Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет им. Александра Григорьевича и Николая Григорьевича Столетовых»

ЧИСЛЕННЫЕ МЕТОДЫ. ЛЕКЦИИ И ПРИМЕРЫ

ВЛАДИМИР 2016

1 Многочлены

1.1 Основные понятия и теоремы

1. Алгебраическим многочленом порядка (степени) n называется функция вида

$$P_n(z) = p_n z^n + p_{n-1} z^{n-1} + \ldots + p_0, \tag{1.1}$$

где $p_n \neq 0$. Иногда индекс n там, где степень многочлена неважна, будем опускать и писать $P(z)$. Числа p_k называются коэффициентами многочлена $P_n(z)$. Алгебраическим уравнением порядка n называется уравнение вида $P_n(z) = 0$. Число z_0 называется корнем кратности $k \geq 1$ уравнения $P_n(z) = 0$ (или, что то же самое, многочлена (1.1)), если

$$P_n(z_0) = 0, \quad P_n'(z_0) = 0, \quad \ldots \quad , P^{(k-1)}(z_0) = 0, \quad P^{(k)}(z_0) \neq 0.$$

В случае $k = 1$ корень z_0 называется простым.

Спарведлива

Основная теорема алгебры. *Уравнение $P_n(z) = 0$ при $n \geq 1$ имеет ровно n комплексных корней с учетом их кратности (то есть каждый корень учитывается столько раз, какова его кратность). Другими словами, сумма кратностей кореней уравнение $P_n(z) = 0$ равна n.*

Следствие основной теоремы алгебры. *Каждый многочлен вида* (1.1) *можно представить в виде произведения простейших множителей*

$$P_n(z) = p_n(z - z_1) \cdot (z - z_2) \cdots (z - z_n). \tag{1.2}$$

В этом произведении некоторые множители, соответствующие кратным корням, могут повторяться. Поэтому разложение на множители (1.2) можно переписать в виде

$$P_n(z) = p_n(z - a_1)^{k_1} \cdot (z - a_2)^{k_2} \cdots (z - a_m)^{k_m}, \tag{1.3}$$

где $a_1, a_2, \ldots, a_m$ — различные корни уравнения $P_n(z) = 0$, а $k_1, k_2, \ldots, k_m$ — кратности этих корней.

Хорошо известна следующая

Теорема Виета. *Коэффициенты многочлена* (1.1) *удовлетворяют следующим равенствам*

$$\frac{p_{n-k}}{p_n} = (-1)^k \sigma_k,$$

где σ_k — элементарные симметрические многочлены порядка k от корней многочлена (1.1)*, которые определяются по формуле*

$$\sigma_k = \sigma_k(z_1, \ldots, z_n) = \sum\nolimits_{1 \leq j_1 < j_2 < \ldots < j_k \leq n} z_{j_1} z_{j_2} \ldots z_{j_k}, \quad k = 1, \cdots, n.$$

2. Свойства многочленов с вещественными коэффициентами. Пусть все коэффициенты многочлена $P(z)$ вещественны. Тогда если $z_0 = x_0 + iy_0$ — корень многочлена (1.1), то либо $y_0 = 0$ (то есть корень вещественный), либо $y_0 \neq 0$ и тогда корнем является и сопряженное с этим корнем число $\overline{z_0} = x_0 - iy_0$.

Это сразу следует из равенства

$$0 = \overline{p_n z_0^n + p_{n-1} z_0^{n-1} + \ldots + p_0} = p_n \overline{z_0}^{\,n} + p_{n-1} \overline{z_0}^{\,n-1} + \ldots + p_0$$

Если n нечетно, то многочлен (1.1) с вещественными коэффициентами всегда имеет хотя бы один вещественный корень.

Это сразу следует из того, что при достаточно больших a знаки величин $P(a)$, $P(-a)$ определяются старшим слагаемым $p_n a^n$, и, следовательно, противоположны. Но непрерывная функция, принимающая на концах отрезка $[-a, a]$ значения с противоположными знаками, имеет на этом отрезке хотя бы один нуль.

1.2 Схема Горнера вычисления значения многочлена. Применение к оценке границы вещественных корней

Вычислим значения многочлена (1.1) в точке $z = c$ по следующей схеме:

$$P_n(c) = p_n c^n + p_{n-1}c^{n-1} + p_{n-2}c^{n-2} + \ldots + p_1 c + p_0 =$$

$$= (p_n c + p_{n-1})c^{n-1} + p_{n-2}c^{n-2} + \ldots + p_1 c + p_0 =$$

$$= [(p_n c + p_{n-1})c + p_{n-2}]c^{n-2} + \ldots + p_1 c + p_0 =$$

$$= (\ldots\{[(p_n c + p_{n-1})c + p_{n-2}]c + p_{n-3}\} + \ldots +)c + p_0.$$

Этот процесс можно записать в виде последовательности операций

$$\begin{array}{rcl} g_n & = & p_n; \\ g_{n-1} & = & g_n c + p_{n-1}; \\ g_{n-2} & = & g_{n-1}c + p_{n-2}; \\ \ldots & \ldots & \ldots \\ g_1 & = & g_2 c + p_1; \\ g_0 & = & g_1 c + p_0. \end{array} \tag{1.4}$$

Числа g_k будем называть коэффициентами Горнера. Последний коэффициент g_0 в этой цепочке равенств равен значению $P_n(c)$.

Пример. Вычислить значение многочлена $P_4(x) = 2x^4 + x^3 - 2x + 3$ по схеме Горнера при $x = c = -2$.

$$\begin{array}{rcl} g_4 & = & 2; \\ g_3 & = & 2c + 1 = -3; \\ g_2 & = & g_3 c + 0 = 6; \\ g_1 & = & g_2 c - 2 = -14; \\ g_0 & = & g_1 c + 3 = 31. \end{array}$$

Указанная схема имеет и другие применения. Справедлива

Теорема. *Если при некотором положительном c все коэффициенты Горнера g_k, $k = 0, \ldots, n$, вещественны и неотрицательны, $p_n = g_n > 0$, то действительные корни не превосходят этого числа c.*

Доказательство. При делении многочлена P_n на $x - c$ получается равенство

$$P_n(x) = (g_n x^{n-1} + g_{n-1}x^{n-2} + \ldots + g_2 x + g_1) \cdot (x - c) + g_0. \tag{1.5}$$

Перемножив выражения справа и собрав коэффициенты при степенях x, получим

$$(g_n x^{n-1} + g_{n-1}x^{n-2} + \ldots + g_2 x + g_1) \cdot (x - c) + g_0 =$$

$$(g_0 - c\,g_1) + (g_1 - c\,g_2)x + (g_2 - c\,g_3)x^2 + \ldots + (g_{n-1} - c\,g_n)x^{n-1} + g_n x^n.$$

Для того, чтобы полученное выражение равнялось P_n необходимо и достаточно равенство соответствующих коэффициентов:

$$g_n = p_n, \quad g_{n-1} - c\,g_n = p_{n-1}, \ldots, g_1 - c\,g_2 = p_1, \quad g_0 - c\,g_1 = p_0.$$

Это совпадает с формулами (1.4).

Пусть все $g_k \geq 0$. Поэтому при $x > c$ оба слагаемых в данном представлении ≥ 0, причем $g_n = p_n > 0$, и, значит, из (1.5) получается $P(x) > 0$ (т.е. не может равняться нулю). Это и доказывает теорему.

Теорему можно применять для оценки границы действительных корней многочлена $P_n(x)$ (считаем $p_n > 0$). Именно, сначала подберем $c > 0$ при котором все $g_k \geq 0$. Тогда все корни $x_k \leq c$. Далее, рассмотрим многочлен $Q(x) = (-1)^n P(-x)$. Этот многочлен имеет положительный старший коэффициент p_n. Снова подберем $\tilde{c} > 0$ при котором все $g_k \geq 0$. Тогда все корни этого многочлена будут удовлетворять неравенству $\tilde{x}_k \leq \tilde{c}$. Но эти корни связаны с корнями многочлена P так: $\tilde{x}_k = -x_k$. Поэтому $-x_k \leq \tilde{c}$, то есть $x_k \geq -\tilde{c}$. В результате получаем оценку границы корней: $-\tilde{c} \leq x_k \leq c$.

Пример. Оценить границу корней x_k многочлена $P(x) = x^4 - 3\,x^3 - 9\,x^2 - 3x + 1$ с помошью метода Горнера.

При $c = 5$ получаем $g_4 = 1$, $g_3 = 2$, $g_2 = 1$, $g_1 = 2$, $g_0 = 11$. Следовательно, все корни меньше 5.

Для оценки снизу берем многочлен

$$Q(x) = (-1)^4 P(-x) = x^4 + 3x^3 - 9x^2 + 3x + 1.$$

При $c = 2$ получаем $g_4 = 1$, $g_3 = 5$, $g_2 = 1$, $g_1 = 5$, $g_0 = 11$. Следовательно, все корни многочлена Q меньше 2, а корни многочлена P больше -2.

Окончательно, $-2 < x_k < 5$.

1.3 Алгоритм Евклида. Наибольший общий делитель двух многочленов.

1. Алгоритм Евклида состоит в следующей процедуре делений с остатком:

$$P(x) = Q(x) \cdot C_1(x) + r_1(x), \tag{1.6}$$

$$Q(x) = r_1(x) \cdot C_2(x) + r_2(x), \tag{1.7}$$

$$r_1(x) = r_2(x) \cdot C_3(x) + r_3(x), \tag{1.8}$$

и так далее. Здесь C_j — целые части, r_j — остатки при делении.

Примечание. Для единообразия положим $P(x) = r_{-1}(x)$, $Q(x) = r_0(x)$. Тогда предыдущие формулы можно переписать в виде

$$r_{m-2}(x) = r_{m-1}(x) \cdot C_m(x) + r_m(x), \qquad m = 1, 2, \ldots \tag{1.9}$$

Хорошо известно, что степень остатка меньше степени соответствующего делителя:

$$\deg Q > \deg r_1 > \deg r_2 > \ldots,$$

поэтому процесс прервется на каком-то конечном шаге $k + 1$, т.е. мы придем к делению без остатка:

$$r_{k-2}(x) = r_{k-1}(x) \cdot C_k(x) + r_k(x), \tag{1.10}$$

$$r_{k-1}(x) = r_k(x) \cdot C_{k+1}(x) + 0. \tag{1.11}$$

Утверждается, что многочлен $r_k(x)$ является наибольшим общим делителем многочленов P и Q, кратко, $r_k = \text{НОД}\,(P, Q)$.

Покажем сначала, что r_k является общим делителем. Из (1.11) следует, что $r_{k-1}(x)$ делится на r_k. Отсюда и из (1.10) следует, что $r_{k-2}(x)$ делится на r_k. И так далее. В результате получим, что $P(x)$ и $Q(x)$ делятся на r_k.

Покажем теперь, что r_k является *наибольшим* из делителей, т.е. делит любой другой делитель. Действительно, если некоторый многочлен $S(x)$ делит одновременно $P(x)$ и $Q(x)$, то он делит и $r_1(x)$ (см. (1.6)). Значит, он делит одновременно $Q(x)$ и $r_1(x)$ и, следовательно, делит $r_2(x)$ (см. (1.7)). Отсюда, аналогично, получается, что он делит $r_3(x)$ (см. (1.8)). И так далее. В результате получаем , что он делит $r_{k-2}(x)$, $r_{k-1}(x)$, $r_k(x)$ (см. (1.10)).

В частности, если r_k является *числом*, т.е. многочленом нулевой степени, то многочлены P и Q взаимно просты.

Примечание. Равенства (1.6)–(1.11) можно домножать на постоянные множители. Поэтому для упрощения счета, например, чтобы иметь дело с целыми коэффициентами, при делении с r_{m-2} на r_{m-1} можно предварительно домножить эти многочлены на некоторые числовые множители M и N и делить $Mr_{m-2}(x)$ на $Nr_{m-1}(x)$. Действительно,

$$r_{m-2}(x) = r_{m-1}(x) \cdot C_m(x) + r_m(x) \quad \Leftrightarrow$$

$$Nr_{m-2}(x) = Nr_{m-1}(x) \cdot C_m(x) + Nr_m(x) \quad \Leftrightarrow$$

$$Nr_{m-2}(x) = Mr_{m-1}(x) \cdot \frac{N}{M}\, C_m(x) + Nr_m(x),$$

и числа $N \neq 0$ и $M \neq 0$ могут быть любыми.

Пример. $P = x^4 + x^2 + x - 3$, $Q = x^3 + x^2 + x - 3$.

$$x^4 + x^2 + x - 3 = (x^3 + x^2 + x - 3) \cdot (x - 1) + (x^2 + 5x - 6);$$
$$x^3 + x^2 + x - 3 = (x^2 + 5x - 6) \cdot (x - 4) + (27x - 27);$$
$$x^2 + 5x - 6 = (x - 1) \cdot (x + 6) + 0.$$

Ответ: $\text{НОД}\,(P, Q) = x - 1$.

2. Устранение кратности корней с помощью алгоритма Евклида. Многочлен (1.1) может иметь кратные корни, что при численном решении уравнения $P_n(z) = 0$ может приводить к большим погрешностям. Существует простой метод получения уравнения

$$Q(z) := q_m z^m + q_{m-1} z^{m-1} + \ldots + q_0 = 0, \tag{1.12}$$

которое имеет те же корни, что и многочлен (1.1), но все они простые. Запишем многочлен P в виде

$$P(z) = p_n (z - a_1)^{k_1} (z - a_2)^{k_2} \ldots (z - a_m)^{k_m},$$

где k_s — кратности соответствующих корней a_s. Поскольку

$$P'(z) = A(z - a_1)^{k_1 - 1} (z - a_2)^{k_2 - 1} \ldots (z - a_m)^{k_m - 1} R(z),$$

где $R(z)$ — многочлен, среди корней которого нет корней многочлена $P(z)$, то наибольшим общим делителем $S(z) = \text{НОД}\,(P(z), P'(z))$ является многочлен вида

$$S(z) = (z - a_1)^{k_1 - 1} (z - a_2)^{k_2 - 1} \ldots (z - a_m)^{k_m - 1}.$$

Значит, многочлен $Q(z) = \frac{P(z)}{S(z)}$ будет иметь те же корни, что и P, причем все эти корни простые.

Для нахождения $S(z) = \text{НОД}\,(P(z), P'(z))$ применяется алгоритм Евклида.

Пример. $P(z) = z(z-1)^2(z+1)^2 = z^5 - 2z^3 + z$. Тогда $P'(z) = 5z^4 - 6z^2 + 1$. Алгоритм Евклида: первый шаг

$$P(z) = \frac{1}{5}z \cdot P'(z) - \frac{4}{5}z^3 + \frac{4}{5}z.$$

Домножим $r_1(z) = -\frac{4}{5}z^3 + \frac{4}{5}z$ на $-5/4$, получим новый остаток, для которого сохраним то же обозначение: $r_1 = z^3 - z$.

Второй шаг

$$P'(z) = 5z \cdot r_1 + 1 - z^2.$$

Получен остаток: $r_2 = 1 - z^2$. Домножим его на -1, получим новый остаток, для которого сохраним то же обозначение: $r_2 = z^2 - 1$.

Третий шаг

$$r_1(z) = z \cdot r_2 + 0.$$

Получен остаток: $r_3 = 0$. Значит $\text{НОД}\,(P, P') = r_2 = z^2 - 1$. Наконец,

$$\frac{P(z)}{z^2-1} = (z^2-1)z$$

есть искомый многочлен без кратных корней.

1.4 Оценка наибольшего корня методом квадрирования корней (методом Лобачевского)

Метод состоит в следующем. Пусть для простоты $p_n = 1$,

$$P(z) = (z - z_1)(z - z_2)\cdots(z - z_n).$$

Тогда

$$P(-z) = (-z - z_1)(-z - z_2)\cdots(-z - z_n) = (-1)^n(z + z_1)(z + z_2)\cdots(z + z_n),$$

$$P_1(z^2) := P(z)P(-z) = (-1)^n(z^2 - z_1^2)(z^2 - z_2^2)\cdots(z^2 - z_n^2).$$

Сделаем замену $z^2 = t$ и получим

$$P_1(t) = (-1)^n(t - z_1^2)(t - z_2^2)\cdots(t - z_n^2).$$

Эта процедура называется квадрированием. Повторив квадрирование, получим

$$P_2(t) = (-1)^n(t - z_1^4)(t - z_2^4)\cdots(t - z_n^4), \quad t = z^4,$$

$$P_3(t) = (-1)^n(t - z_1^8)(t - z_2^8)\cdots(t - z_n^8), \quad t = z^8,$$

и, вообще, после k квадрирований получим

$$P_k(t) = (-1)^n(t - z_1^{2^k})(t - z_2^{2^k})\cdots(t - z_n^{2^k}), \quad t = z^{2^k}.$$

Если нам удалось вычислить приближенно некоторый корень t_1 многочлена $P_k(t)$, то найдем и соответствующий корень исходного многочлена $P(z)$ по формуле

$$z_1 = t_1^{2^{-k}} = \sqrt[2^k]{t_1}.$$

Идея приближенного вычисления корня t_1 многочлена $P_k(t)$ состоит в следующем. Предположим, что модуль первого корня z_1 исходного многочлена $P(z)$ по модулю строго больше модулей других корней, то есть $|z_s| < \varepsilon |z_1|$, $\varepsilon \in (0,1)$, $s = 2, \ldots, n$. Тогда модуль соответствующего корня t_1 квадрированного k раз многочлена $P_k(t)$ *намного* больше:

$$|t_s| = \left| z_s^{2^k} \right| < \varepsilon^{2^k} \left| z_1^{2^k} \right| = \varepsilon^{2^k} |t_1|.$$

Теперь по теореме Виета имеем

$$(-1)^n P_k(t) = t^n + a_{n-1} t^{n-1} + \ldots + a_0,$$

где

$$a_{n-1} = -t_1 - t_2 - \ldots - t_n, \quad a_{n-1} = -(z_1^{2^k} + z_2^{2^k} + \ldots + z_n^{2^k}).$$

Отсюда можно найти приближенно наибольший по модулю корень z_1. Действительно,

$$\sqrt[2^k]{-a_{n-1}} = \sqrt[2^k]{t_1 - E} = \sqrt[2^k]{t_1} \left(1 - \frac{E}{t_1}\right)^{2^{-k}} =$$

$$= \sqrt[2^k]{t_1} + \sqrt[2^k]{t_1} \left(\left(1 - \frac{E}{t_1}\right)^{2^{-k}} - 1 \right) = z_1 + z_1 \left(\left(1 - \frac{E}{t_1}\right)^{2^{-k}} - 1 \right), \tag{1.13}$$

где $E = -t_2 - t_3 - \ldots - t_n$. Оценим модуль последнего слагаемого. Положим $\varepsilon_1 = E/t_1$, $\delta = 2^{-k}$ и оценим модуль величины

$$V := 1 - (1 - \varepsilon_1)^{\delta}.$$

Напишем ряд Тейлора функции V по степеням ε_1 в окрестности нуля

$$V = \delta\, \varepsilon_1 - \frac{1}{2}\delta\, (\delta - 1)\, {\varepsilon_1}^2 + \frac{1}{3!}\delta\, (\delta - 1)\, (\delta - 2)\, {\varepsilon_1}^3 - \frac{1}{4!}\delta\, (\delta - 1)\, (\delta - 2)\, (\delta - 3)\, {\varepsilon_1}^4 + \ldots$$

Будем считать, что номер k удовлетворяет условию $\varepsilon^{2^k}(n-1) < 1$. Тогда из предыдущего с учетом того, что $|\varepsilon_1| \le \varepsilon^{2^k}(n-1) < 1$, $\delta < 1$ получаем

$$|V| \le \delta\, |\varepsilon_1| \left(1 + |\varepsilon_1| + |\varepsilon_1|^2 + \ldots\right) = \delta\, \frac{|\varepsilon_1|}{1 - |\varepsilon_1|} \le 2^{-k} \frac{\varepsilon^{2^k}(n-1)}{1 - \varepsilon^{2^k}(n-1)}.$$

Отсюда и из (1.13) получаем при достаточно больших k приближенную формулу для нахождения корня z_1:

$$\sqrt[2^k]{-a_{n-1}} = z_1 \cdot (1 + R), \qquad |R| \le 2^{-k} \frac{\varepsilon^{2^k}(n-1)}{1 - \varepsilon^{2^k}(n-1)}. \tag{1.14}$$

Правда, отсюда нельзя однозначно определить корень z_1 из-за неоднозначности операции извлечения корней. Но можно однозначно определить модуль числа z_1, вычислив арифметический корень $\sqrt[2^k]{|a_{n-1}|}$. Тогда с указанной в (1.14) относительной погрешностью R имеем

$$|z_1| \approx \sqrt[2^k]{|a_{n-1}|}. \tag{1.15}$$

Можно также однозначно определить само число z_1, если известен его аргумент, в частности, если известно, что z_1 вещественно и > 0. отметим, что если все корни вещественны, то всегда $a_{n-1} < 0$ и

$$|z_1| \le \sqrt[2^k]{|a_{n-1}|}.$$

Формулу (1.15) можно применять и при оценке границы корней.

Пример. Рассмотрим многочлен с корнями 1, 3:

$$P_0(z) = z^2 - 4\,z + 3.$$

В процессе квадрирования получим

$$P_1(t) = z^2 - 10\,z + 9, \quad P_2(t) = z^2 - 82\,z + 81, \quad P_3(t) = z^2 - 6562\,z + 6561.$$

Извлекая корень порядка $2^3 = 8$ из 6562, получим приближенное значение наибольшего корня 3.0000571.

Пример. Рассмотрим многочлен с корнями 1, -2, 3. Теперь

$$P_0(z) = z^3 - 2\,z^2 - 5\,z + 6, \quad P_1(t) = -t^3 + 14\,t^2 - 49\,t + 36;$$

$$P_2(t) = -t^3 + 98\,t^2 - 1393\,t + 1296, \quad P_3(t) = -t^3 + 6818\,t^2 - 1686433\,t + 1679616.$$

Извлекая корень порядка $2^3 = 8$ из 6818, получим приближенное значение наибольшего корня 3.01444333.

1.5 Локализация корней с помощью правила Штурма

С помощью приведенного в этом параграфе алгоритма можно определить количество вещественных корней многочлена и локализовать их. Под локализацией мы здесь понимаем нахождение малого интервала (a, b), на котором лежит единственный корень многочлена.

Построим для многочлена (1.1) следующую последовательность полиномов: $P_0(z) = P(z)$, $P_1(z) = P'(z)$, $P_2(z)$ — остаток от деления $P_0(z)$ на $P_1(z)$, взятый с обратным знаком, $P_3(z)$ — остаток от деления $P_1(z)$ на $P_2(z)$, взятый с обратным знаком, и так далее, пока не получится константный многочлен $P_m(z)$. Эта последовательность называется последовательностью Штурма многочлена (1.1).

Пусть $a \in \mathbb{R}$ не является корнем многочлена (1.1). Рассмотрим последовательность $\{P_0(a), P_1(a), \ldots, P_m(a)\}$ и обозначим через $W(a)$ количество перемен знака в этой последовательности. Имеет место

Теорема (правило Штурма). *Для всех вещественных $a < b$, не являющихся корнями многочлена* (1) *справедливы следующие утверждения:*
1. $W(a) \geq W(b)$;
2. *количество корней многочлена* (1)*, лежащих на интервале (a, b), равно $W(a) - W(b)$.*

Пример. Рассмотрим многочлен $P(x) = x^3 + 3x - 5$. Построим для него последовательность Штурма:

$$P_0(x) = x^3 + 3x - 5, \quad P_1(x) = 3x^2 + 3, \quad P_2(x) = -2x + 5, \quad P_3(x) = -\frac{87}{4}.$$

Определим количество вещественных корней. Для этого подставим в многочлены полученной последовательности $-\infty$ и ∞. При этом нас интересуют только знаки, поэтому запишем результат так:

$$-\infty : \{-, +, +, -\}, \quad \infty : \{+, +, -, -\}.$$

Видим, что $W(-\infty) = 2$, $W(\infty) = 1$. Поэтому многочлен $P(x)$ имеет один вещественный корень.

Теперь найдем $W(0)$. При подстановке 0 в многочлены последовательности Штурма, получаем $\{-,+,+,-\}$, поэтому $W(0)=2$. Поскольку $W(0)-W(\infty)=1$, корень лежит в $(0;\infty)$.

Аналогичным образом, для 2 получаем последовательность $\{+,+,+,-\}$, поэтому $W(2)=1$. Поскольку $W(0)-W(2)=1$, корень лежит в $(0;2)$.

Наконец, при подстановке 1 в многочлены последовательности Штурма, получаем $\{-,+,+,-\}$, поэтому $W(1)=2$. Так как $W(1)-W(2)=1$, корень лежит в $(1;2)$.

Продолжая разбивать интервал пополам, мы можем локализовать корень с любой требуемой точностью.

Примечание. В процессе построения последовательности Штурма во избежание появления дробных коэффициентов можно умножать многочлены на положительные числа.

2 Численное решение функциональных уравнений и систем нелинейных уравнений

2.1 Локализация корней функционального уравнения

Задача численного (приближенного) решения функционального уравнения

$$f(x) = 0 \tag{2.1}$$

состоит в том, чтобы найти приближенно корень x^* этого уравнения с заданной величиной ε допустимой погрешности (короче, с заданной погрешностью ε). Это означает, что требуется найти число $\tilde{x}$ такое, что $|\tilde{x} - x^*| \le \varepsilon$. Всюду считаем, что функция f достаточное число раз дифференцирума, так что для нее можно применять методы классического анализа (теорему Лагранжа о конечных приращениях, формулу Тейлора и т.п.).

Решение обычно начинают с процесса локализации. Локализация это нахождение малого сегмента $[a, b]$, на котором лежит хотя бы один корень x^* уравнения (2.1). При локализации можно использовать следующую теорему.

Теорема Больцано — Коши. *Если непрерывная на отрезке $[a, b]$ функция $f(x)$ принимает на его концах значения противоположных знаков, т.е. $f(a) \cdot f(b) < 0$, то существует точка $c \in (a, b)$, в которой $f(c) = 0$.*

По этой теореме в качестве отрезка локализации можно взять отрезок, на концах которого функция принимает значения противоположных знаков: $f(a) \cdot f(b) < 0$. (В случае равенства $f(a) \cdot f(b) = 0$ корень уже определяется.)

2.2 Метод половинного деления

Для нахождения приближенного значения кореня x^* уравнения (2.1) с заданной погрешностью ε действуем по следующей схеме. Сначала найдем отрезок $[a, b]$, для которого $f(a) \cdot f(b) < 0$. Затем этот отрезок разделим пополам точкой $c = (a + b)/2$ и вычислим значение $f(c)$. Выполняется одно из двух условий: $f(a) \cdot f(c) = 0$, $f(a) \cdot f(c) < 0$ или $f(a) \cdot f(c) < 0$. Если выполнено первое, то решение найдено точно: $x^* = c$. Если выполнено второе, то переходим к следующему отрезку $[a_1, b_1] = [a, c]$, на концах которого функция принимает значения противоположных знаков. Если же выполнено третье, то переходим к следующему отрезку $[a_1, b_1] = [c, b]$. Затем снова делим отрезок $[a_1, b_1]$ пополам и процесс повторяется.

Получим последовательность вложенных отрезков

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \ldots \supset [a_n, b_n]$$

длины $|b_n - a_n| = 2^{-n}(b - a)$, каждый из которых содержит хотя бы один корень уравнения (2.1).

Процесс завершается как только величина $|b_n - a_n|/2$ станет не больше ε. В качестве приближенного решения можно взять

$$\tilde{x} = \frac{a_n + b_n}{2}, \qquad \text{тогда} \qquad |\tilde{x} - x^*| \le \frac{b - a}{2^{n+1}} \le \varepsilon.$$

2.3 Метод итераций

Этим методом решают уравнения вида

$$x = \varphi(x). \tag{2.2}$$

Примечание 1. Уравнение вида (2.1) можно привести к уравнению вида (2.2) и наоборот. Действительно, если задано уравнение (2.1), то, записав его в эквивалентном виде $\alpha f(x) = 0$, $\alpha \neq 0$, а затем в виде $x = \alpha f(x) + x$, получим уравнение типа (2.2), где $\varphi(x) = \alpha f(x) + x$. Число α выбирают из определенных дополнительных соображений (см. примечание 2). Наоборот, уравнение (2.2) сводится к виду (2.1) с $f(x) = x - \varphi(x)$.

Метод итераций основан на формуле

$$x_n = \varphi(x_{n-1}), \qquad n = 1, 2, \ldots, \tag{2.3}$$

где в качестве x_0 берут какое-либо *начальное приближение* истинного корня, например, любую точку из отрезка локализации.

Возникает вопрос о сходимости последовательности x_k к корню x^*. Этот вопрос решает следующая

Теорема 2.1. *Пусть известно, что корень x^* уравнения $x = \varphi(x)$ лежит в некоторой малой окрестности*

$$u = [x_0 - \delta, x_0 + \delta]$$

точки x_0 (т.е. известен отрезок локализации). При этом в более широкой окрестности

$$U = [x_0 - 2\delta, x_0 + 2\delta]$$

функция φ определена и имеет производную, для которой $|\varphi'(x)| \leq \varepsilon < 1$. Тогда итерационный процесс (2.3) сходится к указанному корню x^* и справедлива следующая оценка погрешности

$$|x_n - x^*| \leq \varepsilon^n |x_0 - x^*| \leq \varepsilon^n \delta. \tag{2.4}$$

Доказательство теоремы 2.1. При $n = 0$ получаем $x_1 = \varphi(x_0)$. Причем по теореме Лагранжа имеем

$$|x_1 - x^*| = |\varphi(x_0) - \varphi(x^*)| = |\varphi'(\xi_1)||x_0 - x^*| \leq \varepsilon |x_0 - x^*| \leq \varepsilon\delta,$$

где $\xi_1 \in (x_0, x^*)$. Отсюда, в частности, получаем, что $x_1 \in U$, поскольку по неравенству треугольника

$$|x_1 - x_0| \leq |x_1 - x^*| + |x_0 - x^*| < \varepsilon\delta + \delta = (\varepsilon + 1)\delta < 2\delta.$$

Повторим итерацию и получим $x_2 = \varphi(x_1)$. Тогда

$$|x_2 - x^*| = |\varphi(x_1) - \varphi(x^*)| = |\varphi'(\xi_2)||x_1 - x^*| \leq \varepsilon |x_1 - x^*| \leq \varepsilon^2 \delta.$$

Мы учли, что $x_1 \in U$ и, следовательно, $\xi_2 \in (x_1, x^*) \subset U$, так что справедлива оценка $|\varphi'(\xi_2)| \leq \varepsilon < 1$. Отсюда, в частности, снова получаем, что $x_2 \in U$, поскольку по неравенству треугольника

$$|x_2 - x_0| \leq |x_2 - x^*| + |x_0 - x^*| < \varepsilon^2\delta + \delta = (\varepsilon^2 + 1)\delta < 2\delta$$

Так продолжая, на шаге n получим

$$|x_n - x^*| = |\varphi(x_{n-1}) - \varphi(x^*)| = |\varphi'(\xi_n)||x_{n-1} - x^*| \leq \varepsilon |x_{n-1} - x^*| \leq \varepsilon^n \delta,$$

и $x_n \in U$, поскольку

$$|x_n - x_0| \leq |x_n - x^*| + |x_0 - x^*| \leq (\varepsilon^n + 1)\delta \leq 2\delta.$$

Таким образом, процесс можно продолжать неограниченно; при всех n выполнена оценка (2.4). Теорема 1 доказана.

Примечание 2. Из (2.4) видно, что скорость сходимости итерации увеличивается при уменьшении значения $\varepsilon < 1$. Для увеличения скорости сходимоси можно предварительно провести следующие преобразования. Пусть имеется уравнение (2.2). Тогда при $\alpha \neq 0$ имеем

$$x - \varphi(x) = 0 \quad \Leftrightarrow \quad \alpha(x - \varphi(x)) = 0 \quad \Leftrightarrow \quad x = x + \alpha(x - \varphi(x)).$$

Получили уравнение снова вида (2.2), но с новой правой частью:

$$x = \Phi(x), \qquad \Phi(x) = x + \alpha(x - \varphi(x)).$$

Предположим, что $\varphi'(x_0) \neq 1$ и выберем величину α из условия $\Phi'(x_0) = 0$, т.е.

$$1 + \alpha(1 - \varphi'(x_0)) = 0 \quad \Leftrightarrow \quad \alpha = -\frac{1}{1 - \varphi'(x_0)}.$$

Теперь есть основание ожидать, что абсолютная величина производной $\Phi(x)$ будет достаточно малой во всей окрестности U.

Следует провести хотя бы грубую оценку $|\Phi'(x)|$ в U и убедиться, что она мала. В противном случае следует найти более точное первоначальное приближение x_0 к корню x^*, и уменьшить длину первоначального сегмента $u = [x_0 - \delta, x_0 + \delta]$.

При переходе от уравнения $f(x) = 0$ к равносильному уравнению $x = \varphi(x)$, где $\varphi(x) = x + \alpha f(x)$, величину α также целесообразно выбрать из условия $\varphi'(x_0) = 1 + \alpha f'(x_0) = 0$, т.е.

$$\alpha = -\frac{1}{f'(x_0)}.$$

Таким образом, для уравнения вида (2.1) получаем расчетную формулу

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}.$$

Пример 1. Найти решение уравнения $x^2 = 2$. Сначала составляем уравнение, пригодное для итераций:

$$x = x + \alpha(x^2 - 2), \qquad \varphi(x) = x + \alpha(x^2 - 2), \qquad \varphi'(x) = 1 + 2x\alpha.$$

В качестве начального приближения возьмем $x_0 = 1.5$. Легко проверить, что $\sqrt{2}$ лежит на отрезке $[1.5 - 0.1, 1.5 + 0.1]$, т.е. в данном случае $\delta = 0.1$. Определим α, решая уравнение $\varphi'(x_0) = 0$:

$$\varphi'(x_0) = 1 + 2x_0\alpha = 0, \qquad \alpha = -\frac{1}{2x_0} = -\frac{1}{3}.$$

Оценим производную $\varphi'(x) = 1 + 2x\alpha = 1 - \frac{2}{3}x$ на сегменте $U = [x_0 - 2\delta, x_0 + 2\delta] = (1.3, 1.7]$. Поскольку она является монотонной функцией, то достаточно найти максимум модуля на концах отрезка U. Он равен $\varepsilon = \frac{2}{15}$. Поскольку $\varepsilon < 1$, то можно начинать процесс итераций.

Уравнение для итераций имеет вид

$$x_{n+1} = x_n - \frac{1}{3}(x_n^2 - 2).$$

Отсюда получаем

$$x_1 = \frac{17}{12}, \quad x_1 = \frac{611}{432}, \quad x_3 = \frac{791783}{559872} \approx 1.41422146,$$

причем погрешности $\Delta_n = |x_n - x^*|$ на шаге n оцениваются по формуле $\Delta_n \le \varepsilon^n \delta$ (см. (2.4)):

$$\Delta_1 \le \frac{1}{75}, \quad \Delta_1 \le \frac{2}{1125}, \quad \Delta_3 \le \frac{4}{16875} \approx 0.000237.$$

Точное значение $\sqrt{2} = 1.4142135623730950488$, так что оценка погрешности получена с завышением.

Пример 2. Найти решение уравнения $xe^x = 2$, то есть $f(x) = 0$, где $f(x) = xe^x - 2$. Сначала составляем уравнение, пригодное для итераций:

$$x = x + \alpha\,(xe^x - 2)\,, \qquad \varphi(x) = x + \alpha\,(xe^x - 2)\,, \qquad \varphi'(x) = 1 + \alpha\,(1+x)\,e^x.$$

В качестве начального приближения возьмем $x_0 = 0.9$. Проверяются неравенства $f(x_0 - 0.1) < 0$, $f(x_0 + 0.1) > 0$. Значит, в данном случае можно положить $\delta = 0.1$ и корень лежит на отрезке $[0.8; 1]$. Определим α, решая уравнение $\varphi'(x_0) = 0$:

$$\varphi'(x_0) = 1 + \alpha\,(1+x_0)\,e^{x_0} = 0, \qquad \alpha\,(1+0.9)\,e^{0.9} = -1.$$

Отсюда находим $\alpha = -0.2139840315\ldots$. Чтобы расчетная формула не была громоздкой возьмем $\alpha = -0.2$. Тогда

$$\varphi(x) = x - 0.2\,(xe^x - 2)\,, \qquad \varphi'(x) = 1 - 0.2\,(1+x)\,e^x.$$

Оценим производную $\varphi'(x)$ на сегменте $U = [x_0 - 2\delta, x_0 + 2\delta] = (1.7, 1.1]$. Поскольку

$$\varphi''(x) = -0.2\,e^x\,(2+x) < 0,$$

то функция $\varphi'(x)$ монотонной убывает на указанном сегменте. Следовательно, достаточно найти максимум модуля на концах отрезка U. Вычисляем

$$\varphi'(1.7) = 0.315\ldots, \quad \varphi'(1.1) = -0.261\ldots, \quad \max\{|\varphi'(1.7)|, |\varphi'(1.1)|\} = 0.315\ldots.$$

Итак $\varepsilon < 0.4 < 1$, и можно начинать процесс итераций.

Уравнение для итераций имеет вид

$$x_{n+1} = x_n - 0.2\,(x_n e^{x_n} - 2)\,,$$

погрешность оценивается так: $|x_n - x^*| < 0.1 \cdot (0.4)^n$.

2.4 Метод Ньютона (метод касательных)

Этим методом решают уравнения вида (2.1). Расчетная формула имеет вид

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad n = 0, 1, \ldots. \tag{2.5}$$

Для получения оценки скорости сходимости итерационного процесса (2.5) докажем вспомогательное утверждение.

Лемма. *Пусть x^* — корень уравнения $f(x) = 0$, и пусть x_0 — некоторая точка. Тогда если $m := \min_{x \in [x^*, x_0]} |f'(x)| > 0$, то*

$$|x_0 - x^*| \le \frac{|f(x_0)|}{m}, \tag{2.6}$$

Доказательство. По формуле Тейлора с остаточным членом в форме Лагранжа имеем

$$f(x_0) = f(x^*) + f'(\xi)(x_0 - x^*) = f'(\xi)(x_0 - x^*), \quad \xi \in (x^*, x_0),$$

где $f(x^*) = 0$, откуда

$$|f(x_0)| = |f'(\xi)||x_0 - x^*| \geq m|x_0 - x^*|,$$

что и доказывает неравенство (2.6). Лемма доказана.

Теорема 2.2. *Пусть x^* — корень уравнения $f(x) = 0$, и пусть точка x_{n+1} получена из некоторой точки x_n по формуле Ньютона* (5)*. Тогда*

$$|x_{n+1} - x^*| \leq \frac{1}{2}\frac{M_n}{m_n}|x_{n+1} - x_n|^2, \tag{2.7}$$

где

$$M_n := \max_{x\in[x_n, x_{n+1}]} |f''(x)|, \qquad m_n := \min_{x\in[x^*, x_{n+1}]} |f'(x)|.$$

Доказательство. По формуле Тейлора с остаточным членом в форме Лагранжа имеем

$$f(x_{n+1}) = f(x_n) + f'(x_n)(x_{n+1} - x_n) + \frac{1}{2}f''(\xi)(x_{n+1} - x_n)^2, \quad \xi \in (x_n, x_{n+1}).$$

Поскольку из (2.5) следует, что

$$f(x_n) + f'(x_n)(x_{n+1} - x_n) = 0,$$

то

$$f(x_{n+1}) = \frac{1}{2}f''(\xi)(x_{n+1} - x_n)^2, \quad |f(x_{n+1})| \leq \frac{1}{2}M_n|x_{n+1} - x_n|^2.$$

Отсюда и из (2.6) получается

$$|x_{n+1} - x^*| \leq \frac{|f(x_{n+1})|}{m_n} \leq \frac{M_n}{2m_n}|x_{n+1} - x_n|^2,$$

что и требовалось.

Примечание 3. Процесс решения задачи (2.1), как обычно, начинается с процесса локализации корней. Предположим, что найден малый отрезок $[a, b]$, на котором лежит корень x^* и на котором сохраняют свои знаки производные $f'(x)$, $f''(x)$. Тогда, выбрав в качестве начального приближения x_0 один из концов отрезка (лежащий со стороны выпуклости графика $y = f(x)$) получим, что все точки x_n монотонно приближаются к x^* с одной из сторон, оставаясь на отрезке $[a, b]$ (рис.) Теперь удобно вместо оценки (2.7) использовать более простую оценку

$$|x_{n+1} - x^*| \leq C|x_{n+1} - x_n|^2, \qquad \text{где} \qquad C = \frac{1}{2}\frac{\tilde{M}}{\tilde{m}}, \tag{2.8}$$

где

$$\tilde{M} := \max_{x\in[a,b]} |f''(x)|, \qquad \tilde{m} := \min_{x\in[a,b]} |f'(x)|.$$

Примечание 4. При расчетах по формуле (2.5) надо следить, чтобы число значащих цифр числа x_{n+1} соответствовало малости погрешности. Грубо говоря, следует придерживаться следующего правила.

Предположим на шаге $n+1$ вычислено значение x_{n+1} со всеми возможными для калькулятора значащими цифрами. Вычислим $C|x_{n+1} - x_n|^2$. Если после запятой в полученном

выражении идут m нулей, то в записи числа x_{n+1} оставляем после запятой только $m+1$ цифр, а остальные отбрасываем. Например, если $x_k = 1.234$, $x_{k+1} = 1.23456789...$, $C = 1$, то погрешность оценивается как $(0.00056789...)^2 = 0.0000003...$, и берем $x_{k+1} = 1.234567$.

Пример. Найдем решение уравнения $x^3 = 2$. В данном случае $f(x) = x^3 - 2$. Расчетная формула имеет вид

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},\ x_{n+1} = x_n - \frac{x_n^3 - 2}{3x_n^2},$$

$$x_{n+1} = \frac{2}{3}\frac{x_n^3 + 1}{x_n^2}$$

В качестве начального приближения возьмем точку $x_0 = 1.3$. Тогда на первом шаге получим $x_1 = 1.261143984...$ Вычислим выражение $|x_0 - x_1|^2 = 0.001509....$ Значит, значение x_1 округляем до третьего знака после запятой, т.е. берем $x_1 = 1.261$. Повторим процесс, найдем $x_2 = 1.259921973$. Вычислим $|x_1 - x_2|^2 = 0.00000116...$, и округлим x_2 до шестого знака после запятой, т.е. берем $x_2 = 1.259921$. На следующем шаге $x_3 = 1.259921050$, $|x_2 - x_3|^2 = 2.5 \times 10^{-15}$. Вычисления можно прекращать.

В конце еще можно получить оценку погрешности по формуле (2.8). Мы этого здесь не делаем. Отметим лишь, что полученные значение x_3 отличается от точного $\sqrt[3]{2} = 1.2599210498948..$ меньше чем на 10^{-9}.

2.5 Комбинированный метод хорд и касательных

Этим методом приближенного решения уравнения вида (2.1) строятся вложенные отрезки $[a_k, b_k]$, сожержащие корень x^* как и в методе половинного деления, но длины этих отрезков убывают значительно быстрее. При использовании комбинированного метода получаемые приближения, вообще говоря, сходятся к корню медленнее, чем в методе касательных (Ньютона), но зато здесь не нужно проводить трудоемкие оценки погрешности.

Выпишем расчетные формулы, которые получим ниже:

$$b_{n+1} = b_n - \frac{f(b_n)}{f'(b_n)}, \tag{2.9}$$

$$a_{n+1} = a_n - f(a_n)\frac{a_n - b_{n+1}}{f(a_n) - f(b_{n+1})}. \tag{2.10}$$

Здесь все точки b_k находятся со стороны выпуклости графика, а a_k — со стороны вогнутости графика.

Пусть $[a_0, b_0]$ начальный отрезок, на котором лежит корень уравнения (2.1). Предположим для определенности, что на этом отрезке $f'(x) > 0$, $f''(x) > 0$, то есть график выпуклый вниз. В частности, из монотонного возрастатия функции f следует, что $f(a_0) < 0 < f(b_0)$.

1-й шаг. Проведем касательную к графику $y = f(x)$ в точке b_0. Ее уравнение

$$y - f(b_0) = f'(b_0)(x - b_0),$$

откуда, положив $y = 0$, найдем точку пересечения b_1 касательной и оси OX:

$$-f(b_0) = f'(b_0)(b_1 - b_0), \qquad b_1 = b_0 - \frac{f(b_0)}{f'(b_0)}.$$

Из условий $f'(x) > 0$, $f''(x) > 0$ следует, что $a_0 < x^* < b_1 < b_0$ (рис.).

Теперь проведем хорду, соединяющую точки $(a_0, f(a_0))$ и $(b_1, f(b_1))$. Ее уравнение

$$\frac{x - a_0}{b_1 - a_0} = \frac{y - f(a_0)}{f(b_1) - f(a_0)},$$

откуда, положив $y = 0$, найдем точку пересечения a_1 хорды и оси OX:

$$\frac{a_1 - a_0}{b_1 - a_0} = \frac{-f(a_0)}{f(b_1) - f(a_0)}, \quad a_1 = a_0 - f(a_0)\frac{b_1 - a_0}{f(b_1) - f(a_0)}.$$

Из условий $f'(x) > 0$, $f''(x) > 0$, $f(a_0) < 0$ следует, что (рис.)

$$a_0 < a_1 < x^* < b_1 < b_0, \qquad f(a_0) < f(a_1) < 0 < f(b_1) < f(b_0).$$

Первый шаг завершен.

2-й шаг. Повторяет первый с заменой $[a_0, b_0]$ на $[a_1, b_1]$. В результате получакм новый отрезок $a_2 < b_2$ со свойством

$$a_1 < a_2 < x^* < b_2 < b_1, \qquad f(a_1) < f(a_2) < 0 < f(b_2) < f(b_1).$$

И так далее. В результате получаются указанные формулы (2.9), (2.10).

Аналогично получают расчетные формулы для других случаев.

2.6 Нахождение корней аналитических функций

1. Метод итераций для нахождения комплексных корней. Аналог метода итераций, рассмотренного нами для действительного случая, справедлив и для аналитических функций $\varphi(z)$. В этом случае мы ищем комплексный корень уравнения $z = \varphi(z)$. Сформулируем соответствующее утверждение.

Теорема 2.1'. *Пусть известно, что корень z^* уравнения $z = \varphi(z)$ лежит в некоторой малой окрестности*

$$u: \ |z - z_0| \le \delta$$

начального приближения z_0 этого корня. При этом в более широкой окрестности

$$U: \ |z - z_0| \le 2\delta$$

функция $\varphi(z)$ аналитична и $|\varphi'(z)| \le \varepsilon < 1$. Тогда итерационный процесс $z_n = \varphi(z_{n-1})$ сходится к указанному корню z^* и справедлива следующая оценка погрешности

$$|z_n - z^*| \le \varepsilon^n |z_0 - z^*| \le \varepsilon^n \delta. \tag{2.11}$$

Доказательство этой теоремы аналогично доказательству для действительного случая.

Доказательство теоремы 2.1'. При $n = 0$ получаем $z_1 = \varphi(z_0)$. Причем

$$|z_1 - z^*| = |\varphi(z_0) - \varphi(z^*)| = \left| \int_{z_0}^{z^*} \varphi'(\xi)\, d\xi \right| \le \int_{z_0}^{z^*} |\varphi'(\xi| \ |d\xi| \le$$

$$\le \max_{\xi \in [z_0, z^*]} |\varphi'(\xi)| \int_{z_0}^{z^*} |d\xi| \le \max_{\xi \in U} |\varphi'(\xi)| |z_0 - z^*| \le \varepsilon |z_0 - z^*| \le \varepsilon\delta, \tag{2.12}$$

где интеграл берется по прямолинейному отрезку $[z_0, z^*]$ (лежащему в круге U) и

$$\int_{z_0}^{z^*} |d\xi| = |z_0 - z^*|.$$

Отсюда, в частности, получаем, что $z_1 \in U$, ибо по неравенству треугольника

$$|z_1 - z_0| \le |z_1 - z^*| + |z_0 - z^*| < \varepsilon |z_0 - z^*| + |z_0 - z^*| = (\varepsilon + 1)|z_0 - z^*| < 2\delta.$$

Повторим итерацию и получим $z_2 = \varphi(z_1)$. Как и в (2.12) находим

$$|z_2 - z^*| = |\varphi(z_1) - \varphi(z^*)| \le \int_{z_1}^{z^*} |\varphi'(\xi|\, |d\xi| \le$$

$$\le \max_{\xi\in[z_1,z^*]} |\varphi'(\xi)| \int_{z_1}^{z^*} |d\xi| \le \max_{\xi\in U} |\varphi'(\xi)||z_1 - z^*| \le \varepsilon|z_1 - z^*| \le \varepsilon^2\delta.$$

Здесь в третьем неравенстве при оценке $\max_{\xi\in[z_1,z^*]} |\varphi'(\xi)|$ мы учли, что $z_1 \in U$ и, следовательно, прямолинейный отрезок $[z_1, z^*]$ лежит в круге U (так что на этом отрезке модуль производной по условию теоремы не превосходит ε). Отсюда вновь получаем, что $z_2 \in U$, поскольку по неравенству треугольника

$$|z_2 - z_0| \le |z_2 - z^*| + |z_0 - z^*| < \varepsilon^2|z_0 - z^*| + |z_0 - z^*| = (\varepsilon^2 + 1)|z_0 - z^*| < 2\delta.$$

Так продолжая, на шаге n получим

$$|z_n - z^*| = |\varphi(z_{n-1}) - \varphi(z^*)| \le \max_{\xi\in U} |\varphi'(\xi)||z_{n-1} - z^*| \le \varepsilon|z_{n-1} - z^*| \le \varepsilon^n\delta,$$

и при этом $z_n \in U$, поскольку

$$|z_n - z_0| \le |z_n - z^*| + |z_0 - z^*| \le (\varepsilon^n + 1)|z_0 - z^*| \le 2\delta.$$

Таким образом, процесс можно продолжать неограниченно; при всех n выполнена оценка (2.11). Теорема 2.1' доказана.

2. Описанный выше метод Ньютона можно применять и для нахождения комплексных корней аналитических функций. В этом случае применяется аналогичная расчетная формула:

$$z_{n+1} = z_n - \varepsilon\frac{f(z_n)}{f'(z_n)}, \qquad \varepsilon > 0, \tag{2.13}$$

где ε достаточно малое положительное вещественное число.

Действительно, пусть $f(z) = u(x, y) + iv(x, y)$. Из условий Коши-Римана имиеем

$$\frac{f(z)}{f'(z)} = \frac{u + iv}{u_x + iv_x} = \frac{(uu_x + vv_x) + i(vu_x - uv_x)}{u_x^2 + v_x^2} =$$

$$= \frac{(uu_x + vv_x) + i(vv_y + uu_y)}{u_x^2 + v_x^2} = \frac{1}{2}\frac{(u^2 + v^2)'_x + i(u^2 + v^2)'_y}{u_x^2 + v_x^2}.$$

Значит, получается следующее равенство векторов

$$\frac{f(z_n)}{f'(z_n)} = \left(\mathrm{Re}\frac{f(z)}{f'(z)}, \mathrm{Im}\frac{f(z)}{f'(z)}\right) = \frac{1}{2}\frac{\mathrm{grad}\,|f(z)|^2}{|f'(z)|^2}. \tag{14}$$

Поэтому формула (2.13) — это формула известного метода антиградиента для нахождения минимума функции $|f(z)|$. В данном случае этот минимум равен нулю.

Пример. $f(z) = z^2 + 1$. Решаем уравнение $f(z) = 0$. Возьмем в качестве начального приближения 0 $z_0 = \frac{1}{2} + \frac{1}{2}i$. Тогда по формулам (2.13) получим

$$z_1 = -\frac{1}{4} + \frac{3}{4}i, \quad z_2 = \frac{3}{40} + \frac{39}{40}, \quad z_3 = -\frac{7}{4080} + \frac{4069}{4080}i, \quad z_4 = \frac{217}{46748640} + \frac{46748741}{46748640}\,i.$$

Видно, что $z_k \to i$. Например, $z_4 = 0.00000464 + 1.00000216\,i$.

2.7 Решение систем нелинейных уравнений

Мы ограничимся случаем систем вида

$$\begin{cases} f(x,y) &= 0 \\ g(x,y) &= 0 \end{cases} \tag{2.15}$$

Предположим, что в окрестности искомого решения (x^*, y^*) определитель

$$D(x,y) = \begin{vmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{vmatrix} \neq 0 \tag{2.16}$$

Предположим также, что мы знаем «хорошее» начальное приближение (x_0, y_0). Тогда по формуле Тейлора

$$0 = f(x^*, y^*) = f(x_0, y_0) + f'_x(x_0, y_0) \cdot (x^* - x_0) + f'_y(x_0, y_0) \cdot (y^* - y_0) +$$

$$+\frac{1}{2}\left(f''_{xx}(c_1, c_2) \cdot (x^* - x_0)^2 + 2f''_{xy}(c_1, c_2) \cdot (x^* - x_0)(y^* - y_0) + f''_{yy}(c_1, c_2) \cdot (y^* - y_0)^2\right), \tag{2.17}$$

где $c_1 \in (x^*, x_0)$, $c_2 \in (y^*, y_0)$. Дифференциал второго порядка является бесконечно малой величиной второго порядка. Отбросим его, тогда точку (x^*, y^*) в правой части равенстве (2.17) придется заменить на некоторую (x_1, y_1), близкую к этой точке. Есть основание считать, что эта точка ближе к истинному значению решения, чем исходная точка (x_0, y_0). Получаем равенство

$$f'_x(x_0, y_0) \cdot (x_1 - x_0) + f'_y(x_0, y_0) \cdot (y_1 - y_0) = -f(x_0, y_0).$$

Запишем аналогичное равенство для g:

$$g'_x(x_0, y_0) \cdot (x_1 - x_0) + g'_y(x_0, y_0) \cdot (y_1 - y_0) = -g(x_0, y_0).$$

Получилась система линейных уравнений для определения точки (x_1, y_1). Запишем ее в матричном виде

$$A(x_0, y_0) \cdot \begin{pmatrix} x_1 - x_0 \\ \\ y_1 - y_0 \end{pmatrix} = \begin{pmatrix} -f(x_0, y_0) \\ \\ -g(x_0, y_0) \end{pmatrix},$$

где

$$A(x_0, y_0) = \begin{pmatrix} \frac{\partial f}{\partial x}(x_0, y_0) & \frac{\partial f}{\partial y}(x_0, y_0) \\ \\ \frac{\partial g}{\partial x}(x_0, y_0) & \frac{\partial g}{\partial y}(x_0, y_0) \end{pmatrix}$$

Отсюда найдем

$$\begin{pmatrix} x_1 \\ \\ y_1 \end{pmatrix} = \begin{pmatrix} x_0 \\ \\ y_0 \end{pmatrix} - A^{-1}(x_0, y_0) \cdot \begin{pmatrix} f(x_0, y_0) \\ \\ g(x_0, y_0) \end{pmatrix},$$

где $A^{-1}(x_0, y_0)$ есть матрица, обратная к $A(x_0, y_0)$. Мы учли, что определитель $D(x_0, y_0) \neq 0$ и, следовательно, обратная матрица

$$A^{-1}(x_0, y_0) = \frac{1}{f'_x g'_y - f'_y g'_x} \begin{pmatrix} g'_y & -f'_y \\ \\ -g'_x & f'_x \end{pmatrix}$$

существует (все частные производные вычислены в точке (x_0, y_0)). Заменяя здесь (x_0, y_0) на (x_n, y_n) и (x_1, y_1) на (x_{n+1}, y_{n+1}), получим расчетную формулу для итераций:

$$\begin{pmatrix} x_{n+1} \\ \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ \\ y_n \end{pmatrix} - A^{-1}(x_n, y_n) \cdot \begin{pmatrix} f(x_n, y_n) \\ \\ g(x_n, y_n) \end{pmatrix}. \tag{2.18}$$

Получилась формула, вполне аналогичная формуле Ньютона (2.5).

Пример. Найдем ненулевое решение системы

$$\begin{cases} x + y & = 3 \\ xy & = 1 \end{cases}$$

Вычисляем матрицы

$$A(x, y) = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ \\ y & x \end{pmatrix};$$

$$A^{-1}(x, y) = \frac{1}{x - y} \begin{pmatrix} x & -1 \\ \\ -y & 1 \end{pmatrix};$$

Расчетная формула

$$\begin{pmatrix} x_{n+1} \\ \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ \\ y_n \end{pmatrix} - \frac{1}{x_n - y_n} \begin{pmatrix} x_n & -1 \\ \\ -y_n & 1 \end{pmatrix} \begin{pmatrix} x_n + y_n - 3 \\ \\ x_n y_n - 1 \end{pmatrix}$$

Это можно несколько упростить.

2.8 Решение систем нелинейных комплексных уравнений

Мы ограничимся случаем систем вида

$$\begin{cases} f(\alpha, \beta) & = 0 \\ g(\alpha, \beta) & = 0 \end{cases}$$

Будем считать, что f и g — аналитические функции двух комплексных переменных α, β. Приращение функций приближенно равно дифференциалу первого порядка, то есть

$$f(\alpha + \delta_\alpha, \beta + \delta_\beta) - f(\alpha, \beta) \approx f'_\alpha(\alpha, \beta) \cdot \delta_\alpha + f'_\beta(\alpha, \beta) \cdot \delta_\beta.$$
$$g(\alpha + \delta_\alpha, \beta + \delta_\beta) - g(\alpha, \beta) \approx g'_\alpha(\alpha, \beta) \cdot \delta_\alpha + g'_\beta(\alpha, \beta) \cdot \delta_\beta.$$

Мы должны выбрать комплексные приращения δ_α и δ_β так, чтобы

$$|f(\alpha + \delta_\alpha, \beta + \delta_\beta)| < |f(\alpha, \beta)|, \quad |g(\alpha + \delta_\alpha, \beta + \delta_\beta)| < |g(\alpha, \beta)|.$$

Для этого потребуем, чтобы вектор $f(\alpha + \delta_\alpha, \beta + \delta_\beta) - f(\alpha, \beta)$ был направлен от точки $f(\alpha, \beta)$ к началу координат, то есть, чтобы выполнылось равенство

$$f(\alpha + \delta_\alpha, \beta + \delta_\beta) - f(\alpha, \beta) = -\varepsilon f(\alpha, \beta),$$

где ε — положительное число, меньшее единицы. Такое же равенство запишем и для функции g:

$$g(\alpha + \delta_\alpha, \beta + \delta_\beta) - g(\alpha, \beta) = -\varepsilon g(\alpha, \beta).$$

Тогда будем иметь

$$-\varepsilon f(\alpha,\beta) \approx f'_\alpha(\alpha,\beta)\cdot\delta_\alpha + f'_\beta(\alpha,\beta)\cdot\delta_\beta.$$
$$-\varepsilon g(\alpha,\beta) \approx g'_\alpha(\alpha,\beta)\cdot\delta_\alpha + g'_\beta(\alpha,\beta)\cdot\delta_\beta.$$

Предположим, что в рассматриваемой нами области значений α и β не равен нулю определитель $D(\alpha,\beta)$ матрицы

$$A(\alpha,\beta) = \begin{pmatrix} \frac{\partial f}{\partial\alpha} & \frac{\partial f}{\partial\beta} \\ \\ \frac{\partial g}{\partial\alpha} & \frac{\partial g}{\partial\beta} \end{pmatrix}$$

Тогда для смещений δ_α и δ_β находим

$$-\varepsilon \begin{pmatrix} f(\alpha,\beta) \\ \\ g(\alpha,\beta) \end{pmatrix} \approx A(\alpha,\beta)\cdot \begin{pmatrix} \delta-_\alpha \\ \\ \delta_\beta \end{pmatrix}, \quad \begin{pmatrix} \delta_\alpha \\ \\ \delta_\beta \end{pmatrix} \approx -\varepsilon A^{-1}(\alpha,\beta) \begin{pmatrix} f(\alpha,\beta) \\ \\ g(\alpha,\beta) \end{pmatrix}$$

Отсюда, подставаив $\delta_\alpha = \alpha_{k+1} - \alpha_k$ и $\delta_\beta = \beta_{k+1} - \beta_k$, получим следующий метод итераций Ньютона

$$\begin{pmatrix} \alpha_{k+1} \\ \\ \beta_{k+1} \end{pmatrix} = \begin{pmatrix} \alpha_k \\ \\ \beta_k \end{pmatrix} - \varepsilon A^{-1}(\alpha,\beta) \begin{pmatrix} f(\alpha,\beta) \\ \\ g(\alpha,\beta) \end{pmatrix}.$$

3 Задача аппроксимации многочленами

Задача о приближении (аппроксимации) заданной непрерывной функции на отрезке $[a, b]$ посредством многочленов формулируется следующим образом. Требуется найти многочлен P_n заданной степени n такой, чтобы величина $|P_n(x)-f(x)|$ была как можно меньше во всех точках отрезка $[a, b]$. Другими словами, требуется найти многочлен, для которого мала норма

$$\|f - P_n\|_{C[a,b]} := \max_{x\in[a,b]} |P_n(x) - f(x)|, \tag{3.1}$$

которую естественно назвать *погрешностью аппроксимации*. Погрешность аппроксимации в теории приближений называют называют также уклонением на отрезке $[a, b]$ многочлена $P_n(x)$ от функции $f(x)$. Приближение (аппроксимация), естественно, тем лучше, чем меньше погрешность аппроксимации. Это классическая задача *конструктивной теории функций* или *теории приближения*. Она часто возникает в численном анализе, например, если требуется изучить поведение некоторой трудновычислимой функции, то ее заменяют на более простой объект — аппроксимирующий многочлен и его значения используют при анализе функции.

3.1 Многочлены наилучшего приближения

Одна из основополагающих теорем теории аппроксимаций, теорема Э. Бореля, утверждает, что при каждом натуральном n существует многочлен $P_n^*(x)$ *наилучшего приближения*, для которого погрешность аппроксимации минимальна. Это означает, что

$$\|f - P_n^*\|_{C[a,b]} \le \|f - P_n\|_{C[a,b]},$$

для любого многочлена P_n степени не выше чем n. При этом равенство достигается лишь в случае, когда $P_n = P_n^*$, то есть многочлен наилучшего приближения единствен. В теории приближений принято обозначение:

$$E_n(f) = E_n(f, [a, b]) = \|f - P_n^*\|_{C[a,b]},$$

причем $E_n(f)$ называется величиной наилучшего приближения функции f.

Теорема о существовании многочлена наилучшего приближения не утверждает, что произвольная непрерывная на отрезке $[a, b]$ функции $f(x)$ может быть сколь угодно точно аппроксимирована такими многочленами, то есть что $E_n(f) \to 0$ при $n \to \infty$. Задача о возможности сколь угодно точной аппроксимации решена Вейерштрассом. Ему принадлежит следующая теорема.

Теорема.(К. Вейерштрасс). *Пусть $f(x) \in C[a, b]$ и ε — любое сколь угодно малое положительное число. Тогда существует многочлен $P_n(x)$ такой, что*

$$\|f - P_n\|_{[a,b]} = \max_{x\in[a,b]} |f(x) - P(x)| \le \varepsilon.$$

Поскольку это неравенство тем более выполнено для многочлена P_n^*, то из теоремы Вейерштрасса вытекает, что $E_n(f) \to 0$ при $n \to \infty$.

Существуют алгоритмы построения многочлена наилучшего приближения $P_n^*(x)$, однако они весьма трудоемки. Иногда предпочтительнее строить приближающие многочлены из условий *интерполяции*. Они могут и не быть многочленами наилучшего приближения, но зато получаются по очень простым алгоритмам.

3.2 Полиномиальная интерполяция

3.2.1 Задача интерполяции. Интерполяционные многочлены

Пусть на некотором отрезке $[a,b]$ определена непрерывная функция $f(x)$. Пусть задан набор попарно различных точек $x_1,\dots,x_{n+1}$, $x_k \in [a,b]$, которые называют узлами интерполяции. Требуется найти многочлен P_n, совпадающий с функцией $f(x)$ в узлах интерполяции, т.е. многочлен, для которого $P_n(x_k) = f(x_k) = y_k$ при $k = 1,\dots,n+1$. Многочлен P_n называют интерполяционным. При построении интерполяционного многочлена, как видно из задачи, не учитываются какие-либо свойства функции $f(x)$, а учитываются лишь ее значения в узлах интерполяции. Фактически для построения многочлена надо знать лишь множество пар (x_k, y_k), $k = 1,\dots,n+1$. Это множество называют таблицей интерполяции. Сам процесс построения многочлена называют интерполяцией по заданной таблице. Указанная интерполяция называется простой, поскольку узлы $\{x_k\}$ простые (не кратные).

Оказывается, что интерполяционный многочлен с заданной таблицей интерполяции всегда существует и притом единственен. Действительно, для всех $k = 1,2,\dots,n+1$ подставив x_k в многочлен (1.1) и приравняв его к соответствующему y_k, получим систему линейных алгебраических уравнений относительно неизвестных коэффициентов p_j. Определитель матрицы этой системы — определитель Вандермонда. Он не равен нулю, поскольку узлы попарно различны. Следовательно, по правилу Крамера, эта система имеет единственное решение.

Существует множество способов построения интерполяционных многочленов. Мы остановимся на двух.

1. Метод Ньютона. Положим

$$P_n(x) = c_0 + c_1(x-x_1) + c_2(x-x_1)(x-x_2) + \dots$$

$$+c_{n-1}(x-x_1)(x-x_2)...(x-x_{n-1}) + c_n(x-x_1)(x-x_2)...(x-x_n),$$

где числа c_k подлежат определению. Ясно, что P_n является многочленом степени n. Для определения числа c_1 рассмотрим первое интерполяционное равенство $P_n(x_1) = y_1$. Поскольку $P_n(x_1) = c_0$, то отсюда находим $c_0 = y_1$. Для определения числа c_2 рассмотрим второе интерполяционное равенство $P_n(x_2) = y_2$. Поскольку $P_n(x_2) = c_0 + c_1(x_2 - x_1)$ и c_0 уже определено, то отсюда находим

$$c_0 + c_1(x_2 - x_1) = y_2, \;\; c_1 = \frac{y_2 - c_0}{x_2 - x_1}.$$

И так далее. Для определения числа c_{n-1} рассмотрим n-е интерполяционное равенство $P_n(x_n) = y_n$. Поскольку

$$P_n(x_n) = c_0 + c_1(x_n - x_1) + c_2(x_n - x_1)(x_n - x_2) + \dots$$

$$+c_{n-1}(x_n - x_1)(x_n - x_2)\dots(x_n - x_{n-1}) = y_n,$$

и все c_k с номерами $k = 0,\dots,n-2$ уже определены, то отсюда находим c_{n-1}. Наконец, последний коэффициент c_n определяется из последнего интерполяционного равенства $P_n(x_{n+1}) = y_{n+1}$, т.е.

$$P_n(x_{n+1}) = c_0 + c_1(x_{n+1} - x_1) + . + c_n(x_{n+1} - x_1)(x_{n+1} - x_2)...(x_{n+1} - x_n) = y_{n+1}.$$

Отметим, что в результате может получиться многочлен степени $< n$.

Пример построения многочлена, интерполирующего данную функцию в точках $x_1 = 1$, $x_2 = 2$, $x_3 = 3$. Получается

$$P_2(x) = y_1 + (y_2 - y_1)(x-1) + \frac{1}{2}(y_1 - 2y_2 + y_3)(x-1)(x-2).$$

2. Метод Лагранжа. По простоте построения интерполяционного многочлена этот метод не уступает предыдущему. Построим сначала многочлен $l_k(x)$ степени n, удовлетворяющий равенствам

$$\begin{cases} l_k(x_k) = 1, \\ l_k(x_j) = 0, \quad j \neq k. \end{cases}$$

Такой многочлен имеет вид

$$l_k(x) = \frac{(x-x_1)(x-x_2)\dots(x-x_{k-1})(x-x_{k+1})\dots(x-x_{n+1})}{(x_k-x_1)(x_k-x_2)\dots(x_k-x_{k-1})(x_k-x_{k+1})\dots(x_k-x_{n+1})}.$$

Теперь легко написать интерполяционный многочлен степени $\leq n$

$$P_n(x) = y_1 l_1(x) + y_2 l_2(x) + \dots + y_{n+1} l_{n+1}(x).$$

Здесь мы используем также, что сумма многочленов степени n является многочленом степени $\leq n$.

3.2.2 Погрешность интерполяции

Интерполяционные многочлены и погрешность приближения такими функциями сильно зависят от выбора узлов интерполяции и, к сожалению, при неудачном выборе узлов интерполяции приближение может оказаться неудовлетворительным. Для оценки погрешности установим следующее равенство

$$R_n(x) = f(x) - P_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}\Pi_{n+1}(x), \tag{3.2}$$

где $c = c(x)$ — некоторая зависящая от x точка на отрезке $[a,b]$,

$$\Pi_{n+1}(x) = (x-x_1)\dots(x-x_{n+1}).$$

Действительно, выберем на отрезке $[a,b]$ произвольную точку x_0, отличную от всех узлов интерполяции. Рассмотрим выражение

$$r(x) = f(x) - P_n(x) - \lambda\Pi_{n+1}(x),$$

где коэффициент λ удовлетворяет условию $r(x_0) = 0$. Поскольку в этой точке $\Pi_{n+1}(x_0) \neq 0$, то такой коэффициент однозначно определяется:

$$f(x_0) - P_n(x_0) = \lambda\Pi_{n+1}(x_0), \quad \lambda = \frac{f(x_0) - P_n(x_0)}{\Pi_{n+1}(x_0)}. \tag{3.3}$$

Подсчитаем теперь λ иным способом. Для этого заметим, что при указанном выборе λ функция $r(x)$ (многочлен степени $n+1$) обращается в нуль в точках x_0, x_1, …, x_{n+1}. По теореме Ролля между каждой парой нулей функции $r(x)$ находится по меньшей мере один нуль ее производной, потому $r'(x)$ имеет на отрезке $[a,b]$ не меньше чем $n+1 = (n+2)-1$ нулей. Аналогично, $r''(x)$ имеет на отрезке $[a,b]$ не меньше чем $n = (n+2)-2$ нулей. И так далее. Мы приходим к тому, что производная $r^{(n+1)}(x)$ имеет не менее $1 = (n+2)-(n+1)$ нулей на $[a,b]$. Итак, существует нуль производной $r^{(n+1)}(x)$ на отрезке $[a,b]$. Обозначим его через c. Учитывая, что степень многочлена P_n не выше n и, следовательно, $P_n^{(n+1)}(c) = 0$, получаем

$$r^{(n+1)}(c) = f^{(n+1)}(c) - \lambda\Pi_{n+1}^{(n+1)}(c) = f^{(n+1)}(c) - \lambda(n+1)!.$$

Отсюда

$$f^{(n+1)}(c) - \lambda(n+1)! = 0, \quad \lambda = \frac{f^{(n+1)}(c)}{(n+1)!}$$

Сравнивая это с (3.3), получаем

$$\frac{f(x_0)-P_n(x_0)}{\Pi_{n+1}(x_0)}=\frac{f^{(n+1)}(c)}{(n+1)!},$$

$$f(x_0)-P_n(x_0)=\frac{f^{(n+1)}(c)}{(n+1)!}\Pi_{n+1}(x_0),$$

что ввиду произвольности выбранной точки z_0 и доказывает формулу (3.2) (точнее, мы выбирали точку x_0 отличной от узлов интерполяции, но в узлах интерполяции равенство очевидно).

Оценка погрешности интерполяции получается из (3.2) и имеет вид

$$\|R_n\|_{C[a,b]}=\|f-P_n\|_{C[a,b]}\le\frac{1}{(n+1)!}\|f^{(n+1)}\|_{C[a,b]}\cdot\|\Pi_{n+1}\|_{C[a,b]},\tag{3.4}$$

при всех $x\in[a,b]$, где

$$\Pi_{n+1}(x)=(x-x_1)\dots(x-x_{n+1}).$$

Таким образом, погрешность интерполяции зависит от качества (гладкости) функции f и от выбора узлов интерполяции. Поскольку изменять свойства функции возможности нет, то для уменьшения погрешности интерполяции у нас остается одно средство — выбирать узлы интерполяции наилучшим образом, так, чтобы норма

$$\|\Pi_{n+1}\|_{C[a,b]}=\max_{x\in[a,b]}|\Pi_{n+1}(x)|$$

была наименьшей. Задача о наилучшем выборе узлов решена П. Л. Чебышевым. Такие узлы распределены на отрезке специальным образом, не равномерно, а сгущаются у его концевых точек.

Пример. Постоить многочлен второй степени, интерполирующий функцию $\sin x$ в точках $x_1=0$, $x_2=\pi/2$, $x_3=\pi$ и оценить погрешность на отрезке $[0,\pi]$.

$$P_2(x)=Ax(x-\pi),\quad P_2(\pi/2)=1.$$

Отсюда $A=-4/\pi^2$ и

$$P_2(x)=-\frac{4}{\pi^2}x(x-\pi).$$

$$\Pi_3(x)=x\left(x-\frac{\pi}{2}\right)(x-\pi).$$

Максимум $|\Pi_3(x)|$ на отрезке $[0,\pi]$ легко вычисляется, он равен $\frac{\pi^3\sqrt{3}}{36}<3/2$. Поэтому

$$|P_2(x)-\sin x|\le\frac{1}{3!}\frac{3}{2}=\frac{1}{4},\quad x\in[0,\pi].$$

3.2.1 Выбор узлов интерполяции. Многочлены Чебышева

1. Мы получили оценку **погрешности интерполяции** в виде (3.4). Отсюда следует, что для уменьшения погрешности интерполяции следует выбирать узлы интерполяции так, чтобы была наименьшей норма $\|\Pi_{n+1}\|_{C[a,b]}$. Задача о таком выборе узлов решена П. Л. Чебышевым. Оказывается, что узлы лучше всего брать в нулях *многочлена Чебышева*. Выпишем формулу для их вычисления

$$x_{k+1}=\frac{b+a}{2}+\frac{b-a}{2}\cos\left(\frac{1+2k}{n+1}\frac{\pi}{2}\right),\quad k=0,\dots,n.\tag{3.5}$$

Такие узлы распределены не равномерно на отрезке, а сгущаются у его концевых точек, и при этом имеет место равенство

$$\|\Pi_{n+1}\|_{C[a,b]} = 2\left(\frac{b-a}{4}\right)^{n+1}. \tag{3.6}$$

Тем самым оценка погрешности (3.4) принимает вид

$$\|R_n\|_{C[a,b]} \le \frac{2}{(n+1)!}\left(\frac{b-a}{4}\right)^{n+1}\|f^{(n+1)}\|_{C[a,b]}. \tag{3.7}$$

Чтобы доказать все сказанное остановимся подробнее на функциях вида

$$T_n(x) = \cos(n\arccos x), \tag{3.8}$$

которые играют важную роль в теории аппроксимаций (приближений) и которые мы назвали многочленами Чебышева степени n. Возникает вопрос, почему функция T_n действительно является многочленом. Докажем это. Введем обозначение $\varphi = \arccos x$. Тогда $\cos\varphi = x$, $\sin\varphi = \sqrt{1-x^2}$. Тогда

$$T_n(x) = \cos(n\arccos x) = \cos(n\varphi) = \frac{e^{in\varphi}+e^{-in\varphi}}{2} =$$

$$\frac{(\cos\varphi + i\sin\varphi)^n + (\cos\varphi - i\sin\varphi)^n}{2} = \frac{(x+i\sqrt{1-x^2})^n + (x-i\sqrt{1-x^2})^n}{2} =$$

$$\frac{1}{2}\left(\sum_{k=0}^{n} C_n^k x^{n-k} i^k (1-x^2)^{k/2} + \sum_{k=0}^{n} C_n^k (-1)^k x^{n-k} i^k (1-x^2)^{k/2}\right).$$

В этих суммах слагаемые с нечетными k сокращаются и остаются лишь слагаемые с четными $k = 2m$, $m = 0, 1, \ldots, N/2$, т.е.

$$T_n(x) = \sum_{m=0}^{N/2} C_n^{2m} x^{n-2m} (-1)^m (1-x^2)^m,$$

где $N = n$ при четных n и $N = n-1$ при нечетных n.

2. Унитарные многочлены Чебышева. Несложно показать, что коэффициент при старшей степени x^n многочлена $T_n(x)$ равен 2^{n-1}. Действительно,

$$\frac{T_n(x)}{x^n} = \frac{(1+i\sqrt{\frac{1}{x}-1})^n + (1-i\sqrt{\frac{1}{x}-1})^n}{2} = \frac{(1-\sqrt{1-\frac{1}{x}})^n + (1+\sqrt{1-\frac{1}{x}})^n}{2} =$$

$$\frac{\frac{1}{x\ \ 1+\sqrt{1-\frac{1}{x}}}^n + (1+\sqrt{1-\frac{1}{x}})^n}{2} \to 2^{n-1}, \quad x \to \infty.$$

Поэтому, положив

$$\tilde{T}_n(x) = \frac{1}{2^{n-1}} T_n(x), \tag{3.9}$$

получим многочлен вида $\tilde{T}_n(x) = x^n + t_{n-1}x^{n-1} + \ldots + t_0$, имеющий единичный старший коэффициент. Такой многочлен будем называть *унитарным многочленом Чебышева*.

3. Рекуррентная формула. Вычисление многочленов $T_n(x)$ можно проводить иначе. Непосредственно проверяем, что

$$T_0(x) = 1, \quad T_1(x) = \cos(\arccos x) = x, \quad T_2(x) = \cos(2\arccos x) = 2x^2 - 1. \tag{3.10}$$

Далее, воспользуемся тождеством

$$\cos((n+1)\varphi) + \cos((n-1)\varphi) = 2\cos(n\varphi)\cos(\varphi),$$

откуда, подставив $\varphi = \arccos x$, получим

$$T_{n+1}(x) + T_{n-1}(x) = 2T_n(x)T_1(x) = 2xT_n(x).$$

Мы получили рекуррентное соотношение

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \tag{3.11}$$

позволяющее вычислять функции $T_n(x)$ последовательно, опираясь на (3.10):

$$T_3(x) = 2xT_2(x) - T_1(x) = 4x^3 - 3x, \ldots$$

Из (3.10) и (3.11), очевидно, следует, что все функции $T_n(x)$ являются многочленами со старшими коэффициентами, равными 2^{n-1}.

4. Нули и точки экстремума многочленов Чебышева. Многочлен $T_n(x)$ имеет на отрезке $[-1, 1]$ имеет n нулей (т.е. максимально возможное число). Вычислим их

$$\cos(n \arccos x) = 0, \quad n \arccos x = \frac{\pi}{2} + \pi k, \quad k = 0, \ldots, n-1.$$

Здесь диапазон изменения k выбран так, чтобы выполнялись неравенства $0 \le n \arccos x \le \pi n$

$$\arccos x = \frac{1+2k}{n}\frac{\pi}{2}, \quad x_k = \cos\left(\frac{1+2k}{n}\frac{\pi}{2}\right), \quad k = 0, \ldots, n-1. \tag{3.12}$$

Нули x_k можно изобразить так. Разобьем полуокружность $|z| = 1$, $\operatorname{Im} z \ge 0$, на $2n$ равных дуг точками $z_1, \ldots, z_{2n-1}$. Тогда нули есть проекции на ось OX точек z_1, z_3, …, z_{2n-1}.

Отметим еще одно важное свойство многочлена $T_n(x)$. Он имеет $n+1$ точек α_k экстремума (максимума и минимума), в которых принимает свои наибольшее и наименьшее значения ± 1 попеременно, начиная с правого конца отрезка $[-1, 1]$. Легко вычислить эти точки экстремума: $\alpha_k = \cos\left(\frac{\pi}{n}k\right)$, $k = 0, \ldots, n$. Для этих точек имеем

$$T_n(\alpha_k) = \cos(n \arccos\left(\cos\left(\frac{\pi}{n}k\right)\right)) = \cos(\pi k) = (-1)^k.$$

5. Стандатные многочлены Чебышева обладают следующим свойством.

Теорема. *Среди всех многочленов $P_n(x)$ со старшим коэффициентом 1 многочлен $\tilde{T}_n(x)$ имеет наименьшую норму на отрезке $[-1, 1]$, т.е. величина*

$$\|\tilde{T}_n\|_{C[-1,1]} = \max_{x\in[-1,1]} |\tilde{T}_n(x)| = \frac{1}{2^{n-1}} \tag{3.13}$$

является наименьшей из возможных. При этом многочлен Чебышева (3.9) *является единственным многочленом указанного вида, удовлетворяющим равенству* (3.13).

Поэтому унитарный многочлен Чебышева называется многочленом, наименее уклоняющимся от нуля.

Доказательство. Предположим противное: существует многочлен $P_n(x) = x^n + b_{n-1}x^{n-1} + \ldots + b_0$, для которого

$$\|P_n\|_{C[-1,1]} < \frac{1}{2^{n-1}},$$

в частности, этот многочлен должен быть отличным от многочлена $\tilde{T}_n(x)$. Далее, многочлен $Q(x) = \tilde{T}_n(x) - P_n(x) = (a_{n-1} - b_{n-1})x^{n-1} + \dots$ имеет степень меньше чем n. С другой стороны, эта разность в точках экстремума α_k имеет тот же знак, что и многочлен $\tilde{T}_n$. Поэтому на каждом отрезке между двумя соседними точками экстремума многочлен $Q(x)$ изменяет знак и, следовательно, имеет по меньшей мере один корень. Таких отрезков всего n, а значит и корней у многочлена $Q(x)$ не менее n. Этого возможно только если $Q(x) \equiv 0$, то есть $P_n(x) \equiv \tilde{T}_n(x)$. Противоречие. Утверждение доказано.

Докажем единственность. Пусть имеется еще один многочлен $Q_n(x) = x^n + \dots + b_0$, удовлетворяющий равенству (3.13). Рассмотрим разность $S(x) = \tilde{T}_n(x) - Q_n(x)$, которая представляет собой многочлен степени $\le n-1$.

Обозначим через δ_k, $k = 1, \dots, n$, отрезки, заключенные между всеми парами соседних точек экстремумов многочлена $\tilde{T}_n$ — т.е. между соседними точками его максимумов и минимумов $\alpha_s = \cos\left(\frac{\pi}{n}s\right)$, $s = 0, \dots, n$. В каждой такой точке, отличной от ± 1, многочлен $S(x)$ либо имеет знак функции $\tilde{T}_n(x)$, либо имеет нуль не ниже второго порядка (в такой точке многочлены $\tilde{T}_n$ и Q_n имеют экстремум одного знака одновременно). Теперь объединим все отрезки δ_k, в пересечении которых находятся нули функции S. Получим новый набор отрезков Δ_k, удовлетворяющий следующим свойствам.

Если Δ_k состоит из единственного не крайнего отрезка δ_l, $l \ne 1$, $l \ne n$, то на концах его знаки $S(x)$ противоположны и, следовательно, $S(x)$ имеет внутри Δ_k хотя бы один нуль. Если Δ_k состоит из единственного крайнего отрезка, скажем, δ_1, то нуль Q находится либо в точке -1, либо, в противном случае (из-за перемены знаков на концах), внутри δ_1.

Если же Δ_k состоит из нескольких отрезков, скажем из $m \ge 2$, то внутри его расположено $2(m-1) \ge m$ нулей (с учетом кратности).

Таким образом мы насчитали нулей многочлена S не меньше чем число отрезков δ_k, т.е. $\ge n$. Это противоречит тому, что степень многочлена S не превосходит $n-1$.

6. Многочлен Чебышева на произвольном отрезке $[a, b]$**.** Такой многочлен получается заменой переменной:

$$x = \frac{2\tilde{x}}{b-a} - \frac{b+a}{b-a}.$$

Когда переменная $\tilde{x}$ пробегает отрезок $[a, b]$ переменная x пробегает отрезок $[-1, 1]$. Положим

$$\tau_n(\tilde{x}) = T_n\left(\frac{2\tilde{x}}{b-a} - \frac{b+a}{b-a}\right) = T_n(x) = \cos(n \arccos x)).$$

Функция $\tau_n(\tilde{x})$ является многочленом степени n. Его старший коэффициент, как легко видеть, равен $\frac{2^{2n-1}}{(b-a)^n}$. Поделив на это число многочлен $\tau_n(\tilde{x})$, получим многочлен с единичным старшим коэффициентом

$$\tilde{\tau}_n(\tilde{x}) = 2\left(\frac{b-a}{4}\right)^n T_n(x).$$

Как и в случае отрезка $[-1, 1]$, можно показать, что этот многочлен наименее уклоняется от нуля на отрезке $[a, b]$. Причем, поскольку $\|T_n\|_{C[-1,1]} = 1$, имеем

$$\|\tilde{\tau}_n\|_{C[a,b]} = 2\left(\frac{b-a}{4}\right)^n \|T_n\|_{C[-1,1]} = 2\left(\frac{b-a}{4}\right)^n.$$

Найдем корни уравнения $\tilde{\tau}_n(\tilde{x}) = 0$, т.е. $T_n(x) = 0$. Приравняв аргумент x к выражениям (3.12), получим

$$x_k = \frac{2\tilde{x}_k}{b-a} - \frac{b+a}{b-a} = \cos\left(\frac{1+2k}{n}\frac{\pi}{2}\right),$$

$$\tilde{x}_k = \frac{b+a}{2} + \frac{b-a}{2}\cos\left(\frac{1+2k}{n}\frac{\pi}{2}\right) \quad k = 0, \dots, n-1.$$

Отсюда, заменив n на $n+1$, получим формулу (3.5) для узлов $\tilde{x}_k$ Чебышева на отрезке $[a,b]$. Итак, если $\Pi_{n+1} = \tilde{\tau}_{n+1}$, то получается указанные в (3.6) и (3.7) оценки для погрешности интерполяции.

3.3 Метод наименьших квадратов

Пусть в попарно различных узлах $x_1, \dots x_N$ заданы значения некоторой функции $y_k = f(x_k)$. Требуется найти многочлен

$$P_n(x) = p_n x^n + p_{n-1}x^{n-1} + \dots + p_1 x + p_0$$

степени n, такой что сумма

$$F_N = F_N(p_0, \dots p_n) = \sum_{k=1}^{N} \left(P(x_k) - y_k\right)^2$$

минимальна. Другими словами, при фиксированных узлах x_k и значениях y_k функции f требуется минимизировать функцию $F_N(p_0, \dots p_n)$ по переменным $p_0, \dots p_n$. Если $N \le n+1$, то этот минимум равен нулю, достаточно построить интерполяционный многочлен, совпадающий с функцией во всех узлах. При $N \ge n+2$ этот минимум уже не обязательно равен нулю.

Пусть $N \ge n+2$. Применим необходимый признак экстремума: в точке локального экстремума все частные производные функции $F_N(p_0, \dots p_n)$ должны обращаться в нуль. Запишем сумму в развернутом виде

$$F_N(p_0, \dots p_n) = \sum_{k=1}^{N} \left(p_0 + p_1 x_k + \dots + p_{n-1}x_k^{n-1} + p_n x_k^n - y_k\right)^2, \tag{3.14}$$

и вычислим частные производные

$$\frac{\partial F_N}{\partial p_r} = 2\sum_{k=1}^{N} x_k^r \left(p_0 + p_1 x_k + \dots + p_{n-1}x_k^{n-1} + p_n x_k^n - y_k\right) =$$

$$= 2\sum_{k=1}^{N}\left(\sum_{m=0}^{n}\left(p_m x_k^{m+r}\right) - y_k x_k^r\right) = 2\left(\sum_{m=0}^{n} p_m S_{m+r} - \sum_{k=1}^{N} y_k x_k^r\right), \quad r = 0, \dots, n,$$

где

$$S_l = x_1^l + x_2^l + \dots + x_N^l = \sum_{k=1}^{N} x_k^l, \qquad S_0 = N,$$

есть степенные суммы порядка $r = 0, 1, \dots$ узлов. Получена следующая система уравнений, линейная относительно неизвестных p_k

$$\sum_{m=0}^{n} p_m S_{m+r} = S_r^{(y)}, \qquad S_r^{(y)} := \sum_{k=1}^{N} y_k x_k^r, \quad r = 0, \dots, n.$$

В развернутом виде она имеет вид

$$\begin{cases} S_0 p_0 + & S_1 p_1 + & S_2 p_2 + & \dots + & S_n p_n = & S_0^{(y)} \\ S_1 p_0 + & S_2 p_1 + & S_3 p_2 + & \dots + & S_{n+1} p_n = & S_1^{(y)} \\ S_2 p_0 + & S_3 p_1 + & S_4 p_2 + & \dots + & S_{n+2} p_n = & S_2^{(y)} \\ \dots & \dots & \dots & \dots & \dots & \\ S_n p_0 + & S_{n+1} p_1 + & S_{n+2} p_2 + & \dots + & S_{2n} p_n = & S_n^{(y)} \end{cases} \tag{3.15}$$

Если определитель матрицы этой системы отличен от нуля (а это всегда так, см. примечание 2 ниже), т.е.

$$\Delta = \begin{vmatrix} S_0 & S_1 & S_2 & \dots & S_n \\ S_1 & S_2 & S_3 & \dots & S_{n+1} \\ S_2 & S_3 & S_4 & \dots & S_{n+2} \\ \dots & \dots & \dots & \dots & \dots \\ S_n & S_{n+1} & S_{n+2} & \dots & S_{2n} \end{vmatrix} \neq 0,$$

то система имеет единственное решение $(p_0^*, \dots, p_n^*)$. В точке $(p_0^*, \dots, p_n^*)$ сумма $F_N(p_0, \dots p_n)$ принимает свое наименьшее значение. Интуитивно это ясно: функция $F_N(p_0, \dots p_n)$ должна расти к бесконечности при $\sqrt{p_0^2 + \dots + p_n^2} \to \infty$ и, следовательно, в конечной точке экстремума может иметь только минимум. Итак, задача о нахождении коэффициентов экстремального многочлена $P_n(x)$ решена.

Примечание 1. Покажем, что $F_N(p_0, \dots p_n) \to \infty$ при $\sqrt{p_0^2 + \dots + p_n^2} \to \infty$. Отсюда будет следовать, в частности, что система (3.15) всегда имеет решения, а если решение $(p_0^*, \dots, p_n^*)$ единственно, то оно является точкой минимума функции F_N. Напомним, что $N \geq n+1$ и точки $x_1, \dots x_N$ попарно различны. Поскольку при различных узлах $x_1, \dots, x_{n+1}$ однородная система линейных уравнений

$$V_k(p_0, \dots p_n) := p_0 + p_1 x_k + \dots + p_{n-1} x_k^{n-1} + p_n x_k^n = 0, \qquad k = 1, \dots, n+1,$$

имеет только нулевое решение $p_k = 0$ (определитель Вандермонда отличен от нуля!), то в каждой точке сферы

$$G(1): \ p_0^2 + \dots + p_n^2 = 1$$

хотя бы одна из величин $V_k(p_0, \dots, p_n)$, $k = 1, \dots, n+1$, отлична от нуля. Непрерывная функция $V(p_0, \dots, p_n) = \min_{k=1,\dots,n+1} |V_k(p_0, \dots, p_n)|$ достигает в некоторой точке сферы $G(1)$ своего минимума μ, и из предыдущего следует, что этот минимум положителен $\mu > 0$. Это означает (в силу равенства $V_k(\lambda p_0, \dots, \lambda p_n) = \lambda V_k(p_0, \dots, p_n)$), что на сфере

$$G(r): \ p_0^2 + \dots + p_n^2 = r^2$$

имеем $V(p_0, \dots, p_n) \geq \mu r > 0$. Положим $Y := \max\{|y_1|, \dots, |y_{n+1}|\}$. Выберем сколь угодно большое $C > 1$. При достаточно больших $r = r(C, Y)$ имеем $\mu r > C + Y$. Следовательно, хотя бы для одного слагаемого с некоторым номером k, $1 \leq k \leq n+1$, в сумме (3.14) имеем

$$(V_k(p_0, \dots, p_n) - y_k)^2 \geq (\mu r - Y)^2 > C^2, \qquad (p_0, \dots, p_n) \in G(r).$$

Последнее выражение можно сделать сколь угодно большим. Поэтому сумма (3.14) неограниченно растет с ростом r. Что и требовалось.

Примечание 2. Из примечания 1, в частности, следует, что определитель Δ всегда отличен от нуля. Действительно, допустим противное: $\Delta = 0$. Тогда один из столбцов матрицы системы (3.15), есть линейная комбинация n других столбцов. Отсюда в силу совместности системы (3.15) (по примечанию 1) столбец свободных членов (при любом наборе y_k, $k = 1, \dots, N$) также есть линейная комбинация тех же столбцов. Другими словами пространство всевозможных столбцов свободных членов имеет размерность $\leq n$. Но, с другой стороны, столбец справа из элементов $S_r^{(y)} = \sum_{k=1}^N y_k x_k^r$ заполняет все $n+1$-мерное пространство $\alpha_0, \dots, \alpha_n$, поскольку система уравнений $\sum_{k=1}^N y_k x_k^r = \alpha_r$, $r = 0, \dots, N-1$, для любого столбца $(\alpha_0, \dots, \alpha_n, 0, \dots, 0)^*$ (из N элементов), всегда имеет (и притом единственное) решение $y_1, \dots, y_N$, т.к. ее матрица имеет определитель Вандермонда. Получили противоречие.

Пример. Построить многочлен второй степени по таблице

$$X = [x_1 = -1,\ x_2 = 0,\ x_3 = 1,\ x_4 = 2];$$
$$Y = [y_1 = 7,\ y_2 = 0,\ y_3 = 5,\ y_4 = 2].$$

Здесь $N = 4$, $n = 2$. Получаем систему уравнений (3.15):

$$4\,p_0 + 2\,p_1 + 6\,p_2 = 14;$$
$$2\,p_0 + 6\,p_1 + 8\,p_2 = 2;$$
$$6\,p_0 + 8\,p_1 + 18\,p_2 = 20.$$

Отсюда находим $p_2 = 1$, $p_1 = -2$, $p_0 = 3$; $P_2(x) = x^2 - 2x + 3$.

4 Рациональная аппроксимация. Цепные дроби

4.1 Аппроксимация Паде

Пусть задана аналитическая в окрестности начала координат функция

$$f(x) = f_0 + f_1 x + f_2 x^2 + \ldots + f_n x^n + \ldots$$

Требуется найти рациональную функцию

$$R_{M,N}(x) = \frac{P_M(x)}{Q_N(x)} = \frac{p_0 + p_1 x + p_2 x^2 + \ldots + p_M x^M}{1 + q_1 x + q_2 x^2 + \ldots + q_N \cdot x^N}$$

со свойством

$$f(x) - R_{M,N}(x) = O(x^{M+N+1}), \tag{4.1}$$

то есть в окрестности начала координат имеет место разложение вида

$$f(x) - R_{M,N}(x) = A_{M+N+1} x^{M+N+1} + A_{M+N+2} x^{M+N+2} + \ldots$$

Другими словами, требуется обнулить коэффициенты разложения разности при степенях $x^0, \ldots, x^{M+N}$.

Рациональная функция $R_{M,N}(x)$, а также и сам процесс ее построения, называются $\left(\frac{M}{N}\right)$-аппроксимацией Паде.

Легко видеть, что разложение произведения fQ_N имеет вид

$$f(x)Q_N(x) = \sum_{k=0}^{\infty} A_k x^k, \qquad A_k = \sum_{j=0}^{\min\{N,k\}} q_j f_{k-j}, \quad q_0 = 1,$$

то есть коэффициенты A_k линейно зависят от параметров $q_1, \ldots q_N$. Тогда

$$f(x)Q_N(x) - P_M(x) =$$

$$= A_0 + A_1 x + \ldots + A_M x^M - (p_0 + p_1 x + p_2 x^2 + \ldots + p_M x^M) +$$
$$+ A_{M+1} x^{M+1} + \ldots A_{M+N} x^{M+N} + O(x^{M+N+1}). \tag{4.2}$$

Далее, подберем коэффициенты $q_1, \ldots q_N$ многочлена $Q(N)$ так, чтобы

$$A_{M+1} x^{M+1} + \ldots + A_{M+N} x^{M+N} \equiv 0.$$

Для этого следует решить относительно коэффициентов $q_1, \ldots q_N$ систему линейных уравнений

$$A_{M+1} = 0, \quad \ldots \quad , A_{M+N} = 0.$$

Предположим, что эта система имеет решение $q_1^*, \ldots q_N^*$. Подставим это решение в (4.2) и для многочлена $Q_N^*(x)$ с такими коэффициентами получим

$$f(x)Q_N^*(x) - P_M(x) =$$

$$= A_0^* + A_1^* x + \ldots + A_M^* x^M - (p_0 + p_1 x + p_2 x^2 + \ldots + p_M x^M) +$$
$$+ 0 + \ldots + 0 + O(x^{M+N+1}).$$

Остается положить

$$p_0^* = A_0^*, \quad p_1^* = A_1^*, \quad \ldots \quad p_M^* = A_M^*,$$

и тогда для многочлена $P_M^*(x)$ с такими коэффициентами будем иметь

$$f(x)Q_N^*(x) - P_M^*(x) = O(x^{M+N+1}). \tag{4.3}$$

Это равенство равносильно равенству (4.1), то есть при делении обеих частей (4.3) на $Q_N^*(x)$ получается (4.1).

Пример. Построить $\left(\frac{2}{2}\right)$-аппроксимацию Паде. В данном случае (4.2) принимает вид

$$f(x)Q_2(x) - P_2(x) =$$

$$= (f_0 + f_1x + f_2x^2 + f_3x^3 + \ldots) \cdot (1 + q_1x + q_2x^2) - (p_0 + p_1x + p_2x^2)$$
$$= f_0 + (f_1 + f_0q_1)x + (f_2 + f_1q_1 + f_0q_2)x^2 - (p_0 + p_1x + p_2x^2) +$$
$$+(f_3 + f_2q_1 + f_1q_2)x^3 + (f_4 + f_3q_1 + f_2q_2)x^4 + O(x^5).$$

Решаем систему уравнений

$$\begin{cases} f_2q_1 & +f_1q_2 & = -f_3 \\ f_3q_1 & +f_2q_2 & = -f_4 \end{cases}$$

Находим решение

$$q_2 = \frac{f_3^2 - f_2f_4}{f_2^2 - f_1f_3}; \qquad q_1 = \frac{f_2f_3 - f_1f_4}{f_2^2 - f_1f_3}.$$

Отсюда получаем

$$p_0 = f_0;$$
$$p_1 = f_1 + f_0q_1 = \frac{f_1f_2^2 - f_3f_1^2 + f_0f_1f_4 - f_0f_3f_2}{f_2^2 - f_1f_3};$$
$$p_2 = f_2 + f_1q_1 + f_0q_2 = \frac{-f_0f_2f_4 + f_0f_3^2 + f_1^2f_4 - 2f_1f_3f_2 + f_2^3}{f_2^2 - f_1f_3}.$$

Примечание. Видим, что задача построения указанным методом не всегда разрешима, именно, при $f_2^2 - f_1f_3 = 0$ система для определения q_1, q_2 может не иметь решений.

Для конкретных функций находим

$$\sin x \approx \frac{x}{1 + x^2/6}, \quad \cos x \approx \frac{1 - \frac{5}{12}x^2}{1 + \frac{1}{12}x^2}, \quad e^x \approx \frac{1 + \frac{1}{2}x + \frac{1}{12}x^2}{1 - \frac{1}{2}x + \frac{1}{12}x^2}.$$

Погрешности этих формул на отрезке $[-1, 1]$ не превосходят соответственно 0.0157, 0.0019, 0.004.

4.2 Цепные числовые дроби

Начнем с примеров. Рассмотрим дробь $\frac{34}{15}$. Проведем преобразование

$$\frac{34}{15} = 2 + \frac{4}{15} = 2 + \frac{1}{\frac{15}{4}} = 2 + \frac{1}{3 + \frac{3}{4}} = 2 + \frac{1}{3 + \frac{1}{\frac{4}{3}}} = 2 + \frac{1}{3 + \frac{1}{1 + \frac{1}{3}}}.$$

Полученное справа выражение называется цепной дробью. Полученное равенство можно кратко записать в виде

$$\frac{34}{15} = [2; 3, 1, 3].$$

Сравним указанную процедуру с алгоритмом Евклида:

$$\begin{cases} 34 = 2 \cdot 15 + 4; & \frac{34}{15} = 2 + \frac{4}{15}; \\ 15 = 3 \cdot 4 + 3; & \frac{15}{4} = 3 + \frac{3}{4}; \\ 4 = 1 \cdot 3 + 1; & \frac{4}{3} = 1 + \frac{1}{4}; \\ 3 = 3 \cdot 1 + 0; & \frac{3}{1} = 3 + 0. \end{cases}$$

Видим, что числа $[2; 3, 1, 3]$ это целые части в правых частях делений с остатком. Вообще, для любого рационального числа по аналогичной схеме: «целая часть + остаток, меньший единицы» можно получить представление в виде конечной цепной дроби

$$R = \frac{P}{Q} = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \ldots + \frac{1}{a_n}}},$$

или, кратко,

$$\frac{P}{Q} = [a_0; a_1, a_2, \ldots, a_n].$$

Эту процедуру можно записать в виде цепочки равенств

$$R = a_0 + r_0; \qquad \frac{1}{r_0} = a_1 + r_1, \qquad \frac{1}{r_1} = a_2 + r_2,$$

и т.д. Аналогично можно находить разложения иррациональных чисел. Ясно, что в этом случае не может получиться конечной цепной дроби. Возьмем, например, число $\sqrt{3}$. Тогда

$$\sqrt{3} = 1 + (\sqrt{3} - 1)$$

$$\frac{1}{\sqrt{3} - 1} \equiv \frac{1}{2}(\sqrt{3} + 1) = 1 + \frac{1}{2}(\sqrt{3} - 1)$$

$$\frac{2}{\sqrt{3} - 1} \equiv (\sqrt{3} + 1) = 2 + (\sqrt{3} - 1)$$

$$\frac{1}{\sqrt{3} - 1} = 1 + \frac{1}{2}(\sqrt{3} - 1).$$

Видим, что началось повторение. Значит, числу $\sqrt{3}$ соответствует цепная дробь $[1; 1, 2, 1, \ldots] = [1; (1, 2)]$.

Выражения вида

$$R_k = [a_0; a_1, a_2, \ldots, a_k]$$

называются k-ми подходящими дробями, $k = 0, 1, \ldots$.

В некоторых случаях удобно считать, что в цепных и подходящих дробях a_k — произвольные числа (а не только целые). Пусть имеется некоторая конечная или бесконечная цепная дробь $[a_0; a_1, a_2, \ldots]$. Получим рекуррентную формулу для вычисления подходящих дробей $R_k = P_k/Q_k$.

Теорема 4.1. *Числители и знаменатели подходящих дробей R_k можно вычислять по следующим рекуррентным формулам*

$$P_k = P_{k-1}a_k + P_{k-2};\ Q_k = Q_{k-1}a_k + Q_{k-2}, \quad k \geq 2, \tag{4.4}$$

где считается, что первые две подходящие дроби записаны в виде

$$R_0 = \frac{P_0}{Q_0}, \quad P_0 = a_0;\ Q_0 = 1;$$

$$R_1 = \frac{P_1}{Q_1}, \quad P_1 = a_0 a_1 + 1;\ Q_1 = a_1;$$

Доказательство. Формулы (4.4) получаются по индукции. Действительно, пусть (4.4) выполнено для всех номеров, которые меньше чем k. Заметим, что R_k является также $(k-1)$-й подходящей дробью $\tilde{R}_{k-1}$, если заменить в R_k выражение $a_{k-1} + \frac{1}{a_k}$ формально на $\tilde{a}_{k-1}$. Тогда будем иметь

$$R_k = \tilde{R}_{k-1}, \qquad P_k = \tilde{P}_{k-1}, \qquad Q_k = \tilde{Q}_{k-1},$$

и по предположению индукции можно записать

$$\tilde{P}_{k-1} = P_{k-2}\,\tilde{a}_{k-1} + P_{k-3}; \qquad \tilde{Q}_{k-1} = Q_{k-2}\tilde{a}_{k-1} + Q_{k-3}.$$

Отсюда

$$R_k = \frac{P_k}{Q_k} = \frac{\tilde{P}_{k-1}}{\tilde{Q}_{k-1}} = \frac{P_{k-2}\left(a_{k-1} + \frac{1}{a_k}\right) + P_{k-3}}{Q_{k-2}\left(a_{k-1} + \frac{1}{a_k}\right) + Q_{k-3}} = \frac{P_{k-2}a_{k-1} + P_{k-3} + P_{k-2}/a_k}{Q_{k-2}a_{k-1} + Q_{k-3} + Q_{k-2}/a_k} =$$

$$= \frac{P_{k-1} + P_{k-2}/a_k}{Q_{k-1} + Q_{k-2}/a_k} = \frac{P_{k-1}a_k + P_{k-2}}{Q_{k-1}a_k + Q_{k-2}}.$$

Что и требовалось.

Отметим одно важное соотношение для подходящих дробей, благодаря которому по существу и применяются цепные дроби в теории чисел. Оказывается, что для двух соседних подходящих дробей $R_k = P_k/Q_k$ и $R_{k-1} = P_{k-1}/Q_{k-1}$ справедливы следующие тождества.

Теорема 4.2.

$$\Delta_k := P_k Q_{k-1} - Q_k P_{k-1} = (-1)^{k-1}, \tag{4.5}$$

и, следовательно,

$$R_k - R_{k-1} = \frac{P_k}{Q_k} - \frac{P_{k-1}}{Q_{k-1}} = (-1)^{k-1}\frac{1}{Q_k Q_{k-1}}. \tag{3}$$

Следствие 1. Все четные подходящие дроби меньше нечетных.

Следствие 2. Числитель и знаменатель подходящей дроби взаимно просты. Если бы у них существовал общий делитель $d > 1$, то, поделив обе части равенства (4.5) на d, мы получили бы слева целое, а справа дробное числа.

Это следствие можно использовать для сокращения исходной дроби R, достаточно разложить ее в цепную дробь

$$R = \frac{P}{Q} = \frac{P_n}{Q_n},$$

а затем обратным порядком из раложения восстановить уже сокращенную дробь.

Пример.

$$\frac{3528}{2058} = [1; 1, 2, 2] = \frac{12}{7}.$$

Доказательство теоремы 4.2. Формула (4.6) легко получается с помощью (4.4). Действительно, учитывая (4.4), получаем

$$\Delta_k = (P_{k-1}a_k - P_{k-2})\,Q_{k-1} - P_{k-1}\,(Q_{k-1}a_k - Q_{k-2}) =$$

$$= -P_{k-2}Q_{k-1} + Q_{k-2}P_{k-1} = -\Delta_{k-1}.$$

Так продолжая, находим

$$\Delta_k = -\Delta_{k-1} = \Delta_{k-2} = \ldots = (-1)^{k-1}\Delta_1,$$

но $\Delta_1 = P_1Q_0 - Q_1P_0 = (a_0a_1 + 1)\cdot 1 - a_1a_0 = 1$. Отсюда и получается (4.6):

$$\frac{P_k}{Q_k} - \frac{P_{k-1}}{Q_{k-1}} = \frac{\Delta_k}{Q_kQ_{k-1}} = (-1)^{k-1}\frac{1}{Q_kQ_{k-1}}.$$

Примечание. Приведем одно применение формулы (4.6) для решения диофантовых уравнений, т.е. уравнений в целых числах. Например, рассмотрим уравнение $mx + ny + 1 = 0$ с целыми числами m, n. Для решения разложим дробь $m/n = [a_0; a_1, a_2, \ldots, a_s]$ и отбросим последнее звено $1/a_s$. Вернувшись от полученной цепной дроби к дроби M/N, получаем $(m/n)-(M/N) = \pm(1/nN)$. Домножим обе части на $\mp nN$ и получим $mN-nM+1 = 0$, т.е. искомое (частное) решение есть $(N, -M)$. Для нахождения всех решений нужно найти еще общее решение однородного уравнения $mx + ny = 0$, что делается несложно, и сложить указанное частное и общее решения.

Пример. Решить в целых числах уравнение $53m+35n+1 = 0$. Находим цепную дробь $53/35 = [1; 1, 1, 1, 17]$. Отбрасываем последнее звено $1/17$ и получаем $M/N = [1; 1, 1, 1] = 3/2$. Вычисляем

$$\frac{53}{35} - \frac{3}{2} = \frac{1}{2\cdot 35};\;\; 53\cdot 2 - 35\cdot 3 = 1;\;\; 53\cdot(-2) + 35\cdot 3 + 1 = 0.$$

Итак, найдено частное решение $(-2, 3)$.

Укажем еще одно важное свойство цепных дробей.

Теорема 4.3.

$$\delta_k := P_kQ_{k-2} - Q_kP_{k-2} = (-1)^k a_k, \tag{4.7}$$

и, следовательно,

$$\frac{P_k}{Q_k} - \frac{P_{k-2}}{Q_{k-2}} = (-1)^k\frac{a_k}{Q_kQ_{k-2}}. \tag{4.8}$$

Следствие 3. Четные дроби монотонно возрастают, а нечетные — убывают и сходятся к основной дроби $R = P/Q$. При этом каждая четная дробь меньше каждой нечетной (см. (4.6)). В частности,

$$\frac{P_{2k}}{Q_{2k}} \le R \le \frac{P_{2k\pm1}}{Q_{2k\pm1}},$$

и, значит,

$$0 \le R - \frac{P_{2k}}{Q_{2k}} \le \frac{P_{2k-1}}{Q_{2k-1}} - \frac{P_{2k}}{Q_{2k}} = \frac{1}{Q_{2k}Q_{2k-1}}, \qquad 0 \le \frac{P_{2k+1}}{Q_{2k+1}} - R \le \frac{1}{Q_{2k}Q_{2k+1}}.$$

Поэтому при любом n (четном и нечетном) имеем оценки погрешности аппроксимации

$$\left|R - \frac{P_n}{Q_n}\right| \le \frac{1}{Q_nQ_{n-1}}.$$

Поскльку знаменатели Q_n возрастают не медленнее чем n (см. (4.4)), то $(n+1)$-е подходящие дроби аппроксимируют основную дробь не хуже чем n^{-2}.

Доказательство теоремы 4.3. По формулам (4.4) и (4.5) получаем

$$\delta_k = P_k Q_{k-2} - Q_k P_{k-2} =$$

$$(P_{k-1}a_k + P_{k-2})Q_{k-2} - (Q_{k-1}a_k + Q_{k-2})P_{k-2} = P_{k-1}a_k Q_{k-2} - Q_{k-1}a_k P_{k-2} =$$
$$= \Delta_{k-1}a_k = (-1)^k a_k.$$

Что и требовалось.

Задача о наилучшем приближении действительных чисел рациональными дробями состоит в следующем. Для данного действительного числа α при каждом натуральном Z требуется найти рациональную дробь m/n, знаменатель которой n не превосходит Z, и для которой разность $\alpha - m/n$ имеет наименьшее значение.

Оказывается, цепные дроби дают наилучшее приближение. Приведем без докахательства теорему об оценке погрешности апроксимации действительных чисел.

Теорема 4.4. *Пусть α — положительное действительное число. Тогда при любом n (четном и нечетном)*

$$\left|\alpha - \frac{P_n}{Q_n}\right| \le \frac{1}{Q_n Q_{n-1}}.$$

Для рациональных чисел утверждение получается из следствия 3.

4.3 Цепные рациональные дроби

Их построение аналогично рассмотренному выше построению числовых цепных дробей. Пусть имеется рацилнальная дробь

$$R(x) = \frac{P(x)}{Q(x)},$$

где P и Q — некоторые многочлены. Запишем $R(x)$ в виде суммы целой части и правильной дроби:

$$R(x) = \frac{P(x)}{Q(x)} = C_0(x) + \frac{P_1(x)}{Q(x)} = C_0(x) + \frac{1}{\frac{Q(x)}{P_1(x)}}.$$

На следущем шаге запишем дробь $\frac{Q(x)}{P_1(x)}$ в виде суммы целой части и правильной дроби:

$$R(x) = \frac{P(x)}{Q(x)} = C_0(x) + \frac{P_1(x)}{Q(x)} = C_0(x) + \frac{1}{\frac{Q(x)}{P_1(x)}} =$$

$$= C_0(x) + \frac{1}{C_1(x) + \frac{Q_1(x)}{P_1(x)}} = C_0(x) + \frac{1}{C_1(x) + \frac{1}{\frac{P_1(x)}{Q_1(x)}}}.$$

И так далее. Этот процесс всегда обрывается на некотором шаге n. Сокращенная запись:

$$\frac{P(x)}{Q(x)} = [C_0(x); C_1(x), \ldots, C_n(x)].$$

Здесь, как и в случае чисел, при построении цепной дроби фактически применяется алгоритм Евклида:

$$\begin{cases} P(x) = C_0(x)\cdot Q(x) + r_0(x); & \frac{P(x)}{Q(x)} = C_0(x) + \frac{r_0(x)}{Q(x)}; \\ Q(x) = C_1(x)\cdot r_0(x) + r_1(x); & \frac{Q(x)}{r_0(x)} = C_1(x) + \frac{r_1(x)}{r_0(x)}; \\ r_0(x) = C_2(x)\cdot r_1(x) + r_2(x); & \frac{r_0(x)}{r_1(x)} = C_2(x) + \frac{r_2(x)}{r_1(x)}; \\ \dots & \dots . \end{cases}$$

Здесь степени остатков $r_k(x)$ убывают и поэтому на некотором шаге n получим деление без остатка, т.е. $r_{n-2}(x) = C_n(x)\cdot r_{n-1}(x) + 0$.

Пример.

$$R(x) = \frac{x^4 + x^3 + 2\,x^2 + 4\,x + 1}{x^3 + x^2 + 2\,x + 3} = [x; x^2 + 2, x + 1].$$

4.4 Задача интерполяции рациональными функциями

Пусть задана интерполяционная таблица (x_k, y_k) некоторой функции f, $y_k = f(x_k)$, $k = 1, \dots, n$. Будем искать интерполяционную рациональную функцию $R_{n-1}(x)$ в виде цепной дроби

$$R_{n-1}(x) = A_0 + \frac{x - x_1}{A_1 + \frac{x-x_2}{A_2 + \dots + \frac{x-x_{n-1}}{A_{n-1}}}} \tag{4.9}$$

Отметим, что сумма степеней числителя и знаменателя рациональной функции $R_{n-1}(x)$ равна $n-1$. Например,

$$R_0 = \frac{A_0}{1}, \quad R_1 = \frac{x + A_0A_1 - x_1}{A_1},$$

$$R_2 = \frac{(A_0 + A_2)\,x + A_0A_1A_2 - A_0x_2 - A_2x_1}{x + A_1A_2 - x_2}.$$

Параметры $A_0, \dots, A_{n-1}$ определяются из условия интерполяции $y_k = R_{n-1}(x_k)$ $(k = 1, \dots, n)$ последовательно. Например, при подстановке $x = x_1$ в (4.9) получаем $A_0 = y_1$; при подстановке $x = x_2$, а затем $x = x_3$ в (4.9) получаем

$$y_2 = R_{n-1}(x_2) = R_1(x_2) = A_0 + \frac{x_2 - x_1}{A_1 + 0},$$

$$y_3 = R_{n-1}(x_3) = R_2(x_3) = A_0 + \frac{x_3 - x_1}{A_1 + \frac{x_3 - x_2}{A_2 + 0}}$$

откуда находим A_1, а затем A_2. И так далее.

Можно применять следующие рекуррентные формулы для вычисления подходящих дробей $R_k(x) = \frac{P_k(x)}{Q_k(x)}$:

$$P_0 = A_0,\ Q_0 = 1; \qquad P_1(x) = x + A_0A_1 - x_1,\ Q_1(x) = A_1;$$

и далее

$$P_k(x) = P_{k-1}(x)\,A_k + P_{k-2}(x)(x - x_k),$$
$$Q_k(x) = Q_{k-1}(x)\,A_k + Q_{k-2}(x)(x - x_k),$$

где $k = 2, \ldots, n-1$. Эти формулы напоминают аналогичные формулы (4.4) для числовых цепных дробей. Например,

$$P_2(x) = (x + A_0A_1 - x_1)\, A_2 + A_0\,(x - x_2)\,, \quad Q_2 = x + A_1A_2 - x_2.$$

Отметим одно важное свойство подходящих дробей:

$$R_k(x) - R_{k-1}(x) = \frac{(x-x_1)(x-x_2)\cdots(x-x_k)}{Q_{k-1}(x)Q_k(x)}, \quad k = 1, \ldots, n-1. \tag{4.10}$$

Примечание. Пусть все $A_k > \omega + \delta$, где $\omega > 1$, а δ — длина отрезка интерполяции. Тогда многочлены $Q_k(x)$ положительны и возрастают со скоростью геометрической прогресии во всех точках этого отрезка. Точнее, $Q_k(x) \geq \omega^k$.

Докажем это по индукции. Для $Q_0 = 1$ и $Q_1(x) = A_1$ это очевидно. Допустим утверждение верно при номерах $\leq k-1$. Тогда для номера k имеем

$$Q_k(x) \geq Q_{k-1}(x)\, A_k - Q_{k-2}(x)|x - x_k| > Q_{k-1}(x)\,(\omega + \delta) - Q_{k-2}(x)\delta >$$

$$> Q_{k-1}(x)\,(\omega + \delta) - Q_{k-1}(x)\delta = \omega Q_{k-1}(x) \geq \omega^k.$$

Так что в этом случае равенство (4.10) можно применять для оценки погрешности.

5 Ортогональные системы в эвклидовом пространстве

5.1 Общие понятия

Напомним, что эвклидовым пространством называют линейное пространство E, в котором задано скалярное произведение (f, g), $f, g \in E$. Скалярное произведение это функционал, удовлетворяющий свойствам

$$\begin{array}{ll} \text{I.} & (f,g) = (g,f); \\ \text{II.} & (\lambda f, g) = \lambda (f,g), \quad \lambda \in \mathbb{R}; \\ \text{III.} & (f, g_1 + g_2) = (f,g_1) + (f,g_2); \\ \text{IV.} & (f,f) \geq 0, \quad (f,f) = 0 \quad \Leftrightarrow \quad f = 0. \end{array}$$

В E можно ввести норму (длину вектора) по формуле

$$\|f\| = \sqrt{(f,f)}. \tag{5.1}$$

Ниболее простыми примерами эвклидовых пространств служат n-мерные пространства со скалярным произведением $(f,g) = \rho_1 x_1 y_1 + \ldots + \rho_n x_n y_n$, где $f = (x_1, x_2, \ldots, x_n)$, $g = (y_1, y_2, \ldots, y_n)$ — векторы с действительными координатами, а $(\rho_1, \rho_2, \ldots, \rho_n)$ — фиксированный вектор (вес) с положительными координатами.

Пространство L_2. Пространство $L_2 = L_2([a,b], \rho)$ это множество кусочно-непрерывных на промежутке $[a,b]$ функций со скалярным произведением

$$(f,g) = \int_a^b f(x) g(x) \rho(x)\, dx,$$

где $\rho(x)$ заданная кусочно-непрерывная на промежутке $[a,b]$ положительная функция, которая называется весом. Норму в пространстве L_2 согласно (5.1) можно определить так:

$$\|f\|_2 = \sqrt{(f,f)} = \left(\int_a^b f^2(x) \rho(x)\, dx \right)^{1/2}.$$

Промежуток $[a,b]$ может быть ограниченным отрезком или неограниченным промежутком типа $[a, \infty]$, $[\infty, b]$, $[-\infty, +\infty]$.

5.2 Ортонормированные системы векторов. Полнота систем

В эвклидовых пространствах E важную роль играют ортонормированные конечные или бесконечные системы векторов.

Определение. Система векторов $F = \{f_1, f_2, \ldots\}$ называется ортогональной и нормированной (ортонормированной), если

$$(f_m, f_k) = \delta_m^k.$$

где δ_m^k — символ Кронекера. Отметим, что любая конечная подсистема $F_n = \{f_1, f_2, \ldots f_n\}$ ортонормированной системы F линейно независима. Действительно, если

$$\alpha_1 f_1 + \ldots + \alpha_k f_k + \ldots + \alpha_n f_n = 0,$$

то умножив это равенство скалярно на f_k, получим

$$\alpha_1 \cdot 0 + \ldots + \alpha_k \cdot 1 + \ldots + \alpha_n \cdot 0 = 0,$$

откуда $\alpha_k = 0$ при всех $k = 1, \ldots, n$.

Определение. Система векторов $G = \{g_1, g_2, \ldots\}$ называется полной в соответствующем эвклидовом пространстве E, если любой элемент $f \in E$ можно сколь угодно точно аппроксимировать линейными комбинациями векторов из G. То есть для любого ε существует конечный набор чисел α_k, $k = 1, \ldots, N = N(f, \varepsilon)$, для которых

$$\|f - \sum_{i=1}^{N} \alpha_k g_k\| \leq \varepsilon. \tag{5.2}$$

5.3 Ряды Фурье по ортонормированной системе

Пусть вектор $f \in E$, E — эвклидово пространство. Рядом Фурье вектора f по конечной или бесконечной ортонормированной системе $G = \{g_1, g_2, \ldots\}$ называется $\sum_k (f, g_k) g_k$. При этом пишут

$$f \sim \sum_k (f, g_k) g_k, \tag{5.3}$$

где знак «$\sim$» означает соответствие ряда (5.3) элементу f. Возникает вопрос о сходимости: когда знак «$\sim$» в (5.3) можно заменить знаком равенства для любого элемента $f \in E$. Если пространство n-мерно (конечномерно), то для ортонормированной системы $G = \{g_1, \ldots, g_n\}$ этого пространства равенство выполнено всегда:

$$f = \sum_{k=1}^{n} (f, g_k) g_k,$$

Ниже будет показано, что если бесконечная ортонормированная система $G = \{g_1, g_2, \ldots\}$ является полной, то и в этом случае вместо (5.3) имеем

$$f = \sum_{k=1}^{\infty} (f, g_k) g_k, \quad \text{то есть} \quad \left\| f - \sum_{k=1}^{N} (f, g_k) g_k \right\| \to 0, \quad N \to \infty. \tag{5.4}$$

для любого элемента $f \in E$.

Экстремальное свойство частичных сумм ряда (5.3). Оказывается, что частичные суммы ряда (5.3) аппроксимируют вектор f лучше, чем любая другая линейная комбинация входящих в эту сумму векторов. Другими словами, имеет место

Теорема. *Для любого набора чисел α_k, $k = 1, \ldots, N$, имеем*

$$\left\| f - \sum_{k=1}^{N} (f, g_k) g_k \right\| \leq \left\| f - \sum_{k=1}^{N} \alpha_k g_k \right\|. \tag{5.5}$$

Действительно, положим $\overset{*}{\alpha}_k = (f, g_k)$. Тогда, пользуясь свойствами скалярного произведения и ортонормированностью векторов g_k, получим

$$A := \left\| f - \sum_{k=1}^{N} \alpha_k g_k \right\|^2 = \left(f - \sum_{k=1}^{N} \alpha_k g_k, f - \sum_{k=1}^{N} \alpha_k g_k \right) = \|f\|^2 + \sum_{k=1}^{N} \left(\alpha_k^2 - 2\alpha_k \overset{*}{\alpha}_k \right).$$

В частности, если все $\alpha_k = \overset{*}{\alpha}_k$, то

$$B := \left\| f - \sum_{k=1}^{N} \overset{*}{\alpha}_k g_k \right\|^2 = \|f\|^2 - \sum_{k=1}^{N} \overset{*}{\alpha}_k^2, \quad \text{где} \quad \overset{*}{\alpha}_k = (f, g_k). \tag{5.6}$$

Нам нужно доказать, что $B \le A$. Имеем

$$A = \|f\|^2 + \sum_{k=1}^{N} \left(\alpha_k^2 - 2\alpha_k \overset{*}{\alpha}_k\right) = \|f\|^2 + \sum_{k=1}^{N} \left(\alpha_k - \overset{*}{\alpha}_k\right)^2 - \sum_{k=1}^{N} \overset{*}{\alpha}_k^{\,2} \ge \|f\|^2 - \sum_{k=1}^{N} \overset{*}{\alpha}_k^{\,2} = B,$$

Что и требовалось.

Следствие. *Если ортонормированная система векторов $G = \{g_1, g_2, \ldots\}$ является полной, ряд Фурье сходится к $f \in E$, то есть выполнено (5.4).*

Действительно, из полноты системы G следует, что при заданном сколь угодно малом $\varepsilon > 0$ существует конечный набор чисел α_k, $k = 1, \ldots, N = N(f, \varepsilon)$, для которых выполнено (5.2). Отсюда и из (5.5) получаем

$$\left\| f - \sum_{k=1}^{N} (f, g_k) g_k \right\| \le \|f - \sum_{k=1}^{N} \alpha_k g_k\| \le \varepsilon.$$

Это и означает, что выполнено равенство (5.4).

Устремляя ε к нулю (а N к бесконечности), отсюда и из (5.6) получаем,

$$\|f\|^2 - \sum_{k=1}^{\infty} (f, g_k)^2 = \left\| f - \sum_{k=1}^{\infty} (f, g_k) g_k \right\| = 0$$

и, следовательно, выполняется *равенство Парсеваля*

$$\|f\|^2 = \sum_{k=1}^{\infty} (f, g_k)^2.$$

5.4 Метод ортогонализации Грама — Шмидта

Покажем, что из любой конечной совокупности линейно независимых векторов $F = \{f_1, f_2, \ldots, f_n\}$ путем линейных преобразований можно получить ортонормированную (ортогональную и нормированную) систему векторов $G = \{g_1, g_2, \ldots, g_n\}$ так, что каждый вектор системы F линейно выражается через векторы системы G и наоборот. Построение ортонормированной системы G проводится методом Грама — Шмидта. Опишем его.

Элементы g_k строятся последовательно.

На первом шаге полагаем $g_1 = \frac{f_1}{\|f_1\|}$. Поскольку линейно независимая система не может содержать нулевые векторы, то $\|f_1\| \ne 0$ и, следовательно, $\|g_1\| = 1$.

На втором шаге полагаем

$$\tilde{g}_2 = f_2 - (f_2, g_1) g_1, \quad g_2 = \frac{\tilde{g}_2}{\|\tilde{g}_2\|}.$$

Здесь деление на норму законно, поскольку $\tilde{g}_2 \ne 0$ как нетривиальная линейная комбинация независимых векторов f_2 и g_1. Непосредственно проверяется, что $(g_2, g_1) = 0$, $\|g_2\| = 1$. На третьем шаге полагаем

$$\tilde{g}_3 = f_3 - (f_3, g_1) g_1 - (f_3, g_2) g_2, \quad g_3 = \frac{\tilde{g}_3}{\|\tilde{g}_3\|}.$$

Здесь снова деление на норму законно. Действительно, из построения видно, что векторы g_1, g_2 лежат в линейном подпространстве, образованном векторами f_1, f_2 (то есть каждый из векторов g_1, g_2 линейно выражается через f_1 и f_2). Поэтому $\tilde{g}_3 \ne 0$ как нетривиальная

линейная комбинация независимых векторов f_3 и f_1, f_2. Повторяя указанную процедуру, на шаге k получим формулы

$$\tilde{g}_k = f_k - \sum_{i=1}^{k-1}(f_k, g_i)g_i, \quad g_k = \frac{\tilde{g}_k}{\|\tilde{g}_k\|}. \tag{5.7}$$

Здесь мы снова пользуемся линейной независимостью первоначальной системы F, благодаря которой все $\|\tilde{g}_k\| \neq 0$. Действительно, как легко видеть из построения, векторы $g_1, \ldots, g_{k-1}$ лежат в линейном подпространстве, образованном векторами $f_1 \ldots, f_{k-1}$. Поэтому вектор $\tilde{g}_k$ отличен от нуля как нетривиальная линейная комбинация линейно независимых векторов f_k и $f_1, \ldots, f_{k-1}$. Из построения видим, что все свойства новой системы G, отмеченные выше, выполнены.

Примечание (для запоминания). По формуле (5.7) каждый элемент g_k получается нормировкой разности элемента f_k и его частичной суммы Фурье по уже построенным ортонормированным векторам $g_1, \ldots, g_{k-1}$.

5.5 Ортонормированная система многочленов

Рассмотрим пространство $L_2([a,b],\rho)$ и в этом пространстве линейно независимую систему $f_0 = 1$, $f_k = x^k$, $k = 1, \ldots, n$. Здесь по-прежнему возможны и неограниченные промежутки, например, $a = -\infty$, $b = \infty$. Если провести в этой системе процесс ортогонализации Грама — Шмидта, то получим ортонормированные с весом ρ многочлены

$$\{P_0(x), P_1(x), \ldots, P_n(x)\} \tag{5.8}$$

(индекс указавает степень многочлена). То есть

$$(P_m, P_k) = \int_a^b P_m(x)P_k(x)\rho(x)\,dx = \delta_m^k, \quad m,k = 0, \ldots n. \tag{9}$$

В данном случае процесс ортогонализации Грама — Шмидта выполняется по формулам

$$P_0(x) = P_0 = \frac{1}{\sqrt{\int_a^b \rho(x)dx}}, \tag{5.10}$$

$$\tilde{P}_k(x) = x^k - \sum_{i=1}^{k-1}\left(\int_a^b x^k P_i(x)\rho(x)dx\right)P_i(x), \;\; P_k(x) = \frac{\tilde{P}_k(x)}{\sqrt{\int_a^b \tilde{P}_k^2(x)\rho(x)dx}}. \tag{5.11}$$

Отметим, что старшие коэффициенты μ_k ортонормированных многочленов

$$P_k(x) = \mu_k x^k + a_{k-1}^{(k)}x^{k-1} + \ldots + a_0^{(k)}$$

отличны от нуля, это сразу видно из формул (5.11). Отметим еще, что (5.10), (5.11) определяют совокупность ортогональных многочленов с точностью до знака, т.е. вместо системы многочленов $\{P_0(x), P_1(x), \ldots, P_n(x)\}$ мы можем рассматривать $\{\pm P_0(x), \pm P_1(x), \ldots, \pm P_n(x)\}$ с произвольной расстановкой знаков, такие системы также будут ортонормированными. Несложно показать, что других ортогональных многочленов нет. Если же знаки многочленов выбрать так, чтобы старшие коэффициенты μ_k были положительны, то система ортонормированных многочленов определяется однозначно.

5.6 Задача о наилучшем приближении ортонормированными многочленами

Задача ставится следующим образом. Пусть задана функция $f(x)$ и система ортонормированных в $L_2([a,b],\rho)$ многочленов $\{P_0(x), P_1(x), \ldots, P_n(x)\}$. Требуется найти числа $\overset{*}{\alpha}_0, \ldots, \overset{*}{\alpha}_n$ такие, что

$$e_n(f) = \|f - (\overset{*}{\alpha}_0 P_0 + \overset{*}{\alpha}_1 P_1 + \ldots + \overset{*}{\alpha}_n P_n)\|_2 =$$

$$= \sqrt{\int_a^b \left(f(x) - (\overset{*}{\alpha}_0 P_0(x) + \overset{*}{\alpha}_1 P_1(x) + \ldots + \overset{*}{\alpha}_n P_n(x))\right)^2 \rho(x)\,dx} \tag{5.12}$$

имеет наименьшее значение. Из экстремального свойства (5.5) частичных сумм ряда Фурье следует, что задачу решают коэффициенты Фурье:

$$\overset{*}{\alpha}_k = (f, P_k) = \int_a^b f(x) P_k(x) \rho(x)\,dx.$$

При этом наилучшее приближение вычисляется по формуле (5.6):

$$e_n(f) = \sqrt{\|f\|_2^2 - \sum_{k=0}^{n} \overset{*}{\alpha}_k^{\,2}}. \tag{5.13}$$

Пример 1. Ортогонализировать и нормировать в $L_2 = L_2([0,1],\rho)$ с весом $\rho(x) = 1$ систему функций $F = \{1,\, x,\, x^2\}$. Найти наилучшее приближение в L_2 функции $f(x) = x^3$.

Применим формулы (5.10), (5.11).

$$\tilde{g}_0 = f_0 = 1,\ g_0 = \frac{\tilde{g}_0}{\|\tilde{g}_0\|_2} = \frac{1}{1} = 1.$$

$$\tilde{g}_1 = f_1 - (f_1, g_0) g_0 = x - \left(\int_0^1 x \cdot 1\,dx\right) \cdot 1 = x - \frac{1}{2},$$

$$g_1 = \frac{\tilde{g}_1}{\|\tilde{g}_1\|_2} = \frac{x - \frac{1}{2}}{\left(\int_0^1 \left(x - \frac{1}{2}\right)^2 dx\right)^{\frac{1}{2}}} = \sqrt{12}\left(x - \frac{1}{2}\right).$$

Аналогично,

$$\tilde{g}_2 = f_2 - (f_2, g_0) g_0 - (f_2, g_1) g_1 =$$

$$= x^2 - \left(\int_0^1 x^2 \cdot 1\,dx\right) \cdot 1 - 12 \left(\int_0^1 x^2 \cdot \left(x - \frac{1}{2}\right) dx\right) \cdot \left(x - \frac{1}{2}\right) = x^2 - x + \frac{1}{6},$$

$$g_2 = 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right).$$

Для решения второй части воспользуемся экстремальным свойством (5.5) частичной суммы Фурье. Вычислим коэффициенты Фурье функции $f = x^3$ относительно системы g_k:

$$\overset{*}{\alpha}_k = \int_0^1 x^3 g_k(x)\,dx, \quad k = 0, 1, 2, \quad \overset{*}{\alpha}_0 = \frac{1}{4}, \quad \overset{*}{\alpha}_1 = \frac{3}{20}\sqrt{3}, \quad \overset{*}{\alpha}_2 = \frac{\sqrt{5}}{20}.$$

Вычисляем наилучшее приближение по формуле (5.13)

$$\left\| f - \sum_{k=0}^{2} \overset{*}{\alpha}_k g_k \right\|_2 = \sqrt{\|f\|_2^2 - \sum_{k=0}^{2} \overset{*}{\alpha}_k^{\,2}} = \sqrt{\int_0^1 (x^3)^2 \, dx - \sum_{i=0}^{2} \overset{*}{\alpha}_i^{\,2}} = \sqrt{\frac{1}{2800}} = \frac{1}{140}\sqrt{7}.$$

Пример 2. Ортогонализировать и нормировать в $L_2([0,\infty],\rho)$ с весом $\rho(x) = e^{-x}$ систему функций $A = \{1,\, x,\, x^2, x^3\}$.

Ответ:

$$g_0 = 1, \quad g_1 = x - 1, \quad g_2 = \frac{1}{2}\left(x^2 - 4\,x + 2\right), \quad g_3 = \frac{1}{6}\left(x^3 + 18\,x - 9\,x^2 + 18\,x - 6\right).$$

5.7 Метод наименьших квадратов в L_2

Задача ставится следующим образом. Пусть задана функция $f(x)$ и система произвольных линейно независимых (и не обязательно ортогональных) многочленов $\{P_0, \ldots, P_n\}$. Требуется найти числа $\tilde{\alpha}_0, \ldots, \tilde{\alpha}_n$ такую, что линейная комбинация

$$\tilde{\alpha}_0 P_0 + \tilde{\alpha}_1 P_1 + \ldots + \tilde{\alpha}_n P_n \tag{5.14}$$

наилучшим образом приближает в L_2 функцию f, то есть величина

$$e_n(f) = \|f - (\tilde{\alpha}_0 P_0 + \tilde{\alpha}_1 P_1 + \ldots + \tilde{\alpha}_n P_n)\|_2,$$

определенная как в (5.12), имеет наименьшие значение. В этом случае мы не можем применять методы, связанные с экстремальностью (5.5) сумм Фурье по ортонормированной системе.

Для применения метода Фурье надо сначала по данной системе многочленов построить ортонормированную систему, затем вычислить сумму Фурье; она и будет искомой линейной комбинацией.

Но в некоторых случаях числа $\tilde{\alpha}_0, \ldots, \tilde{\alpha}_n$ выгоднее определить, воспользовавшись методом наименьших квадратов. Положим

$$F(\alpha_0, \ldots, \alpha_n) = \int_a^b (f(x) - \alpha_0 P_0(x) - \alpha_1 P_1(x) - \ldots - \alpha_n P_n(x))^2 \rho(x)\, dx. \tag{5.15}$$

Находим частные производные

$$\frac{\partial F}{\partial \alpha_k} = -2 \int_a^b (f(x) - \alpha_0 P_0(x) - \alpha_1 P_1(x) - \ldots - \alpha_n P_n(x)) P_k(x) \rho(x)\, dx =$$

$$= -2(P_k, f) + 2 \sum_{i=0}^{n} \alpha_i \cdot (P_k, P_i), \quad k = 0, \ldots, n.$$

Приравнивая их к нулю, получим систему линейных уравнений

$$\left\{ \begin{array}{ccccccccc} (P_0,P_0)\,\alpha_0 & + & (P_0,P_1)\,\alpha_1 & +\ldots+ & (P_0,P_n)\,\alpha_n & = & (P_0,f) \\ (P_1,P_0)\,\alpha_0 & + & (P_1,P_1)\,\alpha_1 & +\ldots+ & (P_1,P_n)\,\alpha_n & = & (P_1,f) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ (P_n,P_0)\,\alpha_0 & + & (P_n,P_1)\,\alpha_1 & +\ldots+ & (P_n,P_n)\,\alpha_n & = & (P_n,f) \end{array} \right. \tag{5.16}$$

Введем обозначение $p_{k,l} = (P_k, P_l)$. Тогда матрица этой системы (5.16) примет вид

$$G = \begin{pmatrix} p_{0,0} & p_{0,1} & \dots & p_{0,n} \\ p_{1,0} & p_{1,1} & \dots & p_{1,n} \\ \dots & \dots & \dots & \dots \\ p_{n,0} & p_{n,1} & \dots & p_{n,n} \end{pmatrix}. \tag{5.17}$$

Эта матрица называется матрицей Грама. Ее определитель Δ всегда ≥ 0, причем $\Delta = 0$ тогда и только тогда, когда система $\{P_0(x), P_1(x), \dots, P_n(x)\}$ линейно зависима (см. лемму). В нашем случае, следовательно, $\Delta \neq 0$ и система имеет единственное решение $\tilde{\alpha}_0, \dots, \tilde{\alpha}_n$, которое и дает нужный экстремальный набор коэффициентов и соответствующую линейную комбинацию (5.14).

Примечание. Отметим, что тот же результат получится, если использовать метод Фурье (после ортогонализации системы многочленов).

Лемма. *Определитель Грама равен нулю $\Leftrightarrow$ система $\{P_k(x)\}$ линейно зависима.*

Доказательство. $\Leftarrow$ Пусть система $\{P_k(x)\}$ линейно зависима. Тогда найдется набор чисел $\alpha_0, \dots, \alpha_n$, не равных одновременно нулю такой, что

$$\alpha_0 P_0(x) + \dots + \alpha_n P_n(x) = 0. \tag{5.18}$$

Домножив скалярно обе части этого равенства на $P_k(x)$, $k = 0, \dots, n$, получим однородную линейную систему из $n+1$ уравнения

$$\alpha_0 p_{k,0} + \dots + \alpha_n p_{k,n} = 0, \quad k = 0, \dots, n. \tag{5.19}$$

Поскольку она имеет ненулевое решение, то ее определитель $\Delta = 0$.

$\Rightarrow$ Пусть $\Delta = 0$. Покажем, что система $\{P_k(x)\}$ линейно зависима. Рассмотрим однородную линейную систему (5.19). Ее определитель $\Delta = 0$ и, следовательно, она имеет ненулевое решение $\alpha_0, \dots, \alpha_n$. Покажем, что с этим набором чисел выполняется (5.18). Действительно, домножим скалярно сумму

$$Q(x) = \alpha_0 P_0(x) + \dots + \alpha_n P_n(x)$$

на $P_k(x)$, $k = 0, \dots, n$. Получим $(Q, P_k) = 0$ при всех $k = 0, \dots, n$. Это означает, что и $(Q, Q) = (Q, \alpha_0 P_0 + \dots + \alpha_n P_n) = 0$, то есть $Q(x) \equiv 0$.

Пример 3. Рассмотрим в пространстве $L_2 = L_2([0,1], \rho)$ с весом $\rho(x) = 1$ систему функций $F = \{1, x, x^2\}$. Составить ее матрицу Грама, найти определитель Грама. Написать многочлен наилучшего приближения для функции $f(x) = x^3$ и найти само наилучшее приближение в L_2.

Матрица Грама имеет вид

$$G = \begin{pmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{pmatrix}$$

Находим решение системы (5.16)

$$\alpha_1 = \frac{1}{20}, \quad \alpha_2 = -\frac{3}{5}, \quad \alpha_3 = \frac{3}{2}$$

Многочлен наилучшего приближения в L_2 имеет вид

$$P_2(x) = \frac{3}{2}\, x^2 - \frac{3}{5}\, x + \frac{1}{20}$$

и наилучшее приближение равно $\frac{1}{140}\sqrt{7}$. Это совпадает с результатом примера 1.

6 Численное интегрирование

Пусть $f \in C([a,b])$ и пусть на $[a,b]$ заданы n попарно различных точек $x_1, \ldots, x_n$ (будем называть их квадратурными узлами). Основная задача численного интегрирования (квадратуры) состоит в том, чтобы вычислить приближенно значение интеграла

$$\int_a^b f(x)\,dx$$

с помощью квадратурных сумм вида

$$\sum_{k=1}^{n} A_k f(x_k).$$

Рассмотрим несколько классических методов решения этой задачи.

6.1 Квадратурная формула Симпсона

1. Первый метод численного интегрирования состоит в том, чтобы подобрать числа A_k так, чтобы формула

$$\int_a^b f(x)\,dx \approx \sum_{k=1}^{n} A_k f(x_k), \tag{6.1}$$

была точной для степеней $h_m(x) = x^m$, $m = 0, \ldots, n-1$. Это равносильно тому, что формула (6.1) дает точные значения интегралов от многочленов $P_{n-1}(x) = p_{n-1}x^{n-1} + \ldots + p_0$ степени $\le n-1$.

Примечание. Вместо степеней можно рассматривать и другие пробные функции. Например, $h_m = e^{mx}$.

Рассмотрим сначала случай $[a,b] = [-1,1]$. В данном случае основное условие

$$\int_{-1}^{1} x^m\,dx = \sum_{k=1}^{n} A_k x_k^m, \quad m = 0, \ldots, n-1, \tag{6.2}$$

в развернутом виде дает n линейных уравнений для определения коэффициентов A_k:

$$\sum_{k=1}^{n} A_k x_k^m = \left.\frac{x^{m+1}}{m+1}\right|_{-1}^{1} = \begin{cases} 0 & m = 2s+1 \\ \frac{2}{m+1} & m = 2s \end{cases}, \quad m = 0, \ldots, n-1. \tag{6.3}$$

Определитель этой системы

$$\Delta = \begin{vmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \\ \ldots & x_2 & \ldots & \ldots \\ x_1^{n-1} & x_2^{n-1} & \ldots & x_n^{n-1} \end{vmatrix}$$

есть определитель Вандермонда и, как хорошо известно, вычисляется по формуле

$$\Delta = \prod_{1 \le i < j \le n} (x_j - x_i).$$

Таким образом $\Delta \neq 0$ и, следовательно, система (6.3) имеет и притом единственное решение A_k, $k = 1, \ldots, n$.

Допустим мы решили задачу квадратуры для отрезка $[-1,1]$: при заданных квадратурных узлах $x_k \in [-1,1]$ нашли числа A_k, при которых выполнено (6.2). Перейдем теперь к случаю произвольного отрезка $[a,b]$. В этом случае задача численного интегрирования сводится к уже рассмотренному заменой переменной. Действительно, пусть

$$\tilde{x} = \frac{b-a}{2}x + \frac{b+a}{2}, \qquad d\tilde{x} = \frac{b-a}{2}dx.$$

Тогда

$$\int_a^b f(\tilde{x})\,d\tilde{x} =$$

$$= \frac{b-a}{2}\int_{-1}^{1} f\left(\frac{b-a}{2}x + \frac{b+a}{2}\right)dx \approx \frac{b-a}{2}\sum_{k=1}^{n} A_k f\left(\frac{b-a}{2}x_k + \frac{b+a}{2}\right),$$

то есть

$$\int_a^b f(\tilde{x})\,d\tilde{x} \approx \frac{b-a}{2}\sum_{k=1}^{n} A_k f(\tilde{x}_k), \qquad \tilde{x}_k = \frac{b-a}{2}x_k + \frac{b+a}{2}. \tag{6.4}$$

Эта формула также дает точные значения интегралов от произвольных многочленов степени $\le n-1$.

2. Пример. Найти формулу численного интегрирования по трем равноотстоящим узлам $x_1 = -1$, $x_2 = 0$, $x_3 = 1$.

Составим систему уравнений

$$\begin{cases} A_1 + A_2 + A_3 = 2 \\ -A_1 + 0 + A_3 = 0 \\ A_1 + 0 + A_3 = \frac{2}{3} \end{cases}$$

Отсюда находим $A_1 = A_3 = \frac{1}{3}$, $A_2 = \frac{4}{3}$. Итак,

$$\int_{-1}^{1} f(x)\,dx \approx \frac{1}{3}\left(f(-1) + 4f(0) + f(1)\right) \tag{6.5}$$

Для произвольного отрезка $[a,b]$ отсюда и по формуле (6.4) получаем

$$\int_a^b f(x)\,dx \approx \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right). \tag{6.6}$$

Это равенство называется формулой Симпсона.

3. Погрешность формулы Симпсона. По формуле (6.6) имеем

$$\int_{x_0-h}^{x_0+h} f(x)\,dx \approx \frac{h}{3}(f(x_0-h) + 4f(x_0) + f(x_0+h)).$$

Рассмотрим погрешность

$$\delta(h) := \int_{x_0-h}^{x_0+h} f(x)\,dx - \frac{h}{3}(f(x_0-h) + 4f(x_0) + f(x_0+h))$$

как функцию от h. Найдем производные (проверить):

$$\delta' = \frac{2}{3}f(x_0+h) + \frac{2}{3}f(x_0-h) - \frac{4}{3}f(x_0) - \frac{h}{3}\left(f'(x_0+h) - f'(x_0-h)\right),$$

$$\delta'' = \frac{1}{3}f'(x_0+h) - \frac{1}{3}f'(x_0-h) - \frac{h}{3}\left(f''(x_0+h) + f''(x_0-h)\right),$$

$$\delta''' = -\frac{h}{3}(f'''(x_0+h) - f'''(x_0-h)) = -\frac{2h^2}{3}f''''(c), \quad c \in (x_0-h, x_0+h).$$

Отсюда с учетом того, что $\delta(0) = \delta'(0) = \delta''(0) = 0$ получим

$$|\delta''(h)| \le \left|\delta''(0) + \int_0^h \frac{2t^2}{3}f''''(c)dt\right| \le M_4\frac{2h^3}{9},$$

$$|\delta'(h)| \le \left|\delta'(0) + \int_0^h M_4\frac{2t^3}{9}dt\right| \le M_4\frac{h^4}{18},$$

$$|\delta(h)| \le \left|\delta(0) + \int_0^h M_4\frac{t^4}{18}dt\right| \le M_4\frac{h^5}{90}.$$

4. Комбинированная формула Симпсона для произвольного отрезка с $2n+1$ узлами квадратуры $x_0, x_1, \ldots, x_{2n}$ получается применением (6.6) к каждому промежутку $[x_0, x_2]$, $[x_2, x_4]$, …, $[x_{2n-2}, x_{2n}]$.

Пусть $y_k = f(x_k)$, $k = 0, 1, \ldots, 2n$. Тогда, введя обозначения

$$s_1 = y_1 + y_3 + \ldots + y_{2n-1}, \quad s_2 = y_2 + y_4 + \ldots + y_{2n},$$

получаем вид комбинированной формулы Симпсона:

$$\int_a^b f(x)dx \approx \frac{b-a}{6n}\left(y_0 + y_{2n} + 4s_1 + 2s_2\right).$$

Погрешность оценивается величиной

$$n \cdot M_4\frac{h^5}{90} = nh \cdot M_4\frac{h^4}{90} = \frac{b-a}{180}M_4h^4.$$

6.2 Метод Чебышева

Второй метод численного интегрирования (квадратуры) состоит в том, чтобы при равных числах $A = A_k$ подобрать узлы x_k так, чтобы формула

$$\int_{-1}^{1} f(x)\,dx \approx A\sum_{k=1}^{n} f(x_k), \tag{6.7}$$

была точной на многочленах степени n. В этой задаче имеется $n+1$ параметров $A, x_1, \ldots x_n$, подлежащих определению. Это метод Чебышева.

Составим систему из $n+1$ уравнений для определения величин A и x_k.

$$\begin{cases} A + A + \ldots + A = 2, & A = \frac{2}{n} \\ x_1 + x_2 + \ldots + x_n = & 0 \\ x_1^2 + x_2^2 + \ldots + x_n^2 = & \frac{n}{3} \\ x_1^3 + x_3^2 + \ldots + x_n^3 = & 0 \\ x_1^4 + x_3^4 + \ldots + x_n^4 = & \frac{n}{5} \\ \ldots + \ldots + \ldots + \ldots = & \ldots \\ x_1^n + x_3^n + \ldots + x_n^n = & \frac{n(1+(-1)^n)}{2(n+1)} \end{cases}$$

Основная трудность при решении системы — в ее нелинейности. Но существует способ сведения данной системы к некоторому алгебраическому уравнению порядка n. Еще один недостаток состоит в том, что при $n \geq 10$ и $n = 8$ эта система имеет комплексные решения.

6.3 Метод Гаусса

Третий метод численного интегрирования состоит в том, чтобы подобрать числах A_k и узлы x_k так, чтобы формула

$$\int_{-1}^{1} f(x)\,dx \approx \sum_{k=1}^{n} A_k f(x_k), \tag{6.8}$$

была точной на многочленах степени $2n-1$. В этой задаче имеется $2n$ параметров $A_1, \dots A_n$, $x_1, \dots x_n$, подлежащих определению. Это метод Гаусса.

Запишем систему уравнений

$$\int_{-1}^{1} x^m\,dx = \sum_{k=1}^{n} A_k x_k^m, \quad m = 0, \dots, 2n-1. \tag{6.9}$$

Имеем $2n-1$ нелинейных уравнений для определения указанных параметров. Непосредственное решение системы весьма затруднительно. Гаусс предложил следующий метод ее решения.

Рассмотрим *многочлены Лежандра*

$$L_n(x) = \left(\left(x^2-1\right)^n\right)^{(n)}.$$

Например,

$$L_0 = 1, \quad L_1 = 2x, \quad L_2 = 4(3x^2-1), \quad L_3 = 4x(5x^2-3).$$

Эти многочлены обладают следующими важными свойствами.

1. Многочлен L_n ортогонален на отрезке $[-1,1]$ многочленам степени $k < n$, то есть

$$\int_{-1}^{1} P_k(x) L_n(x)\,dx = 0. \tag{6.10}$$

Это легко проверяется интегрированием по частям.

2. Многочлен L_n имеет ровно n попарно различных действительных корней x_k, $k = 1, \dots, n$, на интервале $(-1,1)$.

Возьмем n корней многочлена L_n в качестве квадратурных узлов. Рассмотрим в системе (6.9) только первые n уравнений, то есть при $m = 0, \dots, n-1$. Решим эту подсистему относительно неизвестных параметров $A_1, \dots A_n$. Покажем, что тогда будут выполнены и оставшиеся равенства в (6.9) при $m = n, \dots, 2n-1$.

Действительно, пусть $m \geq n$ (и $m \leq 2n-1$). Представим x^m в виде

$$x^m = C_p(x) L_n(x) + R_q(x),$$

где $C_p(x)$ — целая часть, а $R_q(x)$ — остаток при делении $x^m / L_n(x)$. Эти многочлены удовлетворяют следующему свойству. И степень p многочлена $C_p(x)$ и степень q многочлена $R_q(x)$ меньше степени многочлена L_n, то есть обе степени $< n$. Первое следует из того, что $p + n = m \leq 2n-1$, $p \leq n-1$, а второе — из того, что степень остатка всегда меньше степени делителя. Отсюда, учитывая (6.10) и свойство чисел A_k, получаем

$$\int_{-1}^{1} x^m\,dx = \int_{-1}^{1} C_p(x) L_n(x)\,dx + \int_{-1}^{1} R_q(x)\,dx =$$

$$= 0 + \sum_{k=1}^{n} A_k R_q(x_k) = \sum_{k=1}^{n} A_k C_p(x_k) L_n(x_k) + \sum_{k=1}^{n} A_k R_q(x_k) =$$

$$= \sum_{k=1}^{n} A_k \left(C_p(x_k) L_n(x_k) + R_q(x_k) \right) = \sum_{k=1}^{n} A_k x_k^m$$

при $m = n, \ldots, 2n-1$. Мы воспользовались также тем, что $L_n(x_k) = 0$ и, следовательно,

$$\sum_{k=1}^{n} A_k C_p(x_k) L_n(x_k) = 0.$$

Что и требовалось.

Таким образом, имеем следующий алгоритм получения квадратурной формулы Гаусса с n узлами:

1. Находим многочлен Лежандра

$$L_n(x) = \frac{1}{2^n n!} \left(\left(x^2 - 1 \right)^n \right)^{(n)}.$$

2. Находим его (попарно различные действительные, лежащие на интервале $(-1, 1)$) корни x_k, $k = 1, \ldots, n$;

3. Составляем систему уравнений (6.2):

$$\sum_{k=1}^{n} A_k x_k^m = \int_{-1}^{1} x^m \, dx = \frac{1 + (-1)^m}{2(m+1)}, \quad m = 0, \ldots, n-1,$$

откуда находим коэффициенты A_k, $k = 1, \ldots, n$.

4. Получаем квадратурную формулу

$$\int_{-1}^{1} f(x) \, dx \approx \sum_{k=1}^{n} A_k f(x_k),$$

точную на многочленах степени $2n - 1$.

Переход к общему случаю отрезка $[a, b]$ проводится по формуле (6.4).

Без доказательства приведем оценку погрешности.

$$\Delta_n = \frac{(b-a)^{2n+1} (n!)^4}{((2n)!)^3 (2n+1)} \max_{x \in [a,b]} |f^{(2n)}(x)|.$$

Приведем значения при $n = 1, 2, 3, 4$ выражения

$$\frac{(n!)^4}{((2n)!)^3 (2n+1)} = \left\{ \frac{1}{24}; \ \frac{1}{4320}; \ \frac{1}{2016000}; \ \frac{1}{1778112000} \right\}.$$

Пример. Получить квадратурную формулу Гаусса численного интегрирования по трем узлам.

Решаем уравнение $L_3(x) = 0$, то есть $x(5x^2 - 3) = 0$. Находим квадратурные узлы $x_1 = -\sqrt{3/5}$, $x_2 = 0$, $x_3 = \sqrt{3/5}$. Решая систему (6.2), находим корни $A_1 = A_3 = 5/9$, $A_2 = 8/9$. В результате имеем

$$\int_{-1}^{1} f(x) \, dx \approx \frac{1}{9} \left(5f \left(-\sqrt{\frac{3}{5}} \right) + 8f(0) + 5f \left(\sqrt{\frac{3}{5}} \right) \right).$$

Для произвольного отрезка $[a, b]$ по формуле (6.4) имеем

$$\int_a^b f(\tilde{x})\, d\tilde{x} \approx$$

$$\approx \frac{b-a}{18}\left(5f\left(-\frac{b-a}{2}\sqrt{\frac{3}{5}}+\frac{b+a}{2}\right)+8f\left(\frac{b+a}{2}\right)+5f\left(\frac{b-a}{2}\sqrt{\frac{3}{5}}+\frac{b+a}{2}\right)\right).$$

7 Численное дифференцирование

1. Рассмотрим задачу о приближенном вычислении значений производных. Будем предполагать, что функция $f(x)$ дифференцируема достаточное число раз. Приближенное вычисление производной $f'(x)$ в точке x основано на формуле Тейлора:

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(c)h^2,$$

где $c \in (x, x+h)$. Отсюда получаем

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \frac{1}{2}f''(c)h,$$

то есть

$$f'(x) \approx \frac{f(x+h) - f(x)}{h},$$

с погрешностью

$$\delta = \frac{h}{2}M_2(f), \quad M_2 = \max_{c\in[x,x+h]} |f''(c)|.$$

2. Эту формулу легко уточнить по порядку величины h. Действительно, поскольку

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2}f''(x)h^2 + \frac{1}{6}f'''(c_1)h^3,$$

$$f(x-h) = f(x) - f'(x)h + \frac{1}{2}f''(x)h^2 - \frac{1}{6}f'''(c_2)h^3,$$

$c_1 \in (x, x+h)$, $c_2 \in (x, x-h)$, то, вычитая из первого равенства второе, получим

$$f(x+h) - f(x-h) = 2f'(x)h + \frac{h^3}{6}\left(f'''(c_1) - f'''(c_2)\right),$$

откуда

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h},$$

с погрешностью

$$\delta = \frac{h^2}{12}M_3(f), \quad M_3 = \max_{c\in[x-h,x+h]} |f'''(c)|.$$

3. Уточнять порядок аппроксимации можно и дальше по следующей схеме. По формуле Тейлора

$$f(x+h) - f(x) = f'(x)h + \sum_{k=2}^{n} \frac{1}{k!}f^{(k)}(x)h^k + \frac{1}{(n+1)!}f^{(n+1)}(c)h^{n+1}, \tag{7.1}$$

где $c \in (x, x+h)$. Пусть фиксированы некоторые попарно различные малые числа h_p, $p = 1, \ldots, n$. Подставим в равенство (7.1) поочередно $h = h_p$, $p = 1, \ldots, n$, и, умножив полученные равенства на некоторые числа A_p, сложим все эти равенства. Тогда

$$\sum_{p=1}^{n} A_p(f(x+h_p) - f(x)) =$$

$$= f'(x)\left(\sum_{p=1}^{n} A_p h_p\right) + \sum_{k=2}^{n}\left(\sum_{p=1}^{n} A_p h_p^k\right)\frac{f^{(k)}(x)}{k!} +$$

$$+\frac{1}{(n+1)!}\sum_{p=1}^{n} A_p f^{(n+1)}(c_p) h_p^{n+1}, \tag{7.2}$$

где $c_p \in (x, x+h_p)$. Далее, пусть выполнены равенства

$$\left\{ \begin{array}{rcl} \sum_{p=1}^{n} A_p h_p & = & 1 \\ \sum_{p=1}^{n} A_p h_p^k & = & 0, \quad k = 2, \ldots, n. \end{array} \right. \tag{7.3}$$

Эта система имеет единственное решение $A_1, \ldots, A_n$ при любом наборе попарно различных и отличных от нуля значениях h_k, поскольку ее определитель есть $h_1 \cdot h_2 \cdots h_n \cdot W$, где W — определитель Вандермонда. Подставив в (7.2) найденное решение, получим

$$f'(x) \approx \sum_{p=1}^{n} A_p f(x+h_p) - Af(x), \qquad A = \sum_{p=1}^{n} A_p, \tag{7.4}$$

с погрешностью

$$\delta_n \le \frac{1}{(n+1)!} M_{n+1} \sum_{p=1}^{n} |A_p||h_p^{n+1}| \le \frac{h^{n+1}}{(n+1)!} M_{n+1} \sum_{p=1}^{n} |A_p|, \tag{7.5}$$

где

$$M_{n+1} = \max_{c \in [x-h, x+h]} |f^{(n+1)}(c)|, \qquad h = \max_{p=1,\ldots,n} \{|h_p|\}.$$

Примечание. Решение системы (7.3) можно получить в явном виде, пользуясь формулами Крамера и Вандермонда:

$$A_p = (-1)^{n+1} \frac{\prod_{q \ne p} h_q}{h_p} \frac{1}{\prod_{q \ne p}(h_p - h_q)} = (-1)^{n+1} \frac{\prod_{q \ne p} h_q}{h_p} \frac{1}{Q'(h_p)},$$

где

$$Q(x) = \prod_{q=1,\ldots,n} (x - h_q).$$

Тогда оценку погрешности (7.5) можно переписать в виде

$$\delta_n \le \frac{M_{n+1}}{(n+1)!} \sum_{p=1}^{n} |A_p||h_p^{n+1}| \le \frac{h^{2n-1}}{(n+1)!} M_{n+1} \sum_{p=1}^{n} \frac{1}{|Q'(h_p)|}.$$

Пример. Пусть $n = 3$ и

$$h_1 = h, \ h_2 = 2h, \ h_3 = 3h,$$

тогда решая систему (7.3), находим

$$A_1 = \frac{3}{h}, \ A_2 = -\frac{3}{2h}, \ A_3 = \frac{1}{3h}, A = \frac{11}{6h},$$

и $\delta_3 \le \frac{9}{4} h^3 M_4$. Здесь и далее погрешности оцениваются по формуле (7.5). То есть получается формула

$$f'(x) \approx \frac{1}{h}\left(3f(x+h) - \frac{3}{2}f(x+2h) + \frac{1}{3}f(x+3h)\right) - \frac{11}{6h}f(x)$$

с указанной погрешностью δ_3.

Пример. Пусть $n = 4$. Тогда при

$$h_{1,2} = \pm\frac{1}{2}h, \qquad h_{3,4} = \pm\frac{3}{2}h,$$

имеем

$$f'(x) \approx \frac{1}{h}\left(\frac{1}{24}f\left(x - \frac{3h}{2}\right) - \frac{9}{8}f\left(x - \frac{h}{2}\right) + \frac{9}{8}f\left(x + \frac{h}{2}\right) - \frac{1}{24}f\left(x + \frac{3h}{2}\right)\right)$$

с погрешностью $\delta_4 \le \frac{3}{512}h^4 M_5$.

4. Вычисление производной второго порядка

Пользуясь тем же методом, можно получить формулу численной производной второго порядка. Именно, снова напишем равенство (7.2)

$$\sum_{p=1}^{n} A_p(f(x + h_p) - f(x)) =$$

$$= f'(x)\left(\sum_{p=1}^{n} A_p h_p\right) + \frac{f''(x)}{2}\left(\sum_{p=1}^{n} A_p h_p^2\right) +$$

$$+ \sum_{k=3}^{n}\left(\sum_{p=1}^{n} A_p h_p^k\right)\frac{f^{(k)}(x)}{k!} + \frac{1}{(n+1)!}\sum_{p=1}^{n} A_p f^{(n+1)}(c_p)h_p^{n+1}. \tag{7.6}$$

Теперь вместо (7.3) составим систему

$$\begin{cases} \sum_{p=1}^{n} A_p h_p &= 0 \\ \sum_{p=1}^{n} A_p h_p^2 &= 2 \\ \sum_{p=1}^{n} A_p h_p^k &= 0, \quad k = 3, \ldots, n. \end{cases} \tag{7.7}$$

Эта система, как отмечено выше, имеет единственное решение при любом наборе попарно различных и отличных от нуля значениях h_k. Решив эту систему и подставив решение в (7.6), получим

$$f''(x) \approx \sum_{p=1}^{n} A_p f(x + h_p) - Af(x), \quad A = \sum_{p=1}^{n} A_p, \tag{7.8}$$

с погрешностью (см. (7.5))

$$\delta_n \le \frac{1}{(n+1)!}M_{n+1}\sum_{p=1}^{n}|A_p||h_p^{n+1}| \le \frac{h^{n+1}}{(n+1)!}M_{n+1}\sum_{p=1}^{n}|A_p|.$$

Аналогично вычисляются производные более высоких порядков.

Пример 2. При $n = 4$ имеем

$$f''(x) \approx$$

$$\frac{1}{h^2}\left(-\frac{1}{18}f\left(x - \frac{3}{2}h\right) + \frac{9}{2}f\left(x + \frac{1}{2}h\right) + \frac{9}{2}f\left(x - \frac{1}{2}h\right) - \frac{1}{18}f\left(x + \frac{3}{2}h\right)\right) - \frac{80}{9h^2}f(0)$$

с погрешностью $\delta_4 \le \frac{3}{320}h^3 M_5$.

Пример 3. Составим аппроксимативную формулу для $f'(x) + f''(x)$ в узлах $x + kh$, $k = 1, 2, 3$.

Составляем систему

$$\begin{cases} hA_1 + 2hA_2 + 3hA_3 = 1, \\ h^2A_1 + 4h^2A_2 + 9h^2A_3 = 2, \\ h^3A_1 + 8h^3A_2 + 27h^3A_3 = 0, \end{cases} \qquad \begin{cases} A_1 + 2A_2 + 3A_3 = 1/h, \\ A_1 + 4A_2 + 9A_3 = 2/h^2, \\ A_1 + 8A_2 + 27A_3 = 0. \end{cases}$$

Отсюда

$$A_1 = \frac{9h-23}{h^2}, \quad A_2 = \frac{-9h+263}{2h^2}, \quad A_3 = \frac{h-3}{h^2}.$$

Тогда

$$f'(x) + f''(x) = \sum_{k=1}^{3} A_k f(x+kh) - f(x) \sum_{k=1}^{3} A_k =$$

$$= \frac{9h-23}{h^2} \cdot f(x+h) + \frac{-9h+263}{2h^2} \cdot f(x+2h) + \frac{h-3}{h^2} \cdot f(x+3h) - f(x) \cdot \frac{11h-26}{2h^2}.$$

8 Численное решение дифференциальных уравнений

Рассмотрим обыкновенное дифференциальное уравнение с начальным условием (задача Коши)

$$y' = f(x, y), \qquad y(x_0) = y_0. \tag{8.1}$$

Заметим, что из (8.1) можно получить все производные решения $y = y(x)$. То есть, исходя из равенства

$$y'(x) = f(x, y(x)), \tag{8.2}$$

получаем

$$y''(x) = f'_x(x, y(x)) + f'_y(x, y(x))f(x, y(x)), \tag{8.3}$$

$$y''' = f''_{xx} + f''_{xy}f + (f'_{xy} + f'_{yy}f)f + f'_y(f'_x + f'_y f), \tag{8.4}$$

и так далее.

Рассмотрим некоторые методы приближенного решения задачи (8.1).

8.1 Метод Эйлера

Метод основан на формуле Тейлора и на равенстве (8.2):

$$y(x + h) = y(x) + y'(x)h + r_1(x) = y(x) + f(x, y(x))h + r_1(x),$$

где

$$r_1(x) = \frac{1}{2}y''(c)h^2, \ c \in (x, x + h).$$

Полагая $x_k = x_0 + kh$, $k = 0, 1, \ldots$, $y(x_k) = y_k$ и отбрасывая остаточный член r_1, получаем расчетную формулу для нахождения приближенного решения:

$$y_{k+1} = y_k + f(x_k, y_k)h, \quad h = x_{k+1} - x_k, \quad y_0 = y(x_0). \tag{8.5}$$

Локальная точность формулы (8.5) имеет порядок h^2. Это означает, что при точном значении y_k следующее значение y_{k+1} решения вычисляется приближенно с абсолютной погрешностью, имеющей порядок h^2 при малых h.

8.2 Уточненный метод Эйлера

Если функция $f(x, y)$ задана в виде аналитического выражения и мы можем вычислить значение ее частных производных первого порядка, то формулу (8.5) можно уточнить, с помощью равенства (8.3). Именно, применив формулу Тейлора, получим

$$y(x + h) = y(x) + y'(x)h + \frac{1}{2}y''(x)h^2 + r_2(x), \ \ r_2(x) = \frac{1}{6}y'''(c)h^3, \ c \in (x, x + h).$$

Отсюда и из (8.2), (8.3) находим

$$y(x + h) = y(x) + f(x, y(x))h + \frac{1}{2}\left(f'_x(x, y(x)) + f'_y(x, y(x))f(x, y(x))\right)h^2 + r_2(x). \tag{8.6}$$

Отсюда получаем расчетную формулу для приближенного решения

$$y_{k+1} = y_k + f(x_k, y_k)h + \frac{1}{2}\left(f'_x(x_k, y_k) + f'_y(x_k, y_k)f(x_k, y_k)\right)h^2. \tag{8.7}$$

Локальная точность формулы (8.7) имеет порядок h^3. По такой схеме можно уточнять формулу Эйлера и далее, добавляя новые слагаемые в формуле Тейлора и используя формулы для производных функции y более высоких порядков (типа (8.4)).

8.3 Метод Рунге — Кутты

Формула (8.7) имеет тот недостаток, что не всегда имеется возможность вычислить частные производные от функции $f(x,y)$, например, если она задана табличными значениями (или же эти производные имеют очень громоздкий вид). Поэтому аппроксимируем входящее в (8.6) выражение

$$F(h) = f(x,y)h + \frac{1}{2}\left(f'_x(x,y) + f'_y(x,y)f(x,y)\right)h^2 \tag{8.8}$$

линейной комбинацией

$$G(h) = h\left(\alpha f(x,y) + \beta f(x+\gamma h, y+\sigma h)\right),$$

не содержащей частных производных. Для этого величины α, β, γ, σ подберем так, чтобы $|F(h)-G(h)| = O(h^3)$. Тогда локальная точность полученной формулы будет по-прежнему иметь порядок h^3.

В дальнейших выкладках будем считать, что x и y — независимые переменные (не имеющие отношения к решению $y = y(x)$). По формуле Тейлора имеем (с точностью до величин порядка h^3):

$$G(h) \approx h\left(\alpha f(x,y) + \beta\left(f(x,y) + f'_x(x,y)\,\gamma h + f'_y(x,y)\,\sigma h\right)\right) =$$

$$= (\alpha+\beta)f(x,y)h + \beta\left(\gamma f'_x(x,y) + \sigma f'_y(x,y)\right)h^2.$$

Сравнивая полученное выражение с (8.8), получаем равенство:

$$f(x,y)\,h + \frac{1}{2}\,\left(f'_x(x,y) + f'_y(x,y)\,f(x,y)\right)\,h^2 =$$

$$= (\alpha+\beta)\,f(x,y)h + \beta\,\left(\gamma\,f'_x(x,y) + \sigma\,f'_y(x,y)\right)\,h^2.$$

Отсюда находим

$$\alpha+\beta = 1, \quad \beta\gamma = \frac{1}{2}, \quad \beta\sigma = \frac{f(x,y)}{2},$$

$$\alpha = 1-\beta, \quad \gamma = \frac{1}{2\beta}, \quad \sigma = \frac{f(x,y)}{2\beta}.$$

В результате вместо (8.7) получаем следующую расчетную формулу:

$$y_{k+1} = y_k + h\left((1-\beta)f(x_k,y_k) + \beta f\left(x_k + \frac{h}{2\beta}, y_k + f(x_k,y_k)\frac{h}{2\beta}\right)\right). \tag{8.9}$$

Обычно полагают $\beta = \frac{1}{2}$. Тогда

$$y_{k+1} = y_k + \frac{h}{2}\left[f(x_k,y_k) + f\left(x_k+h, y_k + f(x_k,y_k)h\right)\right]. \tag{8.10}$$

При расчетах можно применять следующую последовательность вычислений:

$$v_1 = f(x_k,y_k), \quad v_2 = f\left(x_k+h, y_k+v_1h\right), \quad v_3 = v_1+v_2, \quad y_{k+1} = y_k + \frac{h}{2}v_3. \tag{8.11}$$

Как отмечалось, локальная точность формулы (8.10) имеет порядок h^3. Существуют более точные схемы того же типа. Они приводятся в справочниках. Приведем пример схемы, локальная точность которой имеет 4-й порядок.

$$v_1 = f(x_k,y_k);$$

$$v_2 = f(x_k + h/2, y_k + v_1 h/2);$$
$$v_3 = f(x_k + h/2, y_k + v_2 h/2);$$
$$v_4 = f(x_k + h, y_k + v_3 h);$$
$$y_{k+1} = y_k + (h/6)(v_1 + 2v_2 + 2v_3 + v_4).$$

Все эти схемы называется одношаговыми, поскольку при вычислении y_{k+1} использует значение y_k только в одной предыдущей точке.

8.4 Системы дифференциальных уравнений

Рассмотрим задачу

$$\left\{ \begin{array}{rcl} u' &=& F(x,u,v) \\ v' &=& G(x,u,v) \end{array} \right. , \quad u(x_0) = u_0,\ v(x_0) = v_0, \tag{8.12}$$

относительно неизвестных $u = u(x)$, $v = v(x)$.

Если исходить из приближенных формул

$$u(x+h) \approx u(x) + u'(x)h, \qquad v(x+h) \approx v(x) + v'(x)h,$$

то, по аналогии с предыдущим, получим расчетную формулу Эйлера для задачи (8.12):

$$\left\{ \begin{array}{rcl} u_{k+1} &=& u_k + F^{(k)} \cdot h \\ v_{k+1} &=& v_k + G^{(k)} \cdot h \end{array} \right. , \qquad F^{(k)} = F(x_k, u_k, v_k),\ G^{(k)} = G(x_k, u_k, v_k), \tag{8.13}$$

где $x_k = x_0 + k\,h$, $u_k = u(x_k)$, $v_k = v(x_k)$, $k = 0, 1, \ldots$. Если исходить из более точных формул

$$u(x+h) \approx u(x) + u'(x)h + \frac{1}{2}u''(x)h^2 = u(x) + F\,h + \frac{1}{2}\left(F_x + F_u F + F_v G\right)h^2,$$

$$v(x+h) \approx v(x) + v'(x)h + \frac{1}{2}v''(x)h^2 = v(x) + G\,h + \frac{1}{2}\left(G_x + G_u F + G_v G\right)h^2,$$

то получим уточненную формулу Эйлера для задачи (8.12):

$$\left\{ \begin{array}{rcl} u_{k+1} &=& u_k + F^{(k)} \cdot h + \frac{1}{2}\left(F_x^{(k)} + F_u^{(k)} F^{(k)} + F_v^{(k)} G^{(k)}\right) h^2 \\ v_{k+1} &=& v_k + G^{(k)} \cdot h + \frac{1}{2}\left(G_x^{(k)} + G_u^{(k)} F^{(k)} + G_v^{(k)} G^{(k)}\right) h^2 \end{array} \right. , \tag{8.14}$$

Здесь верхний индекс (k) означает, что соответствующая функция вычислена в точке (x_k, u_k, v_k), где $x_k = x_0 + k\,h$, $u_k = u(x_k)$, $v_k = v(x_k)$, $k = 0, 1, \ldots$.

К системе вида (8.12) сводится уравнение второго порядка вида

$$y'' = G(x, y, y'), \quad y(x_0) = u_0, \quad y'(x_0) = v_0.$$

Достаточно ввести обозначения $y = u$, $y' = v$, тогда это уравнение можно переписать в виде системы

$$\left\{ \begin{array}{rcl} u' &=& v \\ v' &=& G(x,u,v) \end{array} \right. , \quad u(x_0) = u_0,\ v(x_0) = v_0.$$

Согласно (8.14) приближенное решение такой системы ищется по формулам

$$\left\{ \begin{array}{rcl} u_{k+1} &=& u_k + v_k \cdot h + \frac{1}{2}G^{(k)}h^2 \\ v_{k+1} &=& v_k + G^{(k)} \cdot h + \frac{1}{2}\left(G_x^{(k)} + G_u^{(k)} v_k + G_v^{(k)} G^{(k)}\right) h^2 \end{array} \right. \tag{8.15}$$

Пример 1. Решить приближенно задачу Коши

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

В этом случае при $y = u$, $y' = v$ уравнение записывается в виде системы

$$\left\{ \begin{array}{rcl} u' & = & v \\ v' & = & u + \frac{v^2}{u} \end{array} \right. , \quad u(0) = 1,\ v(0) = 0.$$

Имеем $G(x, u, v) = u + \frac{v^2}{u}$ и (8.15) принимает вид

$$\left\{ \begin{array}{rcl} u_{k+1} & = & u_k + v_k \cdot h + \frac{1}{2}\left(u_k + \frac{v_k^2}{u_k}\right) h^2 \\ v_{k+1} & = & v_k + \left(u + \frac{v_k^2}{u_k}\right) \cdot h + \frac{1}{2}\left(\left(1 + \frac{v_k^2}{u_k^2}\right) v_k + 2v_k\right) h^2 \end{array} \right.$$

Вычисления с шагом $h = 0.01$ дают приближенное решение на отрезке $[0, 1]$ с погрешностью $\leq 10^{-4}$ (точное решение $y(x) = e^{x^2/2}$)).

Еще более частным случаем является часто возникающее на практике линейное уравнение 2-го порядка, то есть задача

$$A(x)y'' + B(x)y' + C(x)y = 0, \qquad y(x_0) = u_0, \quad y'(x_0) = v_0, \tag{8.16}$$

где $A(x) \neq 0$. Оно записывается в виде

$$\left\{ \begin{array}{rcl} u' & = & Mu + Nv \\ v' & = & Ku + Lv \end{array} \right. , \qquad u(x_0) = u_0,\ v(x_0) = v_0.$$

где $M = 0$, $N = 1$, $K = -C/A$, $L = -B/A$ — заданные функции. Для этой системы можно применять формулы (8.13), (8.14).

Пример 2. Приведем еще один способ решения задачи (8.16). Имеем

$$A\frac{y''}{y} + B\frac{y'}{y} + C = 0, \qquad A = A(x),\ B = B(x),\ C = C(x).$$

Введем новую неизвестную функцию $R = y'/y$. Тогда, поскольку

$$\left(\frac{y'}{y}\right)' = \frac{y''}{y} - \left(\frac{y'}{y}\right)^2,$$

то есть

$$\frac{y''}{y} = R' + R^2,$$

из (8.16) получаем задачу первого порядка

$$AR' + AR^2 + BR + C = 0, \qquad R(x_0) = \frac{v_0}{u_0}, \tag{8.17}$$

или, в другой форме,

$$R' = -R^2 + \mu R + \gamma, \qquad \mu = -\frac{B}{A},\ \gamma = -\frac{C}{A}, \qquad R(x_0) = \frac{v_0}{u_0}. \tag{8.18}$$

Решим ее, методами предыдущих пунктов. Например, при решении задачи (8.18) можно использовать формулу (8.7). Найдя R, получим еще одно уравнение

$$\frac{y'}{y} = R, \quad y(x_0) = u_0. \tag{8.19}$$

Теперь можно использовать, что

$$\ln y = \ln y(x_0) + \int_0^x R(t)dt \quad \Leftrightarrow \quad y = u_0 e^{\int_0^x R(t)dt},$$

и вычислить

$$y_k \approx u_0 e^{h\sum_{m=1}^{k} R_m}, \quad R_m = R(x_0 + hm).$$