A$ (k пъти).

Сходимостта на матричния ред (3) е тясно свързана със сходимостта на числовия ред

(4) $$a_0 + a_1t + a_2t^2 + \ldots\ .$$

С $S_m(t)$, съответно $S(t)$, ще означаваме парциалните суми и сумата на числовия ред, ако той е сходящ.

Теорема 2. *Нека R е радиусът на сходимост на числовия ред (4). Ако $|\lambda| < R$ за всяка собствена стойност λ на A, то матричният ред (3) е сходящ за A. Ако числовият ред (4) е разходящ за някоя собствена стойност λ на A, то матричният ред (3) е разходящ за A.*

Доказателство. Ако $C \in \mathcal{A}_n$ и $\det C \neq 0$, то преобразованието

$$A \to C^{-1}AC$$

се нарича *преобразование на подобие*. Известно е, че преобразованието на подобие запазва собствените стойности на A. Това следва от факта, че характеристичните полиноми на A и на $C^{-1}AC$ са едни и същи:

$$\begin{aligned}\det(C^{-1}AC - \lambda E) &= \det(C^{-1}AC - \lambda C^{-1}EC)\\ &= \det C^{-1}(A - \lambda E)C\\ &= \det C^{-1}.\det(A - \lambda E).\det C = \det(A - \lambda E)\ .\end{aligned}$$

Освен това се вижда, че ако матричният ред (3) е сходящ за A, той е сходящ и за $C^{-1}AC$ и обратно. Наистина, тъй като $S_m(C^{-1}AC) = C^{-1}S_m(A)C$, то

$$S_m(A) \to S(A) \qquad \Longleftrightarrow \qquad S_m(C^{-1}AC) \to C^{-1}S(A)C\ .$$

Следователно теоремата ще бъде доказана напълно, ако докажем твърдението ѝ за някое специално преобразование на подобие на A. Затова ние ще смятаме оттук нататък, че матрицата A е в нормална жорданова форма. (Знаем от линейната алгебра, че всяка матрица $A \in \mathcal{A}_n$ може да бъде приведена чрез преобразование на подобие в нормална жорданова форма.) И така, нека A има собствени стойности $\lambda_1, \dots, \lambda_n$, като някои от тях могат и да съвпадат. Тогава A се представя в нормална жорданова форма от вида

$$A = \begin{pmatrix} \boxed{B_1} & & & \\ & \boxed{B_2} & & O \\ & & \ddots & \\ O & & & \\ & & & \boxed{B_k} \end{pmatrix},$$

където B_i са жордановите клетки с някакви размерности $\nu_i \times \nu_i$, $\nu_1 + \dots + \nu_k = n$, съответстващи на собствените стойности λ_i, т.е.

$$B = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}.$$

Тук, както и по-долу, сме изпуснали индекса i за улеснение. Вижда се, че ако $p(t)$ е алгебричен полином, то

$$p(B_i) = \begin{pmatrix} p(\lambda) & \frac{p'(\lambda)}{1!} & \frac{p''(\lambda)}{2!} & \dots & \frac{p^{(\nu-1)}(\lambda)}{(\nu-1)!} \\ 0 & p(\lambda) & \frac{p'(\lambda)}{1!} & \dots & \frac{p^{(\nu-2)}(\lambda)}{(\nu-2)!} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & p(\lambda) \end{pmatrix}.$$

Това се доказва по индукция. Нека твърдението е вярно за всяко $p_N \in \pi_N$. Произволен полином p_{N+1} от π_{N+1} се представя във вида $p_{N+1}(t) = tp_N(t) + c$, където c е константа. Тогава, очевидно

$$p_{N+1}(\lambda) = \lambda p_N(\lambda) + c$$

и

$$p_{N+1}^{(j)}(\lambda) = \lambda p_N^{(j)}(\lambda) + j\, p_N^{(j-1)}(\lambda), \quad j = 1, 2,$$

Следователно

$$\frac{1}{j!} p_{N+1}^{(j)}(\lambda) = \frac{p_N^{(j)}(\lambda)}{j!} \lambda + \frac{p_N^{(j-1)}(\lambda)}{(j-1)!} 1 \quad \text{при} \quad j \neq 0.$$

Това са точно формулите за пресмятане на елементите на $p_{N+1}(B)$ въз основа на връзката

$$p_{N+1}(B) = B\, p_N(B) + c\, E$$

и индукционното предположение.

Специално при $p(t) = S_m(t)$ ще имаме:

$$S_m(B) = \begin{pmatrix} S_m(\lambda) & \frac{S_m'(\lambda)}{1!} & \dots & \frac{S_m^{(\nu-1)}(\lambda)}{(\nu-1)!} \\ 0 & S_m(\lambda) & \dots & \frac{S_m^{(\nu-2)}(\lambda)}{(\nu-2)!} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S_m(\lambda) \end{pmatrix}.$$

И така, $S_m(A)$ се състои от клетките $S_m(B_i)$, разположени по диагонал, а останалите елементи на $S_m(A)$ са нули. Оттук е ясно, че ако $|\lambda| < R$, то числовите редици $S_m(\lambda), S'_m(\lambda), \ldots, S_m^{(j)}(\lambda)$ са сходящи. Следователно, ако $|\lambda_i| < R$ за всяка собствена стойност λ_i на A, то елементите на $S_m(B_i)$, а оттук и на $S_m(A)$, са сходящи. Това означава, че матричният ред (3) е сходящ за A. Обратно, ако числовият ред (4) е разходящ за поне една собствена стойност λ, то редицата $S_m(\lambda)$ е разходяща и оттук, матричният ред (3) е разходящ. Теоремата е доказана.

Особено важно за нас е следното непосредствено следствие от тази обща теорема.

Следствие 3. *Матричната геометрична прогресия*

$$E + A + A^2 + \ldots \tag{5}$$

е сходяща за A тогава и само тогава, когато всичките собствени стойности на A са по модул по-малки от 1.

Твърдението следва от известния факт, че радиусът на сходимост на числовата геометрична прогресия

$$1 + t + t^2 + \ldots \tag{6}$$

е равен на 1. (По-точно, редът (6) е сходящ само при $|t| < 1$.)

Ще приведеме още едно следствие от доказаната Теорема 2.

Лема 4. *Матричната геометрична прогресия (5) е сходяща тогава и само тогава, когато $A^m \to O$ при $m \to \infty$.*

Доказателство. И тук можем да си мислим, че A е в нормална жорданова форма. Тогава A^m се състои от блокове във вида

$$B^m = \begin{pmatrix} \lambda^m & \binom{m}{1}\lambda^{m-1} & \ldots & \ldots \\ 0 & \lambda^m & \ldots & \ldots \\ \vdots & \vdots & \ldots & \ldots \\ 0 & 0 & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda^m \end{pmatrix}.$$

Нека матричната геометрична прогресия е сходяща за A. Тогава по Следствие 3, $|\lambda_i| < 1$ за всяка собствена стойност λ_i на A. Но ако $|\lambda| < 1$, то $\lambda^m \to 0$. Тогава $B^m \to O$ и следователно $A^m \to O$. Обратно, ако $A^m \not\to O$, то поне един елемент $\binom{m}{j}\lambda^{m-j}$ на B^m не клони към 0. Това може да бъде вярно, само ако $|\lambda| \geq 1$. Оттук, по Следствие 3, матричната геометрична прогресия е разходяща.

Собствените стойности на една матрица от A_n са корени на алгебрични полиноми от степен n. Тяхното намиране е свързано с решаване на алгебрично уравнение, което както знаем вече, не е лесна работа. Затова е удобно за се знаят прости методи за намиране на горна граница за $|\lambda_i|$. Това би ни осигурило прости достатъчни условия за сходимост на матричен ред. По-долу ще приведем една оценка за $|\lambda_i|$ чрез нормата на A.

Лема 5. *Всяка норма на A е по-голяма или равна на абсолютната стойност на всяка собствена стойност на A. С други думи,*

$$|\lambda_i| \leq \|A\| \qquad \textit{за всяка норма} \; .$$

Доказателство. Нека $\|\cdot\|$ е произволна норма в $\mathcal{A}_n$. Ако $\|\cdot\|$ е съгласувана с някаква векторна норма (да я означим пак с $\|\cdot\|$), то твърдението следва веднага. Наистина, нека $\bar{x}$ е собствен вектор, съответстващ на λ_i. Тогава $A\bar{x} = \lambda_i \bar{x}$ и следователно

$$|\lambda_i| \; \|\bar{x}\| = \|\lambda_i \bar{x}\| = \|A\bar{x}\| \leq \|A\|.\|\bar{x}\| \qquad \Rightarrow \qquad |\lambda_i| \leq \|A\| \; .$$

Сега ще дадем друго доказателство на неравенството от лемата, което обхваща общия случай (на произволна матрична норма $\|\cdot\|$). За целта да разгледаме матрицата $B := \frac{A}{\|A\|+\varepsilon}$, където ε е произволно положително число. Очевидно $\|B\| < 1$. Тогава по четвърто свойство на матрична норма,

$$\|B^m\| \leq \|B\|^m \to 0 \quad \text{при} \;\; m \to \infty \; .$$

Следователно $B^m \to O$ и по Лема 4, матричната геометрична прогресия

$$E + B + B^2 + \dots$$

е сходяща. Тогава по Теорема 2 (или Следствие 3), $|\mu| < 1$ за всяка собствена стойност μ на B. Но собствените стойности $\{\mu_i\}$ на B са свързани с $\{\lambda_i\}$ чрез равенството

$$\mu_i = \frac{\lambda_i}{\|A\| + \varepsilon} \, , \qquad i = 1, \dots, m \; .$$

Следователно

$$\left| \frac{\lambda_i}{\|A\| + \varepsilon} \right| < 1 \quad \Rightarrow \quad |\lambda_i| < \|A\| + \varepsilon \quad \Rightarrow \quad |\lambda_i| \leq \|A\| \; .$$

Твърдението е доказано.

Следствие 6. *Ако $\|A\| < 1$ за някоя матрична норма, то матричната геометрична прогресия $E + A + A^2 + \dots$ е сходяща.*

Действително: $\|A\| < 1$ влече $|\lambda| < 1$ за всяка собствена стойност на A. Прилагаме Следствие 3.

Следствие 7. *Ако $\|A\| < 1$, то $A^m \to O$ при $m \to \infty$.*

Лема 8. *Ако всички собствени стойности* $\{\lambda_i\}$ *на* A *са по модул по-малки от 1, то* $E - A$ *е обратима и*

$$(E - A)^{-1} = E + A + A^2 + \dots \; .$$

Доказателство. От равенството

$$(E - A)(E + A + \dots + A^m) = E - A^{m+1},$$

след граничен преход, като използваме Теорема 2, получаваме

$$(E - A)\; S(A) = E \, ,$$

където

$$S(A) = E + A + A^2 + \dots \; .$$

Следователно $E - A$ е обратима и $(E - A)^{-1} = S(A)$.

25. ИТЕРАЦИОННИ МЕТОДИ ЗА РЕШАВАНЕ НА ЛИНЕЙНИ СИСТЕМИ

Приближените методи за решаване на линейни системи от уравнения са главно *итерационни*. При тях се избира подходящо начално приближение $\bar{x}_0 = (x_1^{(0)}, \ldots, x_n^{(0)})$ на решението $\bar{x} = (x_1, \ldots, x_n)$ и след това по формула от вида

$$\bar{x}_{k+1} = B_k \bar{x}_k + \bar{d}_k, \quad k = 0, 1, 2, 3, \ldots$$

се строи редица $\{\bar{x}_k\}$ от точки в $\mathbb{R}^n$, която клони към решението $\bar{x}$.

Тук ние ще разгледаме някои основни итерационни методи за решаване на линейни системи.

Метод на простата итерация. Нека е дадена системата $A\bar{x} = \bar{b}$. Преобразуваме я с помощта на неособената матрица C в еквивалента на нея система

$$\bar{x} = \bar{x} - C\{A\bar{x} - \bar{b}\} .$$

Построяваме итерационния процес

$$\bar{x}_{k+1} = \bar{x}_k - C\{A\bar{x}_k - \bar{b}\}, \quad k = 0, 1, \ldots$$

при някакво начално приближение $\bar{x}_0$. Горната формула може да се запише още така

$$\bar{x}_{k+1} = (E - CA)\bar{x}_k + C\bar{b} =: B\bar{x}_k + \bar{d} .$$

Теорема 1. *Итерационният процес*

$$\bar{x}_{k+1} = B\bar{x}_k + \bar{d}$$

е сходящ при произволен избор на началното приближение $\bar{x}_0$ тогава и само тогава, когато всички собствени стойности на матрицата B са по модул по-малки от 1.

Доказателство. Имаме

$$\begin{aligned} \bar{x}_{k+1} = B\bar{x}_k + \bar{d} &= BB\bar{x}_{k-1} + B\bar{d} + \bar{d} = \cdots \\ &= B^{k+1}\bar{x}_0 + (B^k + B^{k-1} + \ldots + E)\bar{d} . \end{aligned}$$

При направените предположения за B, редът в скобите е сходящ, а $B^{k+1} \to O$ (съгласно Лема 4 от предишната лекция). Следователно редицата $\bar{x}_k$ има граница при $k \to \infty$ и тази граница е $(E - B)^{-1}\bar{d}$. Вижда се, че тази граница е решение на уравнението $\bar{x} = B\bar{x} + \bar{d}$, т.е. решение на нашата система. Да забележим, че ако редът $E + B + \cdots$ не е сходящ, то и редицата $\{\bar{x}_k\}$ може да бъде разходяща, например при $\bar{x}_0 = \bar{o}$. Теоремата е доказана.

Следствие 2. *Ако* $\|B\| < 1$ *за някоя норма* $\|.\|$, *то итерационният процес е сходящ при произволно начално приближение* $\bar{x}_0$.

Твърдението следва веднага от теоремата, като се вземе предвид, че всяка норма на матрицата е по-голяма по абсолютна стойност от всяка нейна собствена стойност. В този случай дори може да се изведе лесно и една оценка на грешката. Наистина, имаме

$$\|\bar{x}_k - \bar{x}\| = \|B\bar{x}_{k-1} - B\bar{x}\| \leq \|B\|\|\bar{x}_{k-1} - \bar{x}\| ,$$

откъдето следва

$$\|\bar{x}_k - \bar{x}\| \leq \|B\|^k\|\bar{x}_0 - \bar{x}\| .$$

Следователно при $\|B\| < 1$ скоростта на сходимост е както при геометрична прогресия.

При итерационни формули от вида

$$\bar{x}_{k+1} = (E - CA)\bar{x}_k + C\bar{b} \tag{1}$$

критериите за сходимост могат да се наложат направо на матрицата $A = \{a_{ij}\}$. Да разгледаме специалния случай на (1), когато C е диагонална матрица с елементи по диагонала $1/a_{ii}$,

$$\begin{aligned} C &= diag\left\{\frac{1}{a_{11}}, \dots, \frac{1}{a_{nn}}\right\} \\ &= diag\{a_{11}, \dots, a_{nn}\}^{-1} . \end{aligned}$$

В този случай системата $A\bar{x} = \bar{b}$ се преобразува по стандартния начин – от i-тото уравнение се определя x_i:

$$x_i = -\frac{a_{i1}}{a_{ii}}x_1 - \dots - \frac{a_{i,i-1}}{a_{ii}}x_{i-1} - \frac{a_{i,i+1}}{a_{ii}}x_{i+1} - \dots - \frac{a_{in}}{a_{ii}}x_n + \frac{b_i}{a_{ii}}, \quad i = 1, \dots, n,$$

и формулите (1) за пресмятане на следващите приближения $x_1^{(k+1)}, \dots, x_n^{(k+1)}$ добиват вида

$$x_i^{(k+1)} = -\sum_{j=1, j\neq i}^{n} \frac{a_{ij}}{a_{ii}}x_j^{(k)} + \frac{b_i}{a_{ii}}, \quad i = 1, \dots, n. \tag{2}$$

Този метод е известен като *метод на простата итерация*.

Да видим как изглеждат достатъчните условия за сходимост на метода на простата итерация, които произлизат от следствието по-горе, при използване на нормата

$$\|B\|_\infty := \max_{1\leq i\leq n} \sum_{j=1}^{n} |b_{ij}| .$$

В нашия случай $B = E - CA$. Нека $\{b_{ij}\}$, $\{c_{ij}\}$ и $\{\delta_{ij}\}$ са елементите на B, C и E, съответно. Тогава

$$\begin{aligned} b_{ij} &= \delta_{ij} - c_{i1}a_{1j} - \cdots - c_{in}a_{nj} \\ &= \delta_{ij} - \frac{1}{a_{ii}}a_{ij} \end{aligned}$$

и следователно

$$\begin{aligned} \|B\|_\infty &= \max_{1\le i\le n} \sum_{j=1}^{n} |b_{ij}| \\ &= \max_{1\le i\le n} \sum_{j=1}^{n} |\delta_{ij} - \frac{a_{ij}}{a_{ii}}| \\ &= \max_{1\le i\le n} \frac{1}{|a_{ii}|} \sum_{j=1, j\neq i}^{n} |a_{ij}| \ . \end{aligned}$$

Оттук се вижда, че условието $\|B\|_\infty < 1$ се записва като

$$\sum_{j=1, j\neq i}^{n} |a_{ij}| < |a_{ii}|, \quad i = 1, \ldots, n \ .$$

Това всъщност е условието A да бъде матрица с *доминиращ главен диагонал*.

Аналогично, условието

$$\|B\|_1 := \max_{1\le j\le n} \sum_{i=1}^{n} |b_{ij}| < 1$$

се свежда към

$$\sum_{i\neq j}^{n} \left|\frac{a_{ij}}{a_{ii}}\right| < 1, \quad j = 1, \ldots, n.$$

Метод на Зайдел. На практика често се използва една естествена модификация на метода на простата итерация, наречена *метод на Зайдел*. При нея в поредното i-то уравнение от (2) за определяне на $x_i^{(k+1)}$ се използват изчислените вече $(k+1)$-ви приближения на $x_1, \ldots, x_{i-1}$. По този начин се получават формулите

$$x_i^{(k+1)} = -\sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^{n} \frac{a_{ij}}{a_{ii}} x_j^{(k)} + \frac{b_i}{a_{ii}}, \quad i = 1, \ldots, n \ .$$

Теорема 3. *Методът на Зайдел е сходящ при произволно начално приближение $\bar{x}_0$ тогава и само тогава, когато всички корени на уравнението*

$$\det \begin{vmatrix} \lambda a_{11} & a_{12} & \ldots & a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{n1} & \lambda a_{n2} & \ldots & \lambda a_{nn} \end{vmatrix} = 0$$

са по модул по-малки от 1.

Доказателство. Да представим A като $A = U + V$, където U е долна триъгълна матрица, включваща главния диагонал, а V е горна триъгълна с елементи 0 по главния диагонал и под него. Тогава системата $A\bar{x} = \bar{b}$ се записва във вида

$$U\bar{x} = -V\bar{x} + \bar{b}$$

и методът на Зайдел се представя чрез итерационния процес

$$U\bar{x}_{k+1} = -V\bar{x}_k + \bar{b}.$$

Решаваме относно $\bar{x}_{k+1}$ и получаваме

$$\bar{x}_{k+1} = -U^{-1}V\bar{x}_k + U^{-1}\bar{b}\ . \tag{3}$$

Но това е изчислителен процес от типа на метода на простата итерация, който беше разгледан в Теорема 1. Съгласно тази теорема методът (3) е сходящ тогава и само тогава, когато собствените стойности на матрицата $-U^{-1}V$ са по модул по-малки от 1, т.е., когато корените на уравнението

$$\det[-U^{-1}V - \lambda E] = (-1)^n \det[\lambda E + U^{-1}V] = 0,$$

или което е същото (като умножим двете страни с $\det U$), на уравнението

$$\det[\lambda U + V] = 0,$$

са по модул по-малки от 1. Теоремата е доказана.

Сравняване на метода на Зайдел и метода на простата итерация. Областите на сходимост на простата итерация и метода на Зайдел се пресичат. Не е трудно да се покаже, че и методът на Зайдел е сходящ за системата $A\bar{x} = \bar{b}$, когато матрицата A е с доминиращ главен диагонал. Подолу ще покажем, че в този случай методът на Зайдел е по-бързо сходящ (в известен смисъл) от метода на простата итерация.

Теорема 4. *Ако матрицата A има доминиращ главен диагонал, то методът на Зайдел е по-бързо сходящ от метода на простата итерация.*

Доказателство. Това твърдение трябва да се разбира в следния смисъл: Ще намерим еднотипни оценки за скоростта на сходимост на метода на простата итерация и метода на Зайдел и ще покажем, че оценката при метода на Зайдел е по-добра от тази при простата итерация.

Ще използваме векторната норма $\|\bar{x}\|_\infty := \max\limits_{1\le i\le n} |x_i|$ и съответната съгласувана матрична норма $\|\cdot\|_\infty$.

Нека $A = \{a_{ij}\}$ е произволна матрица с доминиращ главен диагонал, т.е.

$$\sum_{j=1, j\neq i}^{n} |a_{ij}| < |a_{ii}|, \quad i = 1, \ldots, n.$$

Означаваме:

$$\begin{aligned} c_{ij} &= -\frac{a_{ij}}{a_{ii}}, \\ d_i &= \frac{b_i}{a_{ii}}, \\ \mu &= \max_{1\le i\le n} \frac{\sum_{j\ne i}^n |a_{ij}|}{|a_{ii}|} = \max_{1\le i\le n} \sum_{j=1, j\ne i}^{n} |c_{ij}| \,. \end{aligned}$$

Да отбележим, че съгласно нашето предположение за A, $\mu < 1$. На простата итерация съответства схемата

$$x_i^{(k+1)} = \sum_{j=1, j\ne i}^{n} c_{ij} x_j^{(k)} + d_i,$$

а на метода на Зайдел –

$$x_i^{(k+1)} = \sum_{j=1}^{i-1} c_{ij} x_j^{(k+1)} + \sum_{j=i+1}^{n} c_{ij} x_j^{(k)} + d_i.$$

Нека $\bar{x}$ е решението на системата. Имаме

$$x_i = \sum_{j=1, j\ne i}^{n} c_{ij} x_j + d_i.$$

За грешката по метода на простата итерация получаваме

$$\begin{aligned} \|\bar{x} - \bar{x}_{k+1}\|_\infty &\le \max_{1\le i\le n} \sum_{j=1, j\ne i}^{n} |c_{ij}| \left| x_j^{(k)} - x_j \right| \\ &\le \mu \|\bar{x}_k - \bar{x}\|_\infty \le \cdots \le \\ &\le \mu^{k+1} \|\bar{x}_0 - \bar{x}\|_\infty \,. \end{aligned}$$

Да въведем още означенията:

$$\begin{aligned} \beta_i &= \sum_{j=1}^{i-1} |c_{ij}| \,, \\ \gamma_i &= \sum_{j=i+1}^{n} |c_{ij}| \,, \\ \nu &= \max_i \frac{\gamma_i}{1-\beta_i} \,. \end{aligned}$$

За метода на Зайдел имаме

$$\|\bar{x} - \bar{x}_{k+1}\|_\infty = \max_{1\le i\le n} \left| x_i - x_i^{(k+1)} \right| \le$$

$$
\begin{aligned}
&\leq \max_i \left\{ \sum_{j=1}^{i-1} |c_{ij}| \left| x_j^{(k+1)} - x_j \right| + \sum_{j=i+1}^{n} |c_{ij}| \left| x_j^{(k)} - x_j \right| \right\} \\
&\leq \max_i \{\beta_i \|\bar{x}_{k+1} - \bar{x}\|_\infty + \gamma_i \|\bar{x}_k - \bar{x}\|_\infty\} \ .
\end{aligned}
$$

Последното равенство се достига за някакво i_0. Следователно

$$
\|\bar{x} - \bar{x}_{k+1}\|_\infty \leq \beta_{i_0} \|\bar{x}_{k+1} - \bar{x}\|_\infty + \gamma_{i_0} \|\bar{x}_k - \bar{x}\|_\infty \ .
$$

Оттук

$$
\begin{aligned}
\|\bar{x} - \bar{x}_{k+1}\|_\infty &\leq \frac{\gamma_{i_0}}{1 - \beta_{i_0}} \|\bar{x} - \bar{x}_k\|_\infty \\
&\leq \nu \|\bar{x} - \bar{x}_k\|_\infty \leq \cdots \leq \\
&\leq \nu^{k+1} \|\bar{x}_0 - \bar{x}\|_\infty \ .
\end{aligned}
$$

Но $\beta_i + \gamma_i \leq \mu < 1$. Тогава

$$
\begin{aligned}
\beta_i + \gamma_i - \frac{\gamma_i}{1 - \beta_i} &= \frac{\beta_i(1 - \beta_i) - \gamma_i \beta_i + \gamma_i - \gamma_i}{1 - \beta_i} \\
&= \frac{\beta_i(1 - \beta_i - \gamma_i)}{1 - \beta_i} \geq 0 \ .
\end{aligned}
$$

Следователно

$$
\mu = \max_i (\beta_i + \gamma_i) \geq \max_i \frac{\gamma_i}{1 - \beta_i} = \nu \ .
$$

И така, грешката при метода на Зайдел се оценява с израз, който клони към нула по-бързо от този, получен при простата итерация.

26. ГРАДИЕНТНИ МЕТОДИ ЗА РЕШАВАНЕ НА СИСТЕМИ ОТ ЛИНЕЙНИ УРАВНЕНИЯ

Нека е дадена системата

$$A\bar{x} = \bar{b}. \tag{1}$$

Ще предполагаме, че A е положително определена матрица. Да въведем функционала

$$f(\bar{x}) := (A\bar{x}, \bar{x}) - 2(\bar{b}, \bar{x}), \tag{2}$$

определен за всяко $\bar{x} \in \mathbb{R}^n$. Вижда се, че

$$f(\bar{x}) = \sum_{i,j=1}^{n} a_{ij} x_j x_i - 2 \sum_{i=1}^{n} b_i x_i \ .$$

Следователно $f(\bar{x})$ е полином от втора степен по отношение на x_i, $i = 1, \ldots, n$.

Ще покажем, че решението на (1) минимизира (2) и обратно – минимума на функционала (2) се достига за решение на (1). Наистина, нека $\bar{\xi}$ е решение на (1), т.е. $A\bar{\xi} = \bar{b}$. Нека $\bar{x}$ е произволен елемент от $\mathbb{R}^n$. Като използваме положителната определеност на A, получаваме последователно:

$$\begin{aligned}
f(\bar{x}) - f(\bar{\xi}) &= (A\bar{x}, \bar{x}) - 2(\bar{b}, \bar{x}) - (A\bar{\xi}, \bar{\xi}) + 2(\bar{b}, \bar{\xi}) \\
&= (A\bar{x}, \bar{x}) - 2(A\bar{\xi}, \bar{x}) - (A\bar{\xi}, \bar{\xi}) + 2(A\bar{\xi}, \bar{\xi}) \\
&= (A\bar{\xi}, \bar{\xi}) - (A\bar{\xi}, \bar{x}) + (A\bar{x}, \bar{x}) - (A\bar{\xi}, \bar{x}) \\
&= (A\bar{\xi}, \bar{\xi} - \bar{x}) + (A\bar{x} - A\bar{\xi}, \bar{x}) \\
&= (A\bar{\xi}, \bar{\xi} - \bar{x}) - (A(\bar{\xi} - \bar{x}), \bar{x}) \\
&= (A\bar{\xi}, \bar{\xi} - \bar{x}) - (\bar{\xi} - \bar{x}, A\bar{x}) \\
&= (A(\bar{\xi} - \bar{x}), \bar{\xi} - \bar{x}) \geq 0 \ .
\end{aligned}$$

Равенството се достига само при $\bar{x} = \bar{\xi}$. Следователно $f(\bar{x})$ приема минималната си стойност само при $\bar{x} = \bar{\xi}$.

Да допуснем сега, че $f(\bar{x}) \geq f(\bar{\xi})$ за всяко $\bar{x} \in \mathbb{R}^n$. Тогава

$$\frac{\partial f}{\partial x_i}\Big|_{\bar{x}=\bar{\xi}} = 2 \sum_{j=1}^{n} a_{ij} \xi_j - 2b_i = 0, \quad i = 1, \ldots, n,$$

което означава, че $\bar{\xi}$ е решение на линейната система $A\bar{x} = \bar{b}$. Твърдението е доказано.

Описаната връзка се използва за приближено решение на системата (1) чрез минимизиране на (2). Един от методите за търсене на минимума на (2) е *методът на най-бързото спускане.* При него от едно приближение $\bar{x}_k$ се намира следващото $\bar{x}_{k+1}$ така, че да имаме най-голямо намаление на стойността на $f(\bar{x})$. Всяко следващо приближение се получава от предходното по формулата

$$\bar{x}_{k+1} = \bar{x}_k + \alpha_k \bar{c}_k,$$

където $\bar{c}_k$ има направление на най-бързото намаляване на $f(\bar{x})$, започвайки от точката $\bar{x}_k$, а α_k се избира така, че да се постигне възможно най-голямо намаление на стойността $f(\bar{x})$, ако $\bar{x}$ се движи по това направление. С други думи $\bar{c}_k$ се избира от условието

$$\left| \frac{d}{d\alpha}[f(\bar{x}_k + \alpha\bar{c})] \Big|_{\alpha=0} \right| \to \max ,$$

а при така определеното $\bar{c} = \bar{c}_k$ търсим α_k като решение на уравнението

$$\frac{d}{d\alpha} f(\bar{x}_k + \alpha\bar{c}) = 0.$$

Имаме

$$\begin{aligned} f(\bar{x}_k + \alpha\bar{c}) &= (\bar{x}_k + \alpha\bar{c}, A(\bar{x}_k + \alpha\bar{c})) - 2(\bar{b}, \bar{x}_k + \alpha\bar{c}) \\ &= (\bar{x}_k, A\bar{x}_k) + 2\alpha(\bar{x}_k, A\bar{c}) + \alpha^2(\bar{c}, A\bar{c}) - 2(b, \bar{x}_k) - 2\alpha(\bar{b}, \bar{c}) \\ &= \alpha^2(A\bar{c}, \bar{c}) + f(\bar{x}_k) + 2\alpha(A\bar{x}_k - \bar{b}, \bar{c}). \end{aligned}$$

Оттук, като използваме означението $\bar{r}_k := A\bar{x}_k - \bar{b}$, получаваме

$$\frac{d}{d\alpha}[f(\bar{x}_k + \alpha\bar{c}]|_{\alpha=0} = 2(A\bar{x}_k - \bar{b}, \bar{c}) = 2(\bar{r}_k, \bar{c}),$$

Съгласно неравенството на Коши-Буняковски-Шварц,

$$|(\bar{r}_k, \bar{c})| \leq \|\bar{r}_k\| \cdot \|\bar{c}\|,$$

като равенството се достига когато $\bar{c}$ и $\bar{r}_k$ са колинеарни, т.е. когато $\bar{c} = const. \bar{r}_k$. Следователно направлението на най-бързото намаление на функционала f се дава с направлението на вектора $c_k := \bar{r}_k = A\bar{x}_k - \bar{b}$. Да определим стойността α_k на α, при която f приема минимална стойност по направлението $\bar{c}_k$. Имаме

$$f(\bar{x}_k + \alpha\bar{c}_k) = f(\bar{x}_k + \alpha\bar{r}_k) = \alpha^2(A\bar{r}_k, \bar{r}_k) + f(\bar{x}_k) + 2\alpha(\bar{r}_k, \bar{r}_k).$$

Оттук

$$\frac{d}{d\alpha}f(\bar{x}_k + \alpha\bar{c}_k) = 2\alpha(A\bar{r}_k, \bar{r}_k) + 2(\bar{r}_k, \bar{r}_k) = 0.$$

Определяме решението α_k на горното уравнение. Получаваме

$$\alpha_k = -\frac{(\bar{r}_k, \bar{r}_k)}{(A\bar{r}_k, \bar{r}_k)}.$$

Следователно формулата за пресмятане на следващото приближение приема окончателния вид:

$$\bar{x}_{k+1} = \bar{x}_k - \frac{(\bar{r}_k, \bar{r}_k)}{(A\bar{r}_k, \bar{r}_k)}\bar{r}_k \ . \ .$$

Тъй като функцията $f(\bar{x})$ има единствен локален минимум, то методът е сходящ при всяко начално приближение.

27. РЕШАВАНЕ НА СИСТЕМИ ОТ НЕЛИНЕЙНИ УРАВНЕНИЯ

Метод на простата итерация. Трябва да решим системата

$$f_i(x_1,\dots,x_n)=0,\quad i=1,\dots,n,$$

където f_i са нелинейни функции. Записваме тази система във вида

$$x_i=\varphi_i(x_1,\dots,x_n),\quad i=1,\dots,n,$$

или по-кратко

$$\bar{x}=\bar{\varphi}(\bar{x}),$$

където

$$\begin{aligned}\bar{x}&=(x_1,\dots,x_n),\\ \bar{\varphi}(\bar{x})&=(\varphi_1(\bar{x}),\dots,\varphi_n(\bar{x})).\end{aligned}$$

Както и в едномерния случай (при решаване на нелинейни уравнения) построяваме итерационния процес

$$\bar{x}_{k+1}=\bar{\varphi}(\bar{x}_k),\quad k=0,1,\dots\ ,$$

при някакво начално приближение $\bar{x}_0=(x_{01},\dots,x_{0n})$. В по-подробен запис формулите изглеждат така:

$$x_i^{(k+1)}=\varphi_i(x_1^{(k)},\dots,x_n^{(k)}),\quad i=1,\dots,n. \tag{1}$$

Този метод е известен като *метод на простата итерация*. Аналог пък на *метода на Зайдел* за нелинейни системи е итерационният метод

$$\begin{aligned}x_1^{(k+1)}&=\varphi_1(x_1^{(k)},\dots,x_n^{(k)})\\ x_2^{(k+1)}&=\varphi_2(x_1^{(k+1)},x_2^{(k)},\dots,x_n^{(k)})\\ \dots\ &\ .\ \dots\dots\dots\dots\dots\dots\\ x_n^{(k+1)}&=\varphi_n(x_1^{(k+1)},\dots,x_{n-1}^{(k+1)},x_n^{(k)}).\end{aligned}$$

Да изследваме скоростта на сходимост на метода на простата итерация (1).

Въвеждаме в $\mathbb{R}^n$ нормата $\|\bar{x}\|_\infty=\max\limits_{1\le i\le n}|x_i|$. Нека $S_r:=\{\bar{x}\in\mathbb{R}^n:\|\bar{x}-\bar{\xi}\|_\infty\le r\}$, където $\bar{\xi}$ е решението на системата $\bar{x}=\varphi(\bar{x})$. Да предположим, че

$$\left|\frac{\partial\varphi_i(\bar{x})}{\partial x_j}\right|\le M_{ij}\quad\text{при}\quad \bar{x}\in S_r,\quad\text{за}\quad i,j=1,\dots,n.$$

Теорема 1. *Ако*

$$\max_{1\le i\le n}\sum_{j=1}^{n} M_{ij} < 1,$$

то $\bar{x}_k$ клони към $\bar{\xi}$ със скорост на геометрична прогресия при всяко $\bar{x}_0 \in S_r$.

Доказателство. Ако означим с M матрицата $\{M_{ij}\}$, то условието в теоремата е всъщност $\|M\|_\infty < 1$.

Нека $\bar{x} = (x_1, \dots, x_n) \in S_r$. Тогава

$$\begin{aligned}
\|\bar{\varphi}(\bar{x}) - \bar{\xi}\|_\infty &= \|\bar{\varphi}(\bar{x}) - \bar{\varphi}(\bar{\xi})\|_\infty = \max_{1\le i\le n} |\varphi_i(x_1,\dots,x_n) - \varphi_i(\xi_1,\dots,\xi_n)| \\
&= \max_{1\le i\le n}\left|\sum_{j=1}^{n}\frac{\partial\varphi_i(\eta_{ij})}{\partial x_j}(x_j-\xi_j)\right| \quad \text{(по Тейлър)} \\
&\le \|\bar{x}-\bar{\xi}\|_\infty \max_{1\le i\le n}\sum_{j=1}^{n}\left|\frac{\partial\varphi_i(\eta_{ij})}{\partial x_j}\right|,
\end{aligned}$$

където $\bar{\eta}_i = (\eta_{i1}, \dots, \eta_{in})$ са някакви точки по линията сързваща $\bar{x}$ и $\bar{\xi}$. Тогава очевидно $\bar{\eta}_i \in S_r$, тъй като $\bar{\xi}$ и $\bar{x}$ принадлежат на S_r. Следователно

$$\|\bar{\varphi}(\bar{x}) - \bar{\xi}\|_\infty \le \|\bar{x}-\bar{\xi}\|_\infty \max_{1\le i\le n}\sum_{j=1}^{n} M_{ij} = q\|\bar{x}-\bar{\xi}\|_\infty ,$$

където $q = \|M\|_\infty < 1$. Оттук получаваме (при $\bar{x} = \bar{x}_k$)

$$\|\bar{x}_{k+1} - \bar{\xi}\|_\infty \le q\|\bar{x}_k - \bar{\xi}\|_\infty \le q^{k+1}\|\bar{x}_0 - \bar{\xi}\|_\infty$$

и теоремата е доказана.

Метод на Нютон. Да припомним най-напред метода на Нютон за решаване на уравнението $f(x) = 0$. Нека $f \in C^2[a,b]$. Да предположим, че сме намерили някякво приближение x_k на корена на уравнението $f(x) = 0$ в $[a,b]$. По формулата на Тейлър

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + f''(\xi)\frac{(x-x_k)^2}{2!} ,$$

с някакво ξ, намиращо се между точките x и x_k. Вместо да решаваме уравнението $f(x) = 0$ ние търсим корена на линейното уравнение

$$f(x_k) + f'(x_k)(x - x_k) = 0.$$

Да означим корена му с x_{k+1}. Имаме

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Това е формулата на Нютон.

Да постъпим аналогично и при решаване на система от уравнения от вида

$$F(\bar{x}) = \bar{o}, \tag{1}$$

където $F(\bar{x}) = (f_1(\bar{x}), \dots, f_n(\bar{x}))$, т.е. разглеждаме системата

$$\begin{aligned} f_1(x_1, \dots, x_n) &= 0 \\ \dots\dots\dots\dots\dots & \dots . \\ f_n(x_1, \dots, x_n) &= 0. \end{aligned}$$

Нека $\bar{x}_k = (x_1^{(k)}, \dots, x_n^{(k)})$ е приближено решение на (1), намиращо се в околност на точното решение $\bar{\xi}$ и нека $\frac{\partial^2 f_i}{\partial x_j \partial x_m}$ са непрекъснати в тази околност. По формулата на Тейлър имаме

$$\begin{aligned} f_i(x_1, \dots, x_n) &= f_i(x_1^{(k)}, \dots, x_n^{(k)}) + \sum_{j=1}^{n} \frac{\partial f_i(\bar{x}_k)}{\partial x_j}(x_j - x_j^{(k)}) \\ &+ \text{ нелинейна част }. \end{aligned}$$

И тук, вместо нелинейната система $F(\bar{x}) = \bar{o}$, решаваме линейната

$$f_i(\bar{x}_k) + \sum_{j=1}^{n} \frac{\partial f_i(\bar{x}_k)}{\partial x_j}(x_j - x_j^{(k)}) = 0, \quad j = 1, \dots, n.$$

Означаваме решението с $\bar{x}_{k+1}$. Имаме

$$\sum_{j=1}^{n} \frac{\partial f_i(\bar{x}_k)}{\partial x_j}(x_j^{(k+1)} - x_j^{(k)}) = -f_i(\bar{x}_k). \tag{2}$$

Нека $J(\bar{x}_k)$ е матрицата на горната линейна система, т.е. $J(\bar{x}_k)$ е якобиана $\frac{D(f_1,\dots,f_n)}{D(x_1,\dots,x_n)}$ в точката $\bar{x}_k$. Системата (2) се записва в матрична форма по следния начин

$$J(\bar{x}_k)(\bar{x}_{k+1} - \bar{x}_k) = -F(\bar{x}_k) .$$

За да я решим, умножаваме двете страни с $J^{-1}(\bar{x}_k)$ и получаваме

$$\bar{x}_{k+1} = \bar{x}_k - J^{-1}(\bar{x}_k)\, F(\bar{x}_k).$$

Това е съответната формула за приближено решаване на системи по *метода на Нютон.*

Да се спрем на скоростта на сходимост на този метод.

Нека $\|A\|$ е матричната норма, подчинена на векторната норма $\|\bar{x}\| = \max\limits_{1\le i\le n} |x_i|$. Знаем, че

$$\begin{aligned} \|A\| &= \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}| \, , \\ \|A\bar{x}\| &\le \|A\| \cdot \|\bar{x}\| \, . \end{aligned}$$

Теорема 2. *Нека съществуват числа* $\hat{r}$, c_1, c_2 *такива, че*

(*a*) $\|J^{-1}(\bar{x})\| \le c_1$, *за всяко* $\bar{x} \in S_r$;

(*б*) $\|F(\bar{x}) - F(\bar{y}) - J(\bar{y})(\bar{x} - \bar{y})\| \le c_2 \|\bar{x} - \bar{y}\|^2$, $\bar{x}, \bar{y} \in S_r$.

Тогава при $x_0 \in S_{\hat{r}}$, $\hat{r} = \min(r, \frac{1}{c})$, $c = c_1 c_2$, *методът на Нютон е сходящ със скорост*

$$\|\bar{x}_k - \bar{\xi}\| \le \frac{1}{c}(c\|\bar{x}_0 - \bar{\xi}\|)^{2^k} \, .$$

Доказателство. Най-напред ще докажем по индукция, че от $\bar{x}_0 \in S_{\hat{r}}$ следва $\bar{x}_k \in S_{\hat{r}}$ за всяко k. Наистина, нека $\bar{x}_i \in S_{\hat{r}}$, $i = 0, \dots, k$. Имаме

$$\begin{aligned} F(\bar{\xi}) &- F(\bar{x}_k) - J(\bar{x}_k)(\bar{\xi} - \bar{x}_k) \\ &= -F(\bar{x}_k) - J(\bar{x}_k)(\bar{\xi} - \bar{x}_{k+1} + \bar{x}_{k+1} - \bar{x}_k) \\ &= -F(\bar{x}_k) - J(\bar{x}_k)(\bar{\xi} - \bar{x}_{k+1}) - J(\bar{x}_k)(\bar{x}_{k+1} - \bar{x}_k) \, . \end{aligned}$$

Но $F(\bar{x}_k) + J(\bar{x}_k)(\bar{x}_{k+1} - \bar{x}_k) = 0$, съгласно определението на $\bar{x}_{k+1}$. Следователно

$$F(\bar{\xi}) - F(\bar{x}_k) - J(\bar{x}_k)(\bar{\xi} - \bar{x}_k) = -J(\bar{x}_k)(\bar{\xi} - \bar{x}_{k+1}) \, .$$

Но по (б)

$$\|F(\bar{\xi}) - F(\bar{x}_k) - J(\bar{x}_k)(\bar{\xi} - \bar{x}_k)\| \le c_2 \|\bar{x}_k - \bar{\xi}\|^2 \, .$$

Оттук

$$\|J(\bar{x}_k)(\bar{x}_{k+1} - \bar{\xi})\| \le c_2 \|\bar{x}_k - \bar{\xi}\|^2 \, .$$

Умножаваме двете страни с $\|J^{-1}(\bar{x}_k)\|$. Получаваме

$$\|J^{-1}(\bar{x}_k)\| \cdot \|J(\bar{x}_k)(\bar{x}_{k+1} - \bar{\xi})\| \le c_2 \|J^{-1}(\bar{x}_k)\| \cdot \|\bar{x}_k - \bar{\xi}\|^2 \, .$$

Но матричната норма е съгласувана с векторната. Следователно

$$\begin{aligned} \|(\bar{x}_{k+1} - \bar{\xi})\| &= \|J^{-1}(\bar{x}_k) J(\bar{x}_k)(\bar{x}_{k+1} - \bar{\xi})\| \\ &\le \|J^{-1}(\bar{x}_k)\| \cdot \|J(\bar{x}_k)(\bar{x}_{k+1} - \bar{\xi})\| \le c_2 c_1 \|\bar{x}_k - \bar{\xi}\|^2 \end{aligned}$$

$$\leq \quad c\hat{r}^2 \leq c\hat{r}\hat{r} \leq \hat{r}\ .$$

И така, $\bar{x}_{k+1} \in S_{\hat{r}}$.

В оценката по-горе доказахме неравенството

$$\|\bar{x}_k - \bar{\xi}\| \leq c\,\|\bar{x}_{k-1} - \bar{\xi}\|^2\ .$$

Следователно

$$\begin{aligned} c\|\bar{x}_k - \bar{\xi}\| &\leq c^2\|\bar{x}_{k-1} - \bar{\xi}\|^2 \\ &= [c\|\bar{x}_{k-1} - \bar{\xi}\|]^2 \\ &\leq [c\|\bar{x}_{k-2} - \bar{\xi}\|]^4 \leq \ldots \leq [c\|\bar{x}_0 - \bar{\xi}\|]^{2^k}\ . \end{aligned}$$

Оттук

$$\|\bar{x}_k - \bar{\xi}\| \leq \frac{1}{c}[c\|\bar{x}_0 - \bar{\xi}\|]^{2^k},$$

т.е. методът на Нютон е сходящ със скорост q^{2^k}, където $q < 1$ при достатъчно добро начално приближение. Теоремата е доказана.

28. ЧИСЛО НА ОБУСЛОВЕНОСТ

Тук ще разгледаме един въпрос, свързан с някои проблеми, които могат да възникнат при практическо решаване на линейни системи.

Нека си мислим, се системата е дадена във вида

$$A\bar{x} = \bar{b} \ . \tag{1}$$

Данните тук са елементите $\{a_{ij}\}$ на A и елементите $\{b_i\}$ на дясната страна $\bar{b}$. Обикновено тези данни се задават експериментално или се получават след изчисления с определена точност. В много случаи матрицата A е известна от теоретични изследвания, а само $\bar{b}$ се взимат от конкретни практически задачи. Като добавим и факта, че в компютрите числата се закръгляват с определена точност, стигаме до извода, че на практика вместо системата (1) решаваме всъщност друга система, с променени данни. Практиката показва, че при някои системи малки изменения в данните не влияят съществено върху решението, докато при други това влияние е значително. Дори добре изпитани методи, които в много случаи са давали отлични резултати, се проявяват доста странно при някои конкретни системи и дават резултати, далеч от очакваните. На какво се дължи това? Преди да отговорим на този въпрос, нека най-напред да дадем една теоретична оценка за изменението на решението $\bar{x}$ при изменение на данните A и $\bar{b}$.

Нека $\bar{x}$ е решението на (1), а $\bar{x} + \bar{\varepsilon}$ – решението на системата с матрица $A + \Delta$ и дясна страна $\bar{b} + \bar{\delta}$, т.е.

$$(A + \Delta)(\bar{x} + \bar{\varepsilon}) = \bar{b} + \bar{\delta} \ . \tag{2}$$

Тук Δ е матрица, а $\bar{\varepsilon}$ и $\bar{\delta}$ са вектори. Като вземем предвид, че $A\bar{x} = \bar{b}$, от (2) получаваме

$$A\bar{\varepsilon} + \Delta\bar{x} + \Delta\bar{\varepsilon} = \bar{\delta}$$

и оттук

$$\bar{\varepsilon} = A^{-1}(\bar{\delta} - \Delta\bar{x} - \Delta\bar{\varepsilon}) \ .$$

Следователно

$$\|\bar{\varepsilon}\| \le \|A^{-1}\| \ \|\bar{\delta}\| + \|A^{-1}\| \ \|\Delta\| \ \|\bar{x}\| + \|A^{-1}\| \ \|\Delta\| \ \|\bar{\varepsilon}\| \ .$$

Ако предположим, че $\|A^{-1}\| \ \|\Delta\| < 1$, т.е. че грешката в елементите на A е достатъчно малка, получаваме

$$\|\bar{\varepsilon}\| \le \frac{\|A^{-1}\| \ \|\bar{\delta}\| + \|A^{-1}\| \ \|\Delta\| \ \|\bar{x}\|}{1 - \|A^{-1}\| \ \|\Delta\|} \ .$$

При точното задаване на A (при $\Delta = 0$) ще имаме

$$\|\bar{\varepsilon}\| \le \|A^{-1}\| \ \|\bar{\delta}\| \ . \tag{3}$$

И така, изменението $\bar{\varepsilon}$ на решението се оценява чрез изменението Δ и $\bar{\delta}$ в данните и зависи съществено от нормата на обратната матрица.

Тук $\bar{\delta}$, Δ и $\bar{\varepsilon}$ са абсолютните величини на грешките. Те обаче не дават достатъчно ясна представа за състоянието на нещата. Например, ако $\|\Delta\| = 1$, това голямо изменение на данните ли е? Зависи от величината на $\|A\|$. Ако $\|A\| = 10^6$, изменението е незначително, ако $\|A\| = 10^{-3}$, изменението е катастрофално. Затова при оценка на грешката се разглеждат относителните величини

$$\frac{\|\bar{\varepsilon}\|}{\|\bar{x}\|}, \quad \frac{\|\Delta\|}{\|A\|}, \quad \frac{\|\bar{\delta}\|}{\|\bar{b}\|} .$$

Сега ще изведем оценка чрез тези относителни величини. За целта ще ни бъде необходимо следното неравенство.

Лема 1. *Нека матричната норма $\|\cdot\|$ е съгласувана с векторната $\|\cdot\|$. Тогава за всяка обратима матрица A и вектор $\bar{x}$ е в сила неравенството*

$$\frac{\|\bar{x}\|}{\|A^{-1}\|} \leq \|A\bar{x}\| \leq \|A\| \; \|\bar{x}\| .$$

Доказателство. Дясното неравенство изразява всъщност съгласуваността на матричната и векторната норми. Лявото неравенство следва от факта, че

$$\|\bar{x}\| = \|A^{-1}A\bar{x}\| \leq \|A^{-1}\| \cdot \|A\bar{x}\| .$$

Да проследим влиянието на измененията на данните върху решението в две типични ситуации.

Нека в резултат на някакъв приближен числен метод сме получили приближено решение $\bar{\xi}$ на системата $A\bar{x} = \bar{b}$. Да заместим $\bar{\xi}$ в лявата страна на системата вместо $\bar{x}$. Получаваме $A\bar{\xi}$. Нека $A\bar{\xi}$ е близко до $\bar{b}$. Оттук следва ли, че и $\bar{\xi}$ ще е близко до $\bar{x}$? Струва ни се естествено, че ако $\bar{\delta} := A\bar{\xi} - \bar{b}$ е малко, то и $\bar{\varepsilon} := \bar{\xi} - \bar{x}$ ще е малко. Сега ще видим дали имаме основание за такова твърдение.

Имаме

$$\bar{\delta} = A\bar{\xi} - \bar{b} = A\bar{\xi} - A\bar{x} = A(\bar{\xi} - \bar{x}) = A\bar{\varepsilon} .$$

Следователно $\bar{\varepsilon} = A^{-1}\bar{\delta}$. Тъй като $(A^{-1})^{-1} = A$, то от Лема 1 следва неравенството

$$\frac{\|\bar{\delta}\|}{\|A\|} \leq \|A^{-1}\bar{\delta}\| = \|\bar{\varepsilon}\| \leq \|A^{-1}\| \; \|\bar{\delta}\| . \tag{4}$$

Аналогично,

$$\frac{\|\bar{b}\|}{\|A\|} \leq \|\bar{x}\| = \|A^{-1}\bar{b}\| \leq \|A^{-1}\| \; \|\bar{b}\| . \tag{5}$$

Като следствие от (4) и (5) получаваме следната оценка за относителната грешка

$$(6) \qquad \frac{1}{\|A^{-1}\|\,\|A\|}\frac{\|\bar{\delta}\|}{\|\bar{b}\|} \le \frac{\|\bar{\varepsilon}\|}{\|\bar{x}\|} \le \|A^{-1}\|\,\|A\|\frac{\|\bar{\delta}\|}{\|\bar{b}\|}\ .$$

Числото $\|A^{-1}\|\,\|A\|$ се нарича *число на обусловеност* на матрицата A и се бележи с $cond\ (A)$ (или $\nu(A)$). От (6) следва, че $cond\ (A) \ge 1$ (В противен случай бихме получили абсурд в (6) при $\|\bar{\delta}\| \ne 0$). Това се вижда и директно. Наистина, от равенството $E = A^{-1}A$ получаваме

$$\|E\| \le \|A^{-1}\|.\|A\| = cond\ (A)\ .$$

Но тъй като всички собствени стойности на единичната матрица E са равни на 1 (характеристичният полином на E е $(1-t)^n$) и всяка норма е по-голяма от модула на всяка собствена стойност (виж Лема 24.5), то $\|E\| \ge 1$ и следователно $cond\ (A) \ge 1$.

От (6) се вижда, че ако числото на обусловеност на A е близко до 1, то относителната грешка в решението е близка до относителната грешка в дясната страна. Тогава действително можем да кажем, че ако $A\bar{\xi}$ е близко до $\bar{b}$, то и $\bar{\xi}$ е приближение на решението $\bar{x}$ и дори да дадем добра оценка за грешката чрез (6).

Матрици с число на обусловеност близко до 1 се наричат *добре обусловени матрици*. Тези, при които $cond\ (A)$ е много голямо, се наричат *лошо обусловени*. Именно лошо обусловените матрици могат да предизвикат проблеми при числено решаване на съответната система.

Можем да дадем и по-добра оценка отдолу за $cond\ (A)$ чрез собствените стойности на A. За целта да означим с $\lambda_1, \dots, \lambda_n$ собствените стойности на A, подредени във възходящ ред (по модул),

$$|\lambda_1| \le \dots \le |\lambda_n|\ .$$

Тогава $\frac{1}{|\lambda_n|} \le \dots \le \frac{1}{|\lambda_1|}$ са модулите на собствените стойности на A^{-1} и следователно

$$(7) \qquad cond\ (A) = \|A^{-1}\|\,\|A\| \ge \frac{|\lambda_n|}{|\lambda_1|}\ .$$

В частния случай, когато A е симетрична матрица (т.е. $A = A^T$), $\|A\|_2 = |\lambda_n|$ и $\|A^{-1}\|_2 = \frac{1}{|\lambda_1|}$. Тогава

$$(8) \qquad cond\ (A) = \frac{|\lambda_n|}{|\lambda_1|}.$$

Следователно обусловеността на симетричните матрици зависи от ширината на спектъра (т.е. от частното на най-голямата и най-малка собствени стойности).

Да разгледаме още един случай, в който се проявява характеристиката "обусловеност" на матрицата A. Нека вместо системата $A\bar{x} = \bar{b}$ решаваме системата $\hat{A}\bar{\xi} = \bar{b}$, където $\hat{A} = A + \Delta$. Ще оценим разликата в решенията $\bar{x}$ и $\bar{\xi}$ на тези две системи. Имаме

$$\begin{aligned} \bar{x} = A^{-1}\bar{b} &= A^{-1}(\hat{A}\bar{\xi}) = A^{-1}(A + \hat{A} - A)\bar{\xi} \\ &= \bar{\xi} + A^{-1}(\hat{A} - A)\bar{\xi} = \bar{\xi} + A^{-1}\Delta\bar{\xi}\,. \end{aligned}$$

Оттук получаваме

$$\bar{x} - \bar{\xi} = A^{-1}\Delta\bar{\xi}$$

и следователно

$$\|\bar{x} - \bar{\xi}\| \le \|A^{-1}\|\ \|\Delta\|\ \|\bar{\xi}\| = \|A^{-1}\|\ \|A\|\ \frac{\|\Delta\|}{\|A\|}\ \|\bar{\xi}\|\,.$$

Окончателно

$$\frac{\|\bar{x} - \bar{\xi}\|}{\|\bar{\xi}\|} \le cond\ (A)\ \frac{\|\Delta\|}{\|A\|}\,.$$

Това неравенство показва, че при добре обусловена матрица малките относителни изменения в елементите на матрицата водят до малки изменения в решението.

От тези примери е ясно, че числото на обусловеност е важна характеристика на A. За да намерим това число трябва да знаем $\|A\|$ и $\|A^{-1}\|$. В общия случай изчисляването на последните норми не е проста числена задача. Понякога $cond\ (A)$ може да се оцени с помощта на следната теорема.

Теорема 2. *За произволна норма и произволна неособена матрица A е в сила равенството*

$$\frac{1}{cond\ (A)} = \min\left\{ \frac{\|A - B\|}{\|A\|}\ :\ B \text{ е необратима} \right\}.$$

Теоремата показва, че числото на обусловеност отразява близостта на A до необратима матрица B (т.е. такава, че $\det B = 0$). Ние няма да доказваме тук тази теорема. Ще покажем само, че

$$\frac{1}{cond\ (A)} \le \frac{\|A - B\|}{\|A\|} \quad \text{за всяка } B \text{ с } \det B = 0\,.$$

Наистина, горното неравенство е еквивалентно с

$$\frac{1}{\|A^{-1}\|} \le \|A - B\|\,. \tag{8}$$

Тъй като $\det B = 0$, то съществува ненулев вектор $\bar{x}$ такъв, че $B\bar{x} = \bar{o}$. Тогава

$$\begin{aligned} \|A - B\| \, \|\bar{x}\| &\geq \|A\bar{x} - B\bar{x}\| = \|A\bar{x}\| \\ &\geq \frac{\|\bar{x}\|}{\|A^{-1}\|} \qquad \text{(по Лема 1)} \ . \end{aligned}$$

Оттук следва (8), понеже $\|\bar{x}\| > 0$.

Много от числените методи за решаване на линейни системи се изразяват в следното: матрицата A се преобразува в матрица C с по-специфична структура (триъгълна, лентова, симетрична) и след това се решава системата, съответстваща на C. Понякога тези преобразования могат да доведат до увеличаване на числото на обусловеност на A. (Матрицата A от добре обусловена, може да се превърне в лошо обусловена матрица.)

Да видим какво става например при симетризация на матрицата. Да умножим двете страни на уравнението

$$A\bar{x} = \bar{b}$$

с транспонираната на A. Получаваме $A^T A\bar{x} = A^T\bar{b}$. Това е една друга система, еквивалентна на старата, но вече със симетрична матрица $C = A^T A$. Нека $|\lambda_1| \leq \ldots \leq |\lambda_n|$ са модулите на собствените стойности на A. Да приемем сега, че A е била симетрична. Тогава $A^T A = A^2$ и $\lambda_1^2, \ldots, \lambda_n^2$ са собствените стойности на A^2. Следователно

$$cond\ (C) = \left(\frac{|\lambda_n|}{|\lambda_1|}\right)^2 = [cond\ (A)]^2\ .$$

Но една матрица има число на обусловеност 1 само ако е единичната (с точност до множител). Следователно в общия случай (при $A \neq E$) $cond(A) > 1$ и от (9) следва, че при прилагане на операцията симетризация върху симетрични матрици числото на обусловеност нараства. Това показва (по непрекъснатост), че симетризацията може да развали обусловеността на A, например, когато матрицата е близка до симетрична.

Глава 5

ПРЕСМЯТАНЕ НА СОБСТВЕНИ СТОЙНОСТИ НА МАТРИЦИ

Собствени стойности на матрицата A са онези числа λ, при които уравнението

$$A\bar{x} = \lambda\bar{x}$$

има ненулево решение $\bar{x}$. Тези ненулеви решения се наричат *собствени вектори* на A. Ясно е, че всяка матрица A с размерност $n \times n$ има точно n собствени стойности, които са корените на алгебричното уравнение

$$D(\lambda) := \det(A - \lambda E) = 0.$$

Уравнението $D(\lambda) = 0$ се нарича *характеристично уравнение* на матрицата A. Може да се покаже, че

$$D(\lambda) = (-1)^n[\lambda^n - \sigma_1\lambda^{n-1} + \sigma_2\lambda^{n-2} - \cdots + (-1)^n\sigma_n],$$

където

$$\begin{aligned}
\sigma_1 &= \sum_{k=1}^{n} a_{kk} \\
\sigma_2 &= \sum_{i<k} \begin{vmatrix} a_{ii} & a_{ik} \\ a_{ki} & a_{kk} \end{vmatrix} \\
\sigma_3 &= \sum_{i<j<k} \begin{vmatrix} a_{ii} & a_{ij} & a_{ik} \\ a_{ji} & a_{jj} & a_{jk} \\ a_{ki} & a_{kj} & a_{kk} \end{vmatrix} \\
\dots &\dots \dots\dots\dots\dots\dots\dots \\
\sigma_n &= \det A.
\end{aligned}$$

За намирането на коефициентите на $D(\lambda)$ е необходимо да се пресметнат $2^n - 1 \quad (= \binom{n}{1} + \cdots + \binom{n}{n})$ детерминанти. При големи n това е доста трудоемка задача. Създадени са други, по-прости методи за построяване на характеристичния полином на дадена матрица A. След като е намерен полинома, неговите нули (които са собствените стойности на A) се намират после по известни вече числени методи.

Сега ще се запознаем с един стар универсален метод за построяване на характеристичния полином на дадена матрица.

29. МЕТОД НА ДАНИЛЕВСКИ

Нека $A = \{a_{ij}\}_{i,j=1}^{n}$ е дадената матрица. Нека

$$P = \begin{vmatrix} p_1 & p_2 & \cdots & p_{n-1} & p_n \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{vmatrix}$$

е съответстващата ѝ подобна на нея матрица на Фробениус, т.е.

$$P = C^{-1}AC,$$

където C е неособена матрица. Понеже подобните матрици имат едни и същи характеристични уравнения, то

$$\det(A - \lambda E) = \det(P - \lambda E) = D(\lambda).$$

Когато характеристичното уравнение $\det(A - \lambda E) = 0$ е получено в така наречения нормален вид на Фробениус, т.е. във вида

$$D(\lambda) = \begin{vmatrix} p_1 - \lambda & p_2 & p_3 & \cdots & p_n \\ 1 & -\lambda & 0 & \dots & 0 \\ 0 & 1 & -\lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -\lambda \end{vmatrix},$$

то като развием детерминанта по първия ред получаваме

$$\begin{aligned} D(\lambda) &= (p_1 - \lambda)(-\lambda)^{n-1} - p_2(-\lambda)^{n-2} + p_3(-\lambda)^{n-3} + \cdots + (-1)^{n-1}p_n \\ &= (-1)^n[\lambda^n - p_1\lambda^{n-1} - p_2\lambda^{n-2} - \cdots - p_n]. \end{aligned}$$

И така, ако една матрица е във вид на Фробениус, то нейният характеристичен полином може да се намери веднага. При метода на Данилевски матрицата A се преобразува в подобна на нея матрица P с помощта на $n-1$ преобразования на подобие, последователно преобразуващи редовете на A, започвайки от последния.

Нека си мислим, че с помощта на $n-k$ преобразования на подобие сме получили матрицата (да я означим пак с A)

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k-1} & a_{1k} & \cdots & a_{1n-1} & a_{1n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk-1} & a_{kk} & \cdots & a_{kn-1} & a_{kn} \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 1 & 0 \end{bmatrix},$$

чиито $k+1,\dots,n$-ти ред съвпада с този на P. Искаме k-тият ред $[a_{k1}\ \dots\ a_{k,k-1}\ a_{kk}\ \dots\ a_{kn}]$ да добие вида $[0\ \dots\ 1\ 0\ \dots\ 0]$. За тази цел правим следните преобразования:

1. При $a_{kk-1} \neq 0$, делим всички елементи от $(k-1)$-вия стълб на $a_{k,k-1}$.
2. От i–тия стълб изваждаме $(k-1)$-вия, умножен с a_{ki}, $i \neq k-1$.

Като подложим E на същите преобразования получаваме

$$M_k = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ m_{k-1,1} & m_{k-1,2} & \dots & m_{k-1,n} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

където

$$\begin{aligned} m_{k-1,k-1} &= \frac{1}{a_{k,k-1}} \\ m_{k-1,i} &= -\frac{a_{k,i}}{a_{k,k-1}}, \quad i \neq k-1. \end{aligned}$$

Да означим преобразуваната матрица с B. Съгласно казаното до сега, $B = AM_k$ и k-тият, ..., n-тият ред на B съвпадат с тези на P. За елементите b_{ij} на B намираме

$$\begin{aligned} b_{ij} &= a_{ij} - a_{kj}\frac{a_{i,k-1}}{a_{k,k-1}} \\ &= a_{ij} + m_{k-1,j}a_{i,k-1}, \quad i = 1,\dots,k, \quad j = 1,\dots,k-2,k,\dots,n \\ b_{i,k-1} &= \frac{a_{i,k-1}}{a_{k,k-1}} = a_{i,k-1}m_{k-1,k-1}, \quad i = 1,\dots,k. \end{aligned}$$

Получената матрица B не е подобна на A. За да я преобразуваме в подобна ще я умножим отляво с M_k^{-1}. Получаваме $C = M_k^{-1}B = M_k^{-1}AM_k$.

Вижда се лесно, че

$$M_k^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{kn} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad \rightarrow \quad k-1$$

Наистина, проверява се директно, че $M_k^{-1}M_k = E$.

От формулата $C = M_k^{-1}B$ намираме изразите за елементите c_{ji} на C:

$$\begin{aligned} c_{ji} &= b_{ji}, \quad j \neq k-1 \\ c_{k-1,i} &= a_{k1}b_{1i} + \ldots + a_{kn}b_{ni}, \quad i = 1, \ldots, n. \end{aligned}$$

Вижда се, че C има същите редове от $k-$ти до $n-$ти както B. Продължаваме преобразованието на $k-1$-вия, ..., 2-рия ред по същия начин.

Ако $a_{k,k-1} = 0$ в матрицата, получена след $n-k$ стъпки, то възможни са два случая:

а) Нека $a_{ki} \neq 0$ за някое $i < k-1$.

Тогава заменяме местата на $(k-1)$-вия и i-тия стълб. За да бъде преобразованието преобразование на подобие, разместваме местата и на i-тия и $(k-1)$-вия ред. Продължаваме по описаната по-горе процедура.

б) $a_{ki} = 0$, $i = 1, \ldots, k-1$. Тогава A има вида

$$A = \left[\begin{array}{ccc|cccc} a_{11} & \ldots & a_{1,k-1} & a_{1k} & \ldots & a_{1n-1} & a_{1n} \\ \vdots & \ldots & \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{k-1,1} & \ldots & a_{k-1,k-1} & a_{k-1,k} & \ldots & a_{k-1,n-1} & a_{k-1,n} \\ \hline 0 & \ldots & 0 & a_{kk} & \ldots & a_{k,n-1} & a_{kn} \\ 0 & \ldots & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \ldots & \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & 1 & 0 \end{array}\right]$$

$$= \left[\begin{array}{c|c} D_1 & L \\ \hline O & D_2 \end{array}\right],$$

където D_2 е във вид на Фробениус. Но тогава

$$\det(A - \lambda E) = \det(D_1 - \lambda E)\det(D_2 - \lambda E).$$

Прилагаме метода на Данилевски за матрицата D_1, която е с по-малка размерност.

Намиране на собствените вектори по метода на Данилевски. Нека λ е собствена стойност на A. Тогава λ е собствена стойност и на подобната ѝ матрица P. Ще намерим собствения вектор $\bar{y} = (y_1, \ldots, y_n)$ на P, съответстващ на λ. Имаме $P\bar{y} = \lambda\bar{y}$, т.е.

$$\begin{bmatrix} p_1 - \lambda & p_2 & \dots & p_n \\ 1 & -\lambda & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\lambda \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \cdot \\ \vdots \\ \cdot \\ y_n \end{bmatrix} = \bar{0}.$$

Оттук

$$\begin{array}{rcl} (p_1 - \lambda)y_1 + p_2y_2 + \ldots + p_ny_n & = & 0, \\ y_1 - \lambda y_2 & = & 0, \\ y_2 - \lambda y_3 & = & 0, \\ \dots\dots\dots\dots\dots & \dots & . \\ y_{n-1} - \lambda y_n & = & 0. \end{array}$$

Системата е хомогенна. Тя има безброй много решения, които са пропорционални. Полагаме $y_n = 1$. От системата определяме

$$y_{n-1} = \lambda, \quad y_{n-2} = \lambda^2, \quad \dots , \; y_1 = \lambda^{n-1}.$$

Нека $\bar{x}$ е собственият вектор на A, отговарящ на собствената стойност λ. Понеже

$$M_{n-1}^{-1} \dots M_1^{-1} A M_1 \dots M_{n-1} \bar{y} = \lambda\bar{y},$$

то

$$A M_1 \dots M_{n-1}\bar{y} = \lambda M_1 \dots M_{n-1} \bar{y}$$

и следователно

$$\bar{x} = M_1 \dots M_{n-1} \bar{y}$$

е собственият вектор на A, отговарящ на собствената стойност λ.

30. МЕТОД НА ЯКОБИ

Методът на Якоби е итерационен метод за приближено намиране на собствени стойности и собствени вектори на симетрични матрици. Той е бил предложен от Якоби през 1846 г. Тук ние ще разгледаме само варианта, отнасящ се до реалния случай, т.е. когато елементите на A са реални числа. Най-напред ще припомним някои факти от линейната алгебра.

Лема 1. *Всеки две подобни матрици имат едни и същи характеристични полиноми.*

Това твърдение беше вече доказано в Теорема 2 на лекция 24.

Да означим с $\bar{e}_k = (0, \dots, 1, \dots, 0)$, $k = 1, \dots, n$, съответните базисни вектори.

Лема 2. *Нека $D = diag[\lambda_1, \dots, \lambda_n]$, т.е. D е диагонална матрица с елементи по диагонала $\lambda_1, \dots, \lambda_n$. Тогава $\lambda_1, \dots, \lambda_n$ са собствени стойности на D и $\{\bar{e}_k\}_1^n$ са съответните собствени вектори на D.*

Доказателство. Наистина

$$\det(D - \lambda E) = (\lambda_1 - \lambda) \dots (\lambda_n - \lambda).$$

Освен това, веднага се вижда, че $D\bar{e}_k = \lambda_k \bar{e}_k, \;\; k = 1, \dots, n.$

Тук (за удобство) с A' ще означаваме транспонираната матрица на A. Матрицата T се нарича *ортогонална*, ако $TT' = E$. Ясно е, че $T^{-1} = T'$ при ортогоналните матрици.

Да означим с $S(A)$ *сферичната норма*, а с $S_p(A)$ – *следата* на матрицата $A = \{a_{ij}\}$:

$$S(A) := \left\{ \sum_{i,j=1}^{n} |a_{ij}|^2 \right\}^{\frac{1}{2}},$$

$$S_p(A) := \sum_{i=1}^{n} a_{ii}.$$

Лема 3. *Нека T е ортогонална матрица. Тогава*

$$S^2(A) = S^2(T^{-1}AT).$$

С други думи, преобразованието на подобие с ортогонална матрица не променя сферичната норма.

Доказателство. Ще използваме известните от линейната алгебра свойства: $(AB)' = B'A'$, $S_p(AA') = S^2(A)$. Имаме

$$S^2(T^{-1}AT) \;\; = \;\; S^2(T'AT) = S_p((T'AT)'T'AT)$$

$$
\begin{aligned}
&= S_p(T'A'TT'AT) = S_p(T'A'AT) \\
&= S_p((AT)'AT) = S^2(AT) \\
&= S^2((AT)') = S^2(T'A') \\
&= S_p((T'A')'T'A') = S_p(ATT'A') \\
&= S_p(AA') = S^2(A).
\end{aligned}
$$

Сега ще изложим основната идея, залегнала в метода на Якоби. Нека A е произволна симетрична матрица. Знаем от линейната алгебра, че A може да се сведе чрез ортогонално подобно преобразование до диагонална матрица D, т.е. съществува ортогонална матрица T такава, че

$$T^{-1}AT = D$$

или, което е същото,

$$T'AT = D.$$

Както знаем вече от Лема 1, характеристичният полином не се променя при това преобразование. Съгласно Лема 2, диагоналните елементи на D ще бъдат собствените стойности на A. И така, задачата ще бъде решена, ако намерим това преобразование T.

Тъй като

$$S^2(A) \geq \sum_{i=1}^{n} |a_{ii}|^2 \quad (\ = \text{само когато } A \text{ е диагонална }\ )$$

и $S^2(A)$ не се променя, съгласно Лема 3, при подобно преобразование с ортогонална матрица, то T може да се търси така, че сумата от квадратите на извъндиагоналните елементи на матрицата $T'AT$, при фиксирано A, да достига своя абсолютен нулев минимум или, което е същото, сумата от квадратите на диагоналните елементи на $T'AT$ да достига своя максимум, равен на $S^2(A)$.

Методът на Якоби предлага един итерационен процес за минимизиране на сумата на квадратите на извъндиагоналните елементи. На всяка стъпка се използва подобно преобразование с помощта на матрица от вида

$$
T_{ij}(\varphi) = \begin{array}{c}
\begin{bmatrix}
1 & & & & & & 0 \\
& \ddots & & & & & \\
& & \cos\varphi & \dots & -\sin\varphi & & \\
& & \vdots & & \vdots & & \\
& & \sin\varphi & \dots & \cos\varphi & & \\
& & & & & \ddots & \\
0 & & & & & & 1
\end{bmatrix}
\end{array}
\begin{array}{l}
\\ \\ \leftarrow\ i \\ \\ \leftarrow\ j \\ \\ \\
\end{array}
$$

Тя се различава от единичната матрица само по отбелязяните елементи от редовете i и j. Вижда се веднага, че $T_{ij}(\varphi)$ е ортогонална, т.е. $T'_{ij}(\varphi)\,T_{ij}(\varphi) = E$.

Нека на $(k-1)$-вата стъпка сме стигнали до някаква матрица A_{k-1}. На следващата k-та стъпка се прави преобразованието

$$A_k = T'_{i_k,j_k}(\varphi_k)\,A_{k-1}\,T_{i_k,j_k}(\varphi_k).$$

При това параметрите i_k, j_k и φ_k се избират така, че сумата от квадратите на извъндиагоналните елементи да се намали максимално. Това е идеята на метода на Якоби.

Сега ще намерим в явен вид $T_{i_kj_k}(\varphi_k)$.

Да означим с a_{pq} елементите на A_{k-1}. Нека

$$\begin{aligned} B &:= A_{k-1}\,T_{i_kj_k}(\varphi_k), \\ C &:= T'_{i_kj_k}(\varphi_k)\,B. \end{aligned}$$

По-долу ще изпускаме индекса k в означенията на параметрите i_k, j_k и φ_k. Имаме

$$B = \begin{bmatrix} \\ \quad A_{k-1} \quad \\ \\ \end{bmatrix} \begin{array}{c} \begin{bmatrix} 1 & & & & & 0 \\ & \ddots & & & & \\ & & \cos\varphi & \dots & -\sin\varphi & \\ & & \vdots & & \vdots & \\ & & \sin\varphi & \dots & \cos\varphi & \\ & & & & & \ddots \\ 0 & & & & & 1 \end{bmatrix} \end{array} \begin{array}{l} \\ \\ \leftarrow\ i \\ \\ \leftarrow\ j \\ \\ \\ \end{array}$$

$$C = \begin{bmatrix} 1 & & & & & 0 \\ & \ddots & & & & \\ & & \cos\varphi & \dots & \sin\varphi & \\ & & \vdots & & \vdots & \\ & & -\sin\varphi & \dots & \cos\varphi & \\ & & & & & \ddots \\ 0 & & & & & 1 \end{bmatrix} \begin{bmatrix} \\ \quad B \quad \\ \\ \end{bmatrix}.$$

За елементите $\{b_{pq}\}$ и $\{c_{pq}\}$ на B и C получаваме

$$\begin{aligned} b_{pq} &= a_{pq} \quad \text{при} \quad q \neq i, j, \\ \text{(1)} \qquad b_{pi} &= a_{pi}\cos\varphi + a_{pj}\sin\varphi, \end{aligned}$$

$$
\begin{aligned}
b_{pj} &= -a_{pi}\sin\varphi + a_{pj}\cos\varphi,
\end{aligned}
$$

$$
\begin{aligned}
c_{pq} &= b_{pq} \quad \text{при} \quad p \neq i, j, \\
c_{iq} &= b_{iq}\cos\varphi + b_{jq}\sin\varphi, \qquad (2)\\
c_{jq} &= -b_{iq}\sin\varphi + b_{jq}\cos\varphi.
\end{aligned}
$$

Да означим

$$
\begin{aligned}
\tilde{\sigma}^2 &= \sum_{p\neq q} |c_{pq}|^2, \\
\sigma^2 &= \sum_{p\neq q} |a_{pq}|^2,
\end{aligned}
$$

т.е. това е изследваната сума на преобразуваната и старата матрица. Да забележим, че при направеното преобразование се изменят само елементите от i-тия и j-тия ред и стълб. Като вземем предвид, че A_{k-1} и C са симетрични, получаваме

$$
\begin{aligned}
\tilde{\sigma}^2 = \sigma^2 \; &+ \; \sum_{q\neq i,j} [c_{iq}^2 + c_{jq}^2] + \sum_{p\neq i,j} [c_{pi}^2 + c_{pj}^2] \\
&- \; \sum_{q\neq i,j} [a_{iq}^2 + a_{jq}^2] - \sum_{p\neq i,j} [a_{pi}^2 + a_{pj}^2] \\
&+ \; 2c_{ij}^2 - 2a_{ij}^2.
\end{aligned}
$$

Но съгласно формулите (1) и (2), при $q \neq i, j$, имаме

$$
\begin{aligned}
c_{iq}^2 + c_{jq}^2 &= (b_{iq}\cos\varphi + b_{jq}\sin\varphi)^2 + (-b_{iq}\sin\varphi + b_{jq}\cos\varphi)^2 \\
&= (a_{iq}\cos\varphi + a_{jq}\sin\varphi)^2 + (-a_{iq}\sin\varphi + a_{jq}\cos\varphi)^2 \\
&= a_{iq}^2 + a_{jq}^2 \; .
\end{aligned}
$$

Аналогично, при $p \neq i, j$,

$$
\begin{aligned}
c_{pi}^2 + c_{pj}^2 &= b_{pi}^2 + b_{pj}^2 \\
&= (a_{pi}\cos\varphi + a_{pj}\sin\varphi)^2 + (-a_{pi}\sin\varphi + a_{pj}\cos\varphi)^2 \\
&= a_{pi}^2 + a_{pj}^2.
\end{aligned}
$$

Освен това

$$
2c_{ij}^2 = 2[b_{ij}\cos\varphi + b_{jj}\sin\varphi]^2
$$

$$
\begin{aligned}
&= 2[(-a_{ii}\sin\varphi + a_{ij}\cos\varphi)\cos\varphi + (-a_{ji}\sin\varphi + a_{jj}\cos\varphi)\sin\varphi]^2 \\
&= \frac{1}{2}[-a_{ii}\sin 2\varphi + 2a_{ij}\cos 2\varphi + a_{jj}\sin 2\varphi]^2 \\
&= \frac{1}{2}[-(a_{ii}-a_{jj})\sin 2\varphi + 2a_{ij}\cos 2\varphi]^2.
\end{aligned}
$$

Следователно,

(3) $$\hat{\sigma}^2 = \sigma^2 - 2a_{ij}^2 + \frac{1}{2}[-(a_{ii}-a_{jj})\sin 2\varphi + 2a_{ij}\cos 2\varphi]^2$$

Очевидно $\hat{\sigma}^2$ е минимално при тази трансформация, ако

(4) $$i,j: \quad |a_{ij}| = \max\{|a_{pq}| \ : \ p \neq q\}$$

и

(5) $$-(a_{ii}-a_{jj})\sin 2\varphi + 2a_{ij}\cos 2\varphi = 0.$$

От (4) определяме индексите i_k и j_k на трансформацията $T_{i_k j_k}(\varphi_k)$, а от (5) определяме φ_k.

Получаваме

$$\tan 2\varphi = \frac{2a_{ij}}{a_{ii}-a_{jj}} =: s.$$

Оттук $\varphi = \frac{1}{2}\arctan s$ и следователно

$$
\begin{aligned}
\sin\varphi &= sin\frac{\arctan s}{2} \\
\cos\varphi &= cos\frac{\arctan s}{2}.
\end{aligned}
$$

Като използваме известните връзки

$$
\begin{aligned}
\sin\frac{\alpha}{2} &= \pm\sqrt{\frac{1-\cos\alpha}{2}}, \\
\cos\frac{\alpha}{2} &= \pm\sqrt{\frac{1+\cos\alpha}{2}}, \\
|\cos(\arctan s)| &= \frac{1}{\sqrt{1+s^2}},
\end{aligned}
$$

получаваме

$$\sin\varphi_k = sign\ s\sqrt{\frac{1-\cos\arctan s}{2}}$$

$$= \quad sign\ s\left\{\frac{1}{2}\left(1-\frac{1}{\sqrt{1+s^2}}\right)\right\}^{\frac{1}{2}},$$

$$\cos\varphi_k \quad = \quad \sqrt{\frac{1+\cos\arctan s}{2}}$$

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

В такъв вид формулите са по-удобни за програмиране.

Сега ще покажем, че описаният процес е сходящ.

Наистина, от равенството (3), приложено за $(k+1)$-вата стъпка при съответни параметри i_k, j_k и φ_k получаваме

$$\sigma_{k+1}^2 = \sigma_k^2 - 2a_{i_k,j_k}^2 \ . \tag{6}$$

Но a_{i_k,j_k} е максималният по модул извъндиагонален елемент на A_k. Следователно

$$a_{i_k,j_k}^2 \geq \frac{\sigma_k^2}{n(n-1)}.$$

Оттук и от (6) стигаме до оценката

$$\begin{aligned}
\sigma_{k+1}^2 \quad &\leq \quad \sigma_k^2 - 2\frac{\sigma_k^2}{n(n-1)} = \sigma_k^2\left(1-\frac{2}{n(n-1)}\right) \\
&\leq \quad \sigma_{k-1}^2\left(1-\frac{2}{n(n-1)}\right)^2 \leq \cdots \\
&\leq \quad \sigma_0^2\left(1-\frac{2}{n(n-1)}\right)^{k+1} .
\end{aligned}$$

От това неравенство следва, че $\sigma_{k+1}^2 \to 0$ при $k \to \infty$, което значи, че A_{k+1} клони към диагонална матрица. Да я означим с D. Следователно при големи k $\ A_k \approx D$ и диагоналните елементи на A_k са приближения на собствените стойности на A, а стълбовете на матрицата

$$T_k := \prod_{l=0}^{k-1} T_{i_l,j_l}$$

представляват приближени собствени вектори на A. Наистина, $\bar{l}_k \approx \bar{e}_k$ е приближение на собствения вектор на $A_k = T_k^{-1}AT_k$. Оттук

$$T_k^{-1}AT_k\bar{l}_k \quad \approx \quad \lambda_k\bar{l}_k \ ,$$

$$AT_k\bar{l}_k \quad \approx \quad \lambda_kT_k\bar{l}_k.$$

Следователно $T_k\bar{l}_k$ ($\approx$ k-тия стълб на T_k) е приближение на собствения вектор на A, съответстващ на λ_k.

Глава 6

ЗАДАЧИ ЗА ДОМАШНА РАБОТА

1. Напишете полинома $p(x)$ от π_5, който удовлетворява интерполационните условия $p(-1) = p(0) = p(2) = p(3) = p(5) = 0$, $p(4) = 40$.

2. Нека $x_0 < \cdots < x_n$ и $\omega(x) := (x - x_0)\cdots(x - x_n)$. Покажете, че

$$\frac{1}{\omega(x)} = \sum_{k=0}^{n} \frac{c_k}{x - x_k},$$

като коефициентите в това разложение се определят по формулата

$$c_k = \frac{1}{\omega'(x_k)}, \quad k = 0, \ldots, n.$$

3. Да се докаже тъждеството

$$\sum_{k=0}^{n} (x - x_k)^{n+1} \ell_{nk}(x) = (-1)^n \omega(x),$$

където $x_0, \ldots, x_n$ са различни точки, $\omega(x) = (x - x_0)\cdots(x - x_n)$ и

$$\ell_{nk}(x) = \frac{\omega(x)}{(x - x_k)\omega'(x_k)}.$$

4. Нека a и b са дадени различни числа. Пресметнете детерминантата

$$\begin{bmatrix} x & x-a & x-a & \cdots & x-a \\ x-b & x & x-a & \cdots & x-a \\ x-b & x-b & x & \cdots & x-a \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x-b & x-b & x-b & \cdots & x \end{bmatrix}.$$

5. Нека полиномът $P(x) = a_0 + a_1 x + \cdots + a_n x^n$ приема цели стойности в $n+1$ последователни цели числа. Докажете, че числата $n!\, a_k$, $k = 0, \ldots, n$, са цели.

6. Нека $P_n(x)$ е полином от степен $\leq n$, който удовлетворява интерполационните условия:

$$P_n(x_j) = 0 \ \text{ за } \ j = 0, \ldots k; \quad P_n(x_j) = 1 \ \text{ за } \ j = k+1, \ldots, n.$$

Намерете знака на коефициента пред x^n в $P_n(x)$.

7. Нека $x_0 < x_1 < \cdots < x_n$ и

$$\ell_{nk}(x) := \frac{\omega(x)}{(x - x_k)\omega'(x_k)}, \quad \omega(x) := (x - x_0)\cdots(x - x_n).$$

Докажете, че

$$\ell_{nk}(x) + \ell_{n,k+1}(x) \geq 1 \quad \text{за} \;\; x \in (x_k, x_{k+1}).$$

8. Изведете интерполационната формула на Лагранж с възли – нулите на полинома на Чебишов $T_n(x)$:

$$\xi_k := \cos\frac{(2k-1)\pi}{2n}, \quad k = 1, \ldots, n.$$

9. Нека $\omega(x) := (x - x_1)\cdots(x - x_n)$, $v(x) := (x - t_1)\cdots(x - t_{n-1})$ и нека точките $\{x_k\}_1^n$ и $\{t_j\}_1^{n-1}$ се преплитат, т.е.

$$x_1 < t_1 < x_2 < t_2 < \cdots < t_{n-1} < x_n.$$

Покажете, че $\omega'(x)v(x) - \omega(x)v'(x) \geq 0$ за всяко x.

10. Покажете, че при $f(x) = \cos x$ и възли $x_0 = 0$, $x_1 = \pi/2$, $x_2 = \pi$, грешката при интерполиране в интервала $[0, \pi]$ е по-малка от 0.7, т.е.

$$|f(x) - L_2(f; x)| \leq 0.7 \quad \text{за всяко } x \in [0, \pi].$$

11. Нека $2^{n-1}, a_1, \ldots, a_n$ са коефициентите на полинома на Чебишов $T_n(x)$, т.е.

$$T_n(x) = 2^{n-1}x^n + a_1x^{n-1} + \cdots a_{n-1}x + a_n.$$

Намерете a_{n-1}.

12. Покажете, че $T_n(x)$ е четна функция при четно n и нечетна, при нечетно n.

13. Докажете, че $T_n(T_m(x)) = T_{nm}(x)$ за всички цели неотрицателни числа m и n.

14. Нека ξ_j, $j = 1, \ldots, n$, са нулите на полинома на Чебишов $T_n(x)$. Покажете, че

$$T_n'(\xi_j) = (-1)^{j-1}\frac{n}{\sqrt{1 - \xi_j^2}}, \quad j = 1, \ldots, n.$$

15. Нека ξ_j, $j = 1, \ldots, n$, са нулите на полинома на Чебишов $T_n(x)$. Покажете, че $\xi_1 + \cdots + \xi_n = 0$.

16. Нека $P(x)$ е произволен полином, който удовлетворява условието

$$\max_{-1\le x\le 1} |P(x)| \le 1.$$

Докажете, че при всяко реално t, за което $|t| > 1$, имаме $|P(t)| \le |T_n(t)|$.

17. Покажете, че $|T_n'(x)| \le n^2$ за всяко $x \in [-1, 1]$.
Упътване: Покажете първо, че

$$T_n'(x) = n\frac{\sin n\theta}{\sin\theta}, \quad \text{където} \quad x = \cos\theta.$$

18. Докажете тъждеството

$$n^2(1 - T_n^2(x)) = (1 - x^2)[T_n'(x)]^2.$$

19. Докажете, че

$$T_n(x) = \frac{1}{2}\left\{\left(x + \sqrt{x^2-1}\right)^n + \left(x - \sqrt{x^2-1}\right)^n\right\}.$$

Упътване: Използвайте, че $\cos n\theta = \frac{1}{2}(e^{in\theta} + e^{-in\theta})$.

20. Като използвате интерполационната формула на Нютон с разделени разлики, намерете полинома $p(x)$, който интерполира функцията $f(x) = \sqrt{x} + 3x^2 - x + 2$ в точките $0, 1$ и 4.

21. Нека $f(x) = \frac{1}{x+1}$. Докажете, че

$$f[0, 1, 2, \ldots, n] = \frac{(-1)^n}{(n+1)!}.$$

22. Нека $f(t) = e^t$. Покажете, че $|f[0, 1, \ldots, n]| < 2^n$.

23. Нека $f(x) = \sin x$ и $0 < x_0 < x_1 < \cdots < x_{2n} < \pi$. Покажете, че $(-1)^n f[x_0, \ldots, x_{2n}] > 0$.

24. Нека $f(x) = x^n$. Докажете, че

$$f[x_1, \ldots, x_n] = x_1 + \cdots + x_n.$$

25. Нека $\omega(x) := (x - x_1)\cdots(x - x_n)$, $f(x) := (x - t_1)\cdots(x - t_n)$ и $x_1 < \cdots < x_n$, $t_1 < \cdots < t_n$. Докажете, че

$$\sum_{k=1}^{n} \frac{f(x_k)}{\omega'(x_k)} = \sum_{k=1}^{n}(x_k - t_k).$$

26. Нека $x_0, \ldots, x_n$ са различни точки и $\omega(x) := (x - x_0)\cdots(x - x_n)$. Да се докаже равенството

$$\ell'_{nk}(x_k) = \frac{\omega''(x_k)}{2\omega'(x_k)}.$$

27. Докажете, че

$$\sum_{k=1}^{n+1}(-1)^{n+1-k}\binom{n}{k-1}(2+k)^n = n!\,.$$

28. Представете разделената разлика $f[0,0,1,1]$ във вида

$$f[0,0,1,1] = A_0 f(0) + A_1 f'(0) + B_0 f(1) + B_1 f'(1).$$

Намерете коефициентите A_0, A_1, B_0, B_1.

29. Нека $x_0 < \cdots < x_n$ и $v(x) := (x - x_1)\cdots(x - x_n)$. Покажете, че

$$\frac{1}{v(x_0)} v[x_0, \ldots x_k] = \frac{1}{(x_0 - x_1)\cdots(x_0 - x_k)}.$$

30. Нека $F'(x) = f(x)$. Докажете, че

$$\sum_{k=0}^{n} F[x_0, \ldots, x_{k-1}, x_k, x_k, x_{k+1}, \ldots, x_n] = f[x_0, \ldots, x_n].$$

31. Нека $f(x)$ е произволна непрекъсната функция в $[0,1]$ и $F'(x) = f(x)$. Докажете, че при всеки избор на точките $x_0 < x_1 < \cdots < x_n$ в $[0,1]$,

$$F[x_0, \ldots, x_n, \xi] = \int_0^1 t^n f[x_0(t), \ldots, x_n(t)]\, dt,$$

където $x_k(t) := \xi + (x_k - \xi)t$, $k = 0, \ldots, n$.

Упътване: Покажете, че интегралът $I(f)$ в дясната страна е линейна комбинация от вида

$$I(f) = B \cdot F(\xi) + \sum_{k=0}^{n} A_k \cdot F(x_k)$$

която удовлетворява условията $I(x^m) = \delta_{m,n+1}$. Тези условия определят разделената разлика еднозначно.

32. Докажете, че разделената разлика на f с възли – екстремалните точки на полинома на Чебишов $T_n(x)$ в $[-1, 1]$:

$$\eta_j := \cos\frac{j\pi}{n}, \quad j = 0, \ldots, n,$$

се представя във вида

$$f[\eta_0, \ldots, \eta_n] = \frac{1}{n}\left[\frac{1}{2}(-1)^n f(-1) + \sum_{j=1}^{n-1}(-1)^{n-j} f(\eta_j) + \frac{1}{2} f(1)\right].$$

33. Нека $x_0 < \cdots < x_n$. Докажете, че при дадено ξ, $x_0 < \xi < x_n$, съществува положително число α – такова, че

$$f[x_0, \ldots, x_n] = \alpha f[x_0, \ldots, x_{n-1}, \xi] + (1-\alpha) f[x_1, \ldots, x_n, \xi].$$

34. Нека $x_0 < x_1 < \cdots < x_n$. Докажете, че при дадено m, $0 < m < n$, съществуват положителни числа $\alpha_1, \ldots, \alpha_{n-1}$ такива, че $\alpha_1 + \cdots + \alpha_{n-1} = 1$ и за всяка функция f, определена в тези точки, имаме

$$f[x_0, x_m, x_n] = \sum_{j=1}^{n-1} \alpha_j f[x_{j-1}, x_j, x_{j+1}].$$

35. Нека $f[x_0, \ldots, x_n] = 0$ при всеки избор на различните точки $\{x_k\}_{k=0}^{n}$. Да се докаже, че функцията f е алгебричен полином от степен $n-1$.

Упътване: Нека P_{n-1} е интерполационният полином за f с възли $\{x_k\}_{k=0}^{n-1}$. По формулата на Нютон, за всяко x различно от фиксираните възли,

$$f(x) - P_{n-1}(x) = f[x_0, \ldots, x_{n-1}, x](x - x_0) \cdots (x - x_{n-1}) = 0.$$

36. (Н. Обрешков). Нека $x_0 < x_1 < \cdots < x_n$ и f има непрекъснати производни до $(n+1)$-вата в $[x_0, x_n]$. Докажете, че

$$f[x_0, \ldots, x_n] = \frac{f^{(n)}(x_0)}{n!} + (x_1 + \cdots + x_n - n x_0)\frac{f^{(n+1)}(\xi)}{(n+1)!}$$

за някаква точка $\xi \in (x_0, x_n)$.

37. Нека $a_0, a_1, \ldots, a_{n+1}$ са произволни числа, които удовлетворяват условията $a_0 = a_{n+1} = 0$, $|a_{k+1} - 2a_k + a_{k-1}| \le 1$, $k = 1, \ldots, n$. Да се докаже, че

$$|a_k| \le k(n+1-k)/2, \quad k = 1, \ldots, n.$$

Упътване: Покажете, че в редиците $a_k - b_k$ и $a_k + b_k$, където $b_k = k(n+1-k)/2$, няма смяна на знака.

38. Нека f е непрекъсната функция в $[a, b]$, която удовлетворява условията $f(a) = f(b) = 0$ и

$$|f(x-h) - 2f(x) + f(x+h)| \le 1$$

за всяко x и $h > 0$ такива, че $x - h, x + h \in [a, b]$. Докажете, че $|f(x)| \le 1$ за всяко $x \in (a, b)$.

39. Нека $a < b$ и $f_0, g_0, f_1, g_1, \ldots, f_n, g_n$ са дадени числа. Да се докаже, че съществува единствен полином $P \in \pi_{2n+1}$, който удовлетворява интерполационните условия

$$P^{(2k)}(a) = f_k, \quad P^{(2k)}(b) = g_k, \quad k = 0, 1, \ldots, n.$$

40. Като използвате интерполационната формула на Нютон с разделени разлики, намерете полином $P \in \pi_4$, който удовлетворява интерполационните условия:

$$P(0) = P'(0) = 0, \quad P(1) = 1,\ P'(1) = 2, \quad P(2) = 4.$$

41. Нека

$$h_j(x) := \left(1 - \frac{\omega''(x_j)}{\omega'(x_j)}(x - x_j)\right) \ell_{nj}^2(x), \quad j = 0, \ldots, n.$$

Докажете, че $\sum_{j=0}^{n} h_j(x) = 1$.

42. Нека $x_1 < \cdots < x_n$. Да означим с $D(x_1, \ldots, x_n)$ детерминантата на системата

$$a_0 + a_1 x_k + a_2 x_k^2 + \cdots + a_{2n-1} x_k^{2n-1} = f_k$$
$$\{a_0 + a_1 x + a_2 x^2 + \cdots + a_{2n-1} x^{2n-1}\}'\,|_{x=x_k} = f_k'$$

за $k = 1, 2, \ldots, n$. Докажете, че

$$D(x_1, \ldots, x_n) = 2\prod_{i<j}(x_j - x_i)^4.$$

Упътване: Нека $\omega(x) := (x - x_1)\cdots(x - x_n)$ и

$$q(x) := \frac{\omega^2(x)}{x - x_n} = b_0 + b_1 x + \cdots + b_{2n-2}x^{2n-2} + x^{2n-1}.$$

Добавете към последния стълб на детерминантата D всички останали, умножени съответно с $b_0, b_1, \ldots, b_{2n-2}$. Тогава последния стълб ще добие вида $[0, \ldots, 0, 2\omega(x_n)q'(x_n)]^T$ и формулата следва по индукция.

43. Покажете, че полиномът

$$\begin{aligned} P(x) &= \sum_{j=0}^{n} A_j \frac{(x-b)^{m+1}(x-a)^j}{j!(a-b)^{m+1}} \sum_{i=0}^{n-j} \binom{m+i}{i} \left(\frac{x-a}{b-a}\right)^i \\ &+ \sum_{j=0}^{m} B_j \frac{(x-a)^{n+1}(x-b)^j}{j!(b-a)^{n+1}} \sum_{i=0}^{m-j} \binom{n+i}{i} \left(\frac{b-x}{b-a}\right)^i \end{aligned}$$

удовлетворява интерполационните условия

$$P^{(j)}(a) = A_j,\ j = 0, 1, \ldots, n; \qquad P^{(j)}(b) = B_j, \quad j = 0, 1, \ldots, m.$$

44. Докажете, че функциите $1, \cos x, \cos 2x, \ldots, \cos nx$ образуват система на Чебишов в интервала $[0, \pi]$.

45. Докажете, че функциите 1, $\sqrt{t}$, $t\sqrt{t}$, t^3 образуват система на Чебишов в интервала $[1, 2]$.

46. Нека $v(x)$ е непрекъсната функция и $v(x) > 0$ за всяко $x \in [0, 1]$. Да се докаже, че функциите

$$v(x),\ xv(x),\ x^2v(x),\ \ldots,\ x^n v(x)$$

образуват система на Чебишов в $[0, 1]$.

47. Нека $x_k = \frac{2k\pi}{2n+1}$, $k = 0, 1, \ldots, 2n$. Напишете формулата за тригонометричния полином $\lambda_j(x)$ от ред n, който удовлетворява интерполационните условия $\lambda_j(x_k) = \delta_{kj}$ за $k = 0, \ldots, 2n$.

48. Докажете, че

$$\sin x + \sin 2x + \cdots + \sin nx = \frac{-\cos(n + \frac{1}{2})x + \cos(x/2)}{2\sin(x/2)}.$$

49. Докажете, че ако матрицата $\mathbf{A} = \{a_{ij}\}$ има доминиращ главен диагонал, то $\det\mathbf{A} \neq 0$.

50. Докажете, че отсечените степенни функции

$$(x-x_0)_+^r,\ (x-x_1)_+^r,\ \ldots,(x-x_m)_+^r$$

са линейно независими в (x_m,∞) при $x_0 < x_1 \cdots < x_m$ и $m \leq r$.

51. Нека

$$s(x) := ax + b + \sum_{k=1}^{n} c_k(x-x_k)_+^{r-1}$$

е сплайн-функциата от първа степен с възли $x_1 < \cdots < x_n$, която удовлетворява интерполационните условия $s(x_k) = f(x_k)$, $k = 0,1,\ldots,n$. Намерете коефициентите $a, b, c_1, \ldots, c_n$ на сплайна.

52. Намерете естествения кубичен сплайн $s(x)$, който удовлетворява интерполационните условия: $s(-1) = 0$, $s(0) = 1$, $s(1) = 0$.

53. Нека функцията f има непрекъснати производни до $2n$-тата включително в $[a,b]$ и P е полином от степен $2n-1$, който удовлетворява интерполационните условия

$$P^{(k)}(a) = f^{(k)}(a), \quad P^{(k)}(b) = f^{(k)}(b) \ \text{ за } \ k = 0,1,\ldots,n-1.$$

Докажете, че

$$\int_a^b [P^{(n)}(x)]^2\,dx \leq \int_a^b [f^{(n)}(x)]^2\,dx.$$

Упътване: Покажете първо, че

$$\int_a^b [f^{(n)}(x) - P^{(n)}(x)]P^{(n)}(x)\,dx = 0.$$

54. Изчислете $B(0,1,2,3;t)$ за $t=1$.

55. Нека $x_0 < x_1 < \cdots < x_r$. Докажете, че $B(x_0,\ldots,x_r;t) = 0$ за $t < x_0$.

56. Знаем, че $B(1,2,3,4;t)$ съвпада с полином $p(t)$ от степен 2 при $2 < t < 3$. Намерете този полином.

57. Нека $a \leq x_0 < \cdots < x_r \leq b$. Докажете, че

$$\int_a^b B(x_0,\ldots,x_r;t)\,dt = \frac{1}{r}.$$

Упътване: Използвайте, че

$$(x-x_0)^r = r\int_{x_0}^{x}(x-t)^{r-1}\,dt.$$

58. Нека $a \le x_0 < \cdots < x_r \le b$. Докажете, че

$$\int_a^b tB(x_0,\ldots,x_r;t)\,dt = \frac{1}{r(r+1)}(x_0+\cdots+x_r).$$

59. Докажете връзката

$$B(x_0,\ldots,x_r;t) = \frac{t-x_r}{r-1}B'(x_0,\ldots,x_r;t) + B(x_0,\ldots,x_{r-1};t).$$

Упътване: По правилото на Лайбниц-Попович,

$$\begin{aligned} B(x_0,\ldots,x_r;t) &= (x-t)_+^{r-1}[x_0,\ldots,x_r] = \left((x-t)(x-t)_+^{r-2}\right)[x_0,\ldots,x_r] \\ &= (x_r-t)(x-t)_+^{r-2}[x_0,\ldots,x_r] + (x-t)_+^{r-2}[x_0,\ldots,x_{r-1}]. \end{aligned}$$

60. Докажете, че $\|\mathbf{x}\| = \max_{1\le k\le n}|x_k|$ е норма.

61. Докажете, че $\|p\| := \max_{0\le k\le n}|a_k|$ определя норма в пространството

$$\pi_n := \left\{ p(x) = \sum_{k=0}^{n} a_k x^k \;:\; (a_0,\ldots,a_n)\in\mathbb{R}^{n+1} \right\}.$$

62. Нека X е произволно линейно нормирано пространство. Докажете, че единичното кълбо $\mathbf{B} := \{\mathbf{x} \,:\, \|\mathbf{x}\| \le 1\}$ е изпъкнало множество.

63. Нека $x_0 < x_1 < \cdots < x_n$ са произволни фиксирани точки. Докажете, че пространството π_n от алгебрични полиноми от степен n може да се нормира по следния начин:

$$\|p\| = \max_{0\le k\le n}|p(x_k)|.$$

64. Докажете, че ако функцията $f(x)$ е четна, т.е. $f(x) = f(-x)$ за $x\in[-1,1]$, то и полиномът $P(x)\in\pi_n$ на най-добро равномерно приближение на f в $[-1,1]$ е четен.

65. Намерете полинома p от степен 4 на най-добро равномерно приближение в $[0,2\pi]$ за функцията $f(x) = \sin 3x$.

66. Намерете полинома от степен 2 на най-добро равномерно приближение в $[-1,1]$ за функцията $f(x) = 2^3x^4$.

67. Нека $x_0 < \cdots < x_{n+1}$ са точки на алтернанс за непрекъсната в $[a, b]$ функция f. Покажете, че за най-доброто приближение $E_n(f)$ на f с алгебрични полиноми от степен n имаме:

$$E_n(f) = \left| \frac{f[x_0, \ldots, x_{n+1}]}{s[x_0, \ldots, x_{n+1}]} \right|,$$

където s е произволна функция, удовлетворяваща условието $s(x_k) = (-1)^k$, $k = 0, \ldots, n+1$.

68. (Н. Обрешков). Нека функциите f и g имат непрекъснати производни до $(n+1)$ -вата включително в $[a, b]$ и $a \le x_0 < \cdots < x_{n+1} \le b$ са произволни точки. Тогава съществува точка $\xi \in (a, b)$ такава, че

$$\frac{f[x_0, \ldots, x_{n+1}]}{g[x_0, \ldots, x_{n+1}]} = \frac{f^{(n+1)}(\xi)}{g^{(n+1)}(\xi)}.$$

69. (С. Н. Бернщайн). Нека функциите f и g имат непрекъснати производни до $(n+1)$-вата включително в $[a, b]$ и удовлетворяват там неравенството $0 \le f^{(n+1)}(x) \le g^{(n+1)}(x)$. Докажете, че тогава за най-добрите им приближения в $[a, b]$ имаме $E_n(f) \le E_n(g)$.

70. Покажете директно, че

$$\max_{0 \le x \le 1} |f(x) - B_n(f; x)| \quad \to \quad 0 \text{ при } n \to \infty$$

при $f(t) = t^3$.

Упътване: За да намерите $B_n(f; x)$ приложете 3 пъти оператора $\frac{p}{n}\frac{d}{dp}$ към тъждеството

$$(p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

и слажете $p = x$, $q = 1 - x$ в полученото равенство.

71. Нека $f(x) = 0$ за $0 \le x \le 1/2$ и $f(x) = x - 1/2$ за $1/2 \le x \le 1$. Напишете $B_4(f; x)$ и покажете, че $B_4(f; x) \ge f(x)$ в $[0, 1]$.

72. Докажете, че коефициентът пред x^n в полинома на Бернщайн $B_n(f; x)$ е равен на нула за всяко $f \in \pi_{n-1}$.

73. Покажете, че $B_n(e^t; x) = [1 + (e^{1/n} - 1)x]^n$. Оттук докажете директно, че $B_n(e^t; x)$ клони равномерно в $[0, 1]$ към e^x при $n \to \infty$.

74. Нека $L_r(t) = \alpha_r t^r + \cdots$, $\alpha_r > 0$, е полиномът на Льожандър (т.е. полиномът от степен r с водещ коефициент α_r, който е ортогонален в $[-1,1]$ с тегло 1 на всеки полином от степен $r-1$). Докажете, че функцията

$$f(x) := \int_{-1}^{1} (x-t)_+^{(r-1)} L_r(t)\, dt$$

не си сменя знака в $(-1,1)$ и намерете този знак.

75. Докажете, че системата от полиноми $1, x, x^2, \ldots$ не може да бъде ортогонална система в $[-1,1]$ при никакво тегло $\mu(x)$.

76. Като използвате, че полиномът на Чебишов $T_n(x)$ е ортогонален в $[-1,1]$ с тегло $1/\sqrt{1-x^2}$ на всеки полином от степен $\le n-1$, докажете, че

$$\int_{-1}^{1} \frac{T_n^3(x)}{\sqrt{1-x^2}} Q(x)\, dx = 0$$

за всеки полином Q от степен $\le n-1$.

77. Нека $U_n(x) = T'_{n+1}(x)$. Докажете, че

$$\int_{-1}^{1} \sqrt{1-x^2}\, U_n(x) Q(x)\, dx = 0$$

за всеки полином Q от π_{n-1}.

78. (Формула на Кристофел–Дарбу). Нека $\{P_n\}$ е редица от ортогонални полиноми – такива, че

$$P_n(x) = \alpha_n x^n + \text{полином от } \pi_{n-1}, \quad \alpha_n > 0, \quad (P_n, P_n) = 1.$$

Тогава

$$\sum_{k=0}^{n} P_k(x) P_k(y) = \frac{\alpha_n}{\alpha_{n+1}} \frac{P_{n+1}(x)P_n(y) - P_{n+1}(y)P_n(x)}{x-y}.$$

79. Нека в линейното пространство H е определено скаларно произведение (x,y) и $x_1, \ldots, x_n$ са дадени елементи от H. Намерете елемента x от вида $x = a_1 x_1 + \cdots + a_n x_n$, където $\{a_k\}$ са реални числа, който минимизира израза

$$\|x - x_1\|^2 + \cdots + \|x - x_n\|^2.$$

80. Нека P_n е полиномът на Льожандър от степен n, удовлетворяващ условието $P_n(1) = 1$. Пресметнете интеграла $\int_{-1}^{1} P_n^2(x)\, dx$.

81. Нека функцията $f(x)$ е непрекъсната в $[-1,1]$ и $p(x)$ е полиномът на най-добро средноквадратично приближение за f от степен n в $[-1,1]$ с тегло 1. Докажете, че разликата $f(x)-p(x)$ има поне една смяна на знака в $(-1,1)$ или $f=p$.

82. Намерете приближение на таблицата

x_i :	0	1	2	3
y_i :	2	4	2	1

с полином p от първа степен по метода на най-малките квадрати.

83. Нека $x_k = x_0 + kh$, $k = 0,1,2$. Намерете приближение на $f'(x_0)$ от вида

$$f'(x_0) \approx \alpha f(x_0) + \beta f(x_1) + \gamma f(x_2)$$

по метода на неопределените коефициенти.

84. Нека $f''(x) \geq 0$ в $[a,b]$. Докажете, че

$$f(\frac{a+b}{2})(b-a) \leq \int_a^b f(x)\,dx \leq \frac{f(a)+f(b)}{2}(b-a).$$

85. Пресметнете $\int_0^1 \frac{1}{1+x}\,dx$ по квадратурната формула на Симпсън. Докажете, че грешката е по-малка от 0.01.

86. Нека $\xi_1,\ldots,\xi_n$ са нулите на полинома на Чебишов $T_n(x)$. Покажете, че квадратурната формула

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\,dx \approx \frac{\pi}{n}\sum_{k=1}^{n} f(\xi_k)$$

има АСТ $= 2n-1$.

87. Нека $-1 = \eta_0 < \eta_1 < \cdots < \eta_n = 1$ са екстремалните точки на полинома на Чебишов $T_n(x)$ в $[-1,1]$. Покажете, че квадратурната формула

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} f(x)\,dx \approx \frac{\pi}{n}\left(\frac{1}{2}f(-1) + \sum_{k=1}^{n-1} f(\eta_k) + \frac{1}{2}f(1)\right)$$

има АСТ $= 2n-1$.

88. Намерете коефициентите B_1 и B_2 в квадратурната формула на Лобато в $[-1,1]$ с тегло $\mu(x) \equiv 1$.

Отговор: $B_1 = B_2 = \frac{2}{(n+1)(n+2)}$

89. Нека $P_n(x) = \alpha_n x^n + \cdots$ е полином, ортогонален в $[-1,1]$ с тегло $\mu(x)$ на всички полиноми от степен $< n$ и такъв, че $\int_{-1}^{1} \mu(x) P_n^2(x)\,dx = 1$. Нека

$$\int_{-1}^{1} \mu(x) f(x)\,dx = \sum_{k=0}^{n} b_k f(t_k)$$

е квадратурната формула на Лобато. Тогава за всеки полином $f \in \pi_n$ е изпълнено равенството

$$\sum_{k=0}^{n} b_k P_n(t_k) f(x_k) = C \cdot f[t_0, \ldots, t_n],$$

където

$$C = \frac{1}{\alpha_n}\left[1 + \alpha_n^2 \int_{-1}^{1} \mu(x)(1-x^2)(x-t_1)^2 \cdots (x-t_{n-1})^2\,dx\right].$$

90. Нека квадратурната формула

$$\int_a^b \mu(x) f(x)\,dx \approx \sum_{k=1}^{n} a_k f(x_k)$$

има алгебрическа степен на точност $2n-1$. Дали съществува поне един полином от степен точно $2n$, за който тя да е точна ?

91. Нека квадратурната формула

$$\int_a^b \mu(x) f(x)\,dx \approx B f(a) + \sum_{k=1}^{n} a_k f(x_k)$$

има алгебрическа степен на точност $2n$. Докажете, че $B > 0$.

92. Нека

$$\int_a^b \mu(x) f(x)\,dx \approx \sum_{k=1}^{n} a_k f(x_k)$$

е квадратурната формула на Гаус и $\omega(x) := (x-x_1)\cdots(x-x_n)$. Докажете, че полиномите $\omega_k(x) := \omega(x)/(x-x_k)$, $k = 1, \ldots, n$, са ортогонални един на друг в $[a,b]$ с тегло $\mu(x)$.

93. Докажете, че

$$\int_{-1}^{1} f(x)\,dx \approx \sum_{k=1}^{n} a_k f(x_k)$$

е квадратурната формула на Гаус тогава и само тогава, когато

$$x_k^n \int_{-1}^{1} \omega_k(x)\,dx = \int_{-1}^{1} x^n \omega_k(x)\,dx, \quad \forall\, k = 1, \dots, n.$$

Тук $\omega(x) := (x - x_1)\cdots(x - x_n)$, $\omega_k(x) := \omega(x)/(x - x_k)$.

94. Нека

$$\int_{-1}^{1} f(x)\,dx \approx \sum_{k=1}^{n} a_k f(x_k) =: Q_n(f)$$

е квадратурната формула на Гаус. Докажете, че за всяка непрекъсната в $[-1,1]$ функция имаме

$$\lim_{n\to\infty} Q_n(f) = \int_{-1}^{1} f(x)\,dx.$$

95. (Л. Чакалов). Нека

$$\int_a^b \mu(x) f(x)\,dx \approx \sum_{k=1}^{n} a_k f(x_k)$$

е квадратурната формула на Гаус. Ако $P \in \pi_{2n-1}$ и $\int_a^b \mu(x)P(x)\,dx = 0$, докажете че P има поне една нула в интервала (x_1, x_n).

96. Нека $-1 = t_0 < t_1 < \cdots < t_n = 1$ и квадратурната формула

$$\int_{-1}^{1} f(x)\,dx \approx \sum_{j=0}^{n} b_j f(t_j)$$

има АСТ $= 2n - 1$. Нека

$$\int_a^b f(x)\,dx \approx \sum_{k=1}^{n} a_k f(x_k)$$

е квадратурната формула на Гаус. Докажете, че

$$\int_{-1}^{1} (1 - x^2) \prod_{j=1}^{n-1} (x - t_j)^2\,dx > \int_{-1}^{1} \prod_{k=1}^{n} (x - x_k)^2\,dx.$$

97. При условията от предишната задача покажете, че

$$\sum_{j=0}^{n} b_j f(t_j) - \sum_{k=1}^{n} a_k f(x_k)$$

съвпада с разделената разлика на f в точките $x_1, \ldots, x_n, t_0, t_1, \ldots, t_n$ с точност до постоянен множител. Докажете, че точките $\{x_k\}$ и $\{t_j\}$ се преплитат, т.е. $t_0 < x_1 < t_1 < x_1 < \cdots < x_n < t_n$.

98. Докажете, че равенството

$$\|A\| := n \max_{ij} |a_{ij}|,$$

където максимумът е по всички $i, j \ : \ 1 \leq i \leq n, \ 1 \leq j \leq n$, определя матрична норма в пространството от всички матрици $n \times n$.

99. Нека E е единичната матрица. Докажете, че $\|E\| \geq 1$ за всяка матрична норма.

100. Покажете, че за всяка матрична норма $\|\cdot\|$ може да се намери векторна норма $\mu(\cdot)$, която е съгласувана с матричната.

Упътване: Например, $\mu(\mathbf{x}) := \|\mathbf{X}\|$, където матрицата $\mathbf{X}$ има за първи стълб елементите на $\mathbf{x}$, а останалите стълбове на $\mathbf{X}$ са нулеви.

101. Докажете, че уравнението $x^3+x-27=0$ има единствен положителен корен ξ. Намерете интервал $[a, b]$, който съдържа ξ и постройте подходящ итерационен процес $x_{m+1} = \varphi(x_m)$ (т.е. намерете $\varphi(x)$) така, че процесът да е сходящ със скорост на геометрична прогресия при всяко начално приближение x_0 от $[a, b]$.

102. Постройте итерационен метод за пресмятане на $\sqrt{5}$ чрез решаване на уравнението $x^2 - 5 = 0$.

103. Нека $\mathbf{x}_0 := (1, 1, -1)$. Намерете първото приближение

$$\mathbf{x}_1 = (x_1^{(1)}, x_2^{(1)}, x_3^{(1)})$$

на решението на системата

$$\left| \begin{array}{lllllll} x_1 & + & x_2 & + & x_3 & = 1 \\ x_1 & - & x_2 & & & = 0 \\ 2x_1 & + & x_2 & + & x_3 & = 2 \end{array} \right.$$

по метода на простата итерация.

СЪДЪРЖАНИЕ

Глава 1

ПРИБЛИЖАВАНЕ НА ФУНКЦИИ

Глава 2

ЧИСЛЕНО ДИФЕРЕНЦИРАНЕ И ИНТЕГРИРАНЕ

Глава 3

ЧИСЛЕНО РЕШАВАНЕ НА УРАВНЕНИЯ

Глава 4

РЕШАВАНЕ НА СИСТЕМИ ОТ УРАВНЕНИЯ

Глава 5

ПРЕСМЯТАНЕ НА СОБСТВЕНИ СТОЙНОСТИ НА МАТРИЦИ

Глава 6