1 Решение нелинейных уравнений

В этом разделе будут рассматриваться методы приближенного вычисления действительных корней алгебраических и трансцендентных уравнений вида

$$F(x) = 0. \tag{1}$$

Нелинейная функция $F(x)$ в своей области определения может иметь конечное или бесконечное множество нулей или не иметь их вовсе. Большинство методов нахождения корней уравнения (1) требуют знания промежутков, где имеется только лишь один корень. Тем самым, до поиска корней требуется ответить на вопрос о том, какое количестве корней имеет уравнение и где они расположены (*локализованы*). Для решения этой задачи можно использовать как точные аналитические методы, например, методы математического анализа, так и приближенные методы, например, графический способ. После того, как определены границы расположения корней, следует найти приближенные значения корней и уточнить их одним из численных методов с требуемой точностью. Все рассматриваемые далее методы, за исключением метода деления отрезка пополам, являются итерационными методами, требующими знания одного или нескольких начальных приближений к искомому корню.

1.1 Локализация корней

Применим для отделения корней некоторые теоремы математического анализа. Если на концах некоторого отрезка $[a, b]$ непрерывная функция $F(x)$ принимает значения разных знаков $F(a)F(b) < 0$, то на (a, b) она хотя бы один раз обращается в нуль. Вообще говоря, на этом отрезке функция

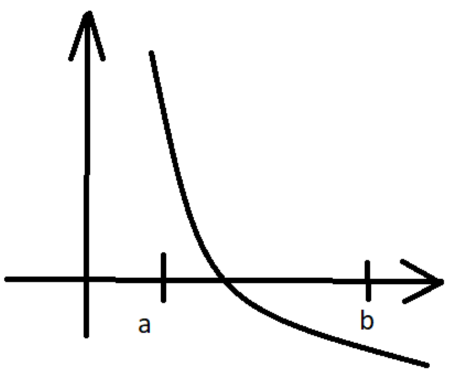


Рисунок 1

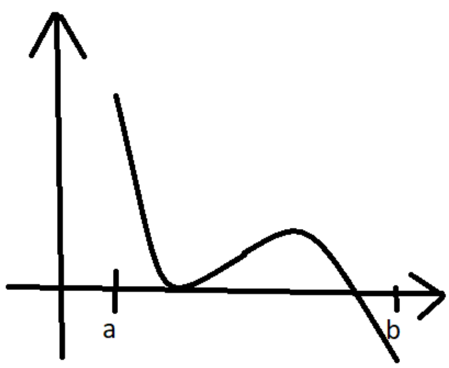


Рисунок 2

будет иметь нечетное число нулей, посчитанных с учетом их кратности. Если же дополнительно предположить, что функция дифференцируема и ее производная сохраняет знак на $[a, b]$, то уравнение $F(x) = 0$ будет иметь единственный корень на (a, b).

Алгоритм аналитического отделения корней:

1) Найдите область определения функции
2) Найдите интервалы монотонности
3) Определите знак функции на концах каждого интервала
4) Сделайте вывод о числе корней
5) Уменьшите интервал локализации корня до отрезка единичной длины

Аналитическое отделение корней. Рассмотрим уравнение

$$x^2 e^x = \pi.$$

Будем разыскивать нули функции $F(x) = x^2 e^x - \pi$. Найдем $F'(x)$.

$$F'(x) = 2xe^x + x^2e^x = x\,e^x(2+x) = 0 \quad x_1 = -2\; x_2 = \; 0$$

Составим таблицу поведения функции $F(x)$

x	$(-\infty;-2)$	-2	$(-2;0)$	0	$(0;+\infty)$
$F'(x)$	$+$	0	$-$	0	$+$
$F(x)$	- возрастает	$-$	убывает	$-$	возрастает +
корень	нет корней		нет корней		есть корень

$$F(-2) = \frac{4}{e^2} - \pi < 0; F(0) = -\pi < 0$$

Так как производная меняет знак в двух точках, то есть три отрезка, где функция сохраняется монотонность. Определим знаки функций на концах этих отрезков

$$\lim_{x\to+\infty} F(x) = \lim_{x\to+\infty}(x^2e^x - \pi) = +\infty$$

$$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty}(x^2e^x - \pi) = \lim_{x\to-\infty}\frac{x^2}{e^{-x}} - \pi = -\pi$$

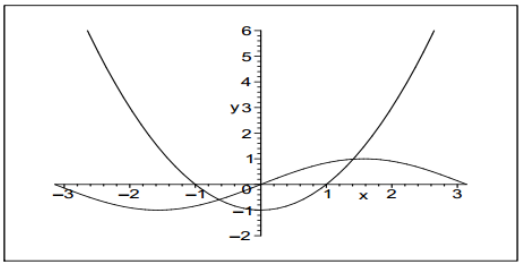
$$F(1) = e - \pi < 0 \quad , F(2) = 4e^2 - \pi > 0$$

Вывод: Функция $F(x) = x^2e^x - \pi$ имеет единственный корень $\in (1; 2)$

Графический способ отделения корней. Пусть требуется исследовать корни уравнения

$$x^2 - \sin x = 1.$$

Представим его в виде $x^2 - 1 = \sin x$ и построим графики функции $f(x) = x^2 - 1$ и $g(x) = \sin x$. Задача отделения корней исходного уравнения свелась к задаче отделения абсцисс точек пересечения двух графиков функций $f(x)$ и $g(x)$.



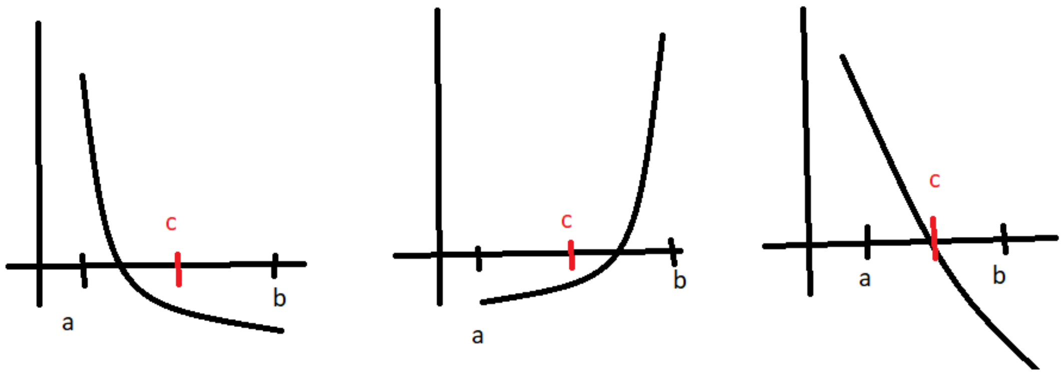
На рис. 1 видно, что графики функций имеют две точки пересечения, расположенные на промежутках $(1, 2)$ и $(-1, 0)$.

Вывод: Нелинейное уравнение $x^2 - \sin x - 1 = 0$ имеет два корня $(-1; 0)$, $(1,2)$.

1.2 Метод половинного деления

Для уточнения корня, локализованного на отрезке $[a, b]$ используем *метод половинного деления* (другие названия — *метод дихотомии* или *метод бисекций*). Пусть функция $F(x)$ определена и непрерывна при всех $x \in [a, b]$ и выполняется условие $F(a)F(b) < 0$. И пусть функция $F(x)$ имеет на $[a, b]$ только один корень.

Метод половинного деления заключается в том, что последовательно определяется середина отрезка $c = \frac{a+b}{2}$ и вычисляется значение функции $F(c)$ и значение функции, например, $F(a)$ для сравнения. При этом возможна одна из трех ситуаций:



1. $F(c) = 0$
2. $F(a)F(c) < 0$
3. $F(a)F(c) > 0$.

В первом случае корень уравнения найден и равен $x = c$. Это утверждение справедливо только лишь при проведении точных (!!!), то есть аналитических, вычислений. При численной реализации считается, что $F(c) = 0$, если $|F(c)| < \varepsilon$, где ε, например, 10^{10}. Вообще говоря, из того, что $F(c) = 0$ не следует, что c — корень уравнения. Приведем контрпример. Пусть имеется функция вида $F(x) = 0.5\,(x-1)$ и $\varepsilon = 10^{-6}$. Число $c = 1 + 10^{-6}$ удовлетворяет условию $|F(c)| < \varepsilon$ и может считаться корнем уравнения. Теперь рассмотрим функцию $G(x) = 0.5 \cdot 10^{-6}(x-1)$, которая будет иметь те же корни, что и $F(x)$. При $x = 2$ для функции $G(x)$ выполняется условие $|G(c)| < \varepsilon$, значит, $x = 2$ — корень $G(x) = 0$, тогда как корнем является $x = 1$. Для контроля вычисления простого корня на отрезке $[a, b]$ следует, вообще говоря, проверять условие

$$\frac{|F(c)|}{m_1} < \varepsilon, \quad \text{здесь } |F'(x)| < m_1, \quad x \in [a, b].$$

Во втором случае корень следует искать на (a, c), так как теперь на концах этого отрезка функция принимает значения разных знаков. Поэтому считаем $b = c$ и повторяем алгоритм. В третьем случае корень попадает на (c, b), считаем $a = c$ и повторяем алгоритм.

Этот процесс повторяется до тех пор, пока длина промежутка $(b - a)$ не станет меньше 2ε, где ε — требуемая точность определения корня, в этом случае корнем считается середина отрезка $c = \frac{a+b}{2}$.

Таким образом, критерием окончания итерационного процесса метода является выполнение одного из условий

$$|F(c)| < \varepsilon \quad \text{или} \quad |b - a| < 2 * \varepsilon.$$

Метод половинного деления сводится к построению последовательности вложенных друг в друга отрезков $[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \ldots \supset [a_n, b_n]$, таких что $F(a_n)F(b_n) < 0$. Левые концы отрезков a, a_1, $a_2, \ldots$, a_n представляют собой ограниченную неубывающую последовательность, правые концы b, b_1, $b_2, \ldots$, b_n — ограниченную невозрастающую последовательность. Такие последовательности являются сходящимися и имеют предел. Кроме этого выполнено соотношение

$$|b_n - a_n| = \frac{1}{2^n}(b - a).$$

Устремляя $n \to \infty$, будем иметь $|b_n - a_n| \to 0$. Это означает, что последовательности a_i, b_i имеют одинаковый предел, который и будет корнем уравнения

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = x_*.$$

Если c_n — середина отрезка $[a_n, b_n]$, на котором есть корень x_*, то погрешность корня, найденного методом половинного деления, удовлетворяет оценке

$$|x_* - c_n| \leqslant \frac{b_n - a_n}{2} = \frac{1}{2^{n+1}}(b-a). \tag{2}$$

Метод половинного деления прост и надежен: он сходится к простому корню[1] уравнения (1) для любых непрерывных функций, в том числе и для недифференцируемых. Этот метод устойчив к ошибкам округления и гарантирует точность результата. Недостатком метода является невысокая скорость сходимости. Для того, чтобы длина отрезка $[a, b]$ стала меньше ε, требуется выполнить $\approx \log_2 \frac{b-a}{\varepsilon}$ итераций (см. (2)). Если уравнение имеет несколько корней на $[a, b]$, то метод обязательно сойдется к одному из них, хотя заранее неизвестно, к какому. Если на отрезке $[a, b]$ будет, например, три корня, но один из них двухкратный, то, скорее всего метод сойдется

[1]Корень $x = c$ называется простым, если $F(c) = 0$, $F'(c) \neq 0$.

к простому корню уравнения. Таким образом, следует применять метод половинного деления, когда несущественна скорость и требуется высокая надежность результата при условии, что корень на искомом отрезке один.

Пример 1. Уточним корень уравнения

$$x^2\, e^x = \pi$$

методом половинного деления с точностью $\varepsilon = 10^{-5}$. Заметим, что единственный корень этого уравнения локализован на отрезке $(1, 2)$ (см.п.(1.1)).

Зададим функцию

$$F(x) = x^2\, e^x - \pi.$$

Для заданной точности $\varepsilon = 10^{-5}$ потребуется $\geqslant \log_2 10^5 \approx 17$ итераций. Выполним расчет и результаты занесем в таблицу.

N	a	b	c	$F(a)$	$F(c)$	$F(b)$
1	1.000000	2.000000	1.500000	-0.423311	6.942208	26.414632
2	1.000000	1.500000	1.250000	-0.423311	2.312068	6.942208
3	1.000000	1.250000	1.125000	-0.423311	0.756807	2.312068
4	1.000000	1.125000	1.062500	-0.423311	0.125006	0.756807
5	1.000000	1.062500	1.031250	-0.423311	-0.158999	0.125006
6	1.031250	1.062500	1.046875	-0.158999	-0.019529	0.125006
...						
17	1.049011	1.049026	1.049019	-0.000071	-0.000001	0.000068

В результате за 17 итераций корень 1.04902 найден с точностью $\varepsilon = 10^{-5}$.

1.5 Метод Ньютона и его модификации

Если известно хорошее начальное приближение к решению уравнения $F(x) = 0$, то самым эффективным методом уточнения этого приближения является метод Ньютона.

Пусть $x = c$ — точное решение уравнения $F(x) = 0$, $x = x_n$ — некоторое приближение к нему. В предположении, что существуют непрерывные производные $F'(x)$, $F''(x)$ разложим функцию $F(x)$ по формуле Тейлора в окрестности x_n

$$F(x) = F(x_n) + F'(x_n)(x - x_n) + \frac{F''(\theta_n)}{2}(x - x_n)^2, \qquad (18)$$

где $\theta_n \in [a, b]$ — некоторая точка, лежащая между x и x_n.

Это разложение справедливо и для $x = c$

$$0 = F(c) = F(x_n) + F'(x_n)(c - x_n) + \frac{F''(c_n)}{2}(c - x_n)^2, \qquad (19)$$

здесь $c_n \in [a, b]$ лежит между c и x_n. Если бы значение c_n известно, то полученное уравнение (19) могло бы служить для определения c. Считая x_n достаточно близким к корню $x = c$ и полагая малым последнее слагаемое в (19), получим уравнение для нового приближения к корню x_{n+1}

в (19), получим уравнение для нового приближения к корню x_{n+1}

$$F(x_n) + F'(x_n)(x_{n+1} - x_n) = 0 \tag{20}$$

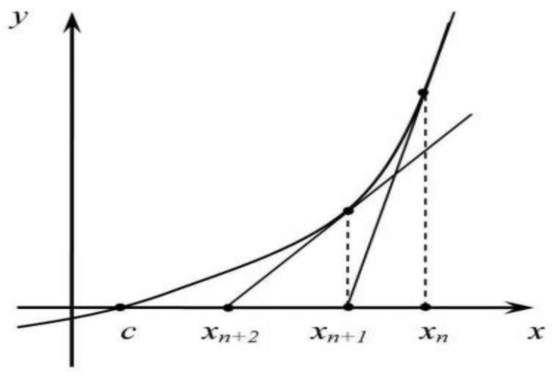
или

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}, \quad n = 0, 1, 2, \ldots$$

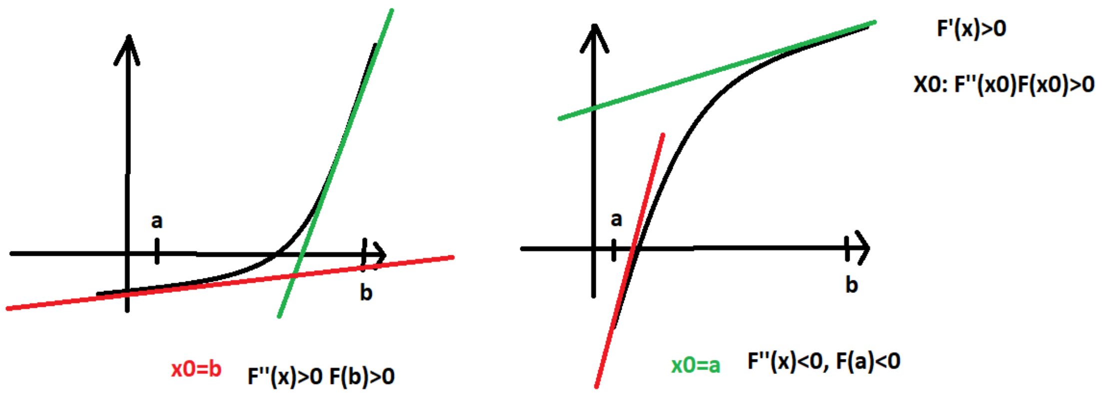
Геометрический смысл полученного метода становится ясен, если записать уравнение касательной к графику функции $F(x)$ в точке x_n:

$$y = F(x_n) + F'(x_n)(x - x_n).$$

В точке пересечения с осью Ox касательная имеет абсциссу x_{n+1} (см. (20)). Таким образом, в методе Ньютона каждое новое приближение x_{n+1} определяется как абсцисса точки пересечения касательной к графику $F(x)$, проведенной в точке x_n, с осью Ox. Отсюда и второе название метода — *метод касательных*.



Выбор начального приближения. Рассмотрим монотонно возрастающую функцию с единственным нулем на отрезке (a, b)



Сформулируем условия сходимости метода Ньютона. Пусть функция $F(x) \in C^2_{[a,b]}$, $F(a)F(b) < 0$, а производные $F'(x)$, $F''(x)$ сохраняют знак на $[a, b]$. Тогда, исходя из начального приближения $x_0 \in [a, b]$, удовлетворяющего условию

$$F(x_0)F''(x_0) > 0 \tag{21}$$

последовательные приближения

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}, \quad n = 0, 1, 2, \dots \tag{22}$$

монотонно сходятся к единственному корню $x = c \in [a, b]$ уравнения $F(x) = 0$.

Для оценки скорости сходимости метода Ньютона рассмотрим соотношение

$$x_{n+1} - c = x_n - \frac{F(x_n)}{F'(x_n)} - c, \tag{23}$$

Исходя из разложения (19), имеем

$$c - x_n = -\frac{F(x_n)}{F'(x_n)} - \frac{1}{2}\frac{F''(c_n)}{F'(x_n)}(c - x_n)^2.$$

Подставляя последнее в (23), получим

$$x_{n+1} - c = (x_n - c) - \frac{F(x_n)}{F'(x_n)} = \frac{1}{2}\frac{F''(c_n)}{F'(x_n)}(c - x_n)^2.$$

Эта оценка демонстрирует быструю, квадратичную скорость сходимости метода Ньютона и позволяет считать его методом 2-го порядка [1].

[1] Метод имеет p-ый порядок сходимости, если выполняется $|c - x_{n+1}| \leqslant C_1 |c - x_n|^p$, где $|c - x_n| \leqslant C_2$, здесь $0 < C_1, C_2 < +\infty$

Для окончания итерационного процесса можно использовать условие $|F(x_{n+1})| < \varepsilon$, или условие близости двух последовательных приближений:

$$|x_{n+1} - x_n| < \varepsilon, \quad n = 0, 1, 2, \dots. \tag{27}$$

На самом деле, можно доказать следующую оценку

$$|x_{n+1} - c| \leqslant \frac{M_2}{2\,m_1}|x_{n+1} - x_n|^2, \quad n = 0, 1, 2, \dots. \tag{28}$$

Недостатком этой оценки является то, что оценить вторую и первую производные функции $F(x)$ не всегда представляется возможным. Если же выполняется условие (27), то оценка (28) дает следующее

$$|x_{n+1} - c| \leqslant \frac{M_2}{2\,m_1}\,\varepsilon^2.$$

Тем самым, если $\frac{2\,m_1}{M_2} \geqslant \varepsilon$, то корень x_{n+1} найден с точностью ε.

Пример 3. Уточним корень уравнения

$$x^2\,e^x = \pi \tag{33}$$

методом Ньютона с точностью $\varepsilon = 10^{-5}$. Найдем производную функции

$$F(x) = x^2e^x - \pi; F'(x) = \; 2xe^x + x^2e^x$$

Запишем итерационную расчетную формулу метода Ньютона

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}, n = 0,1,2,\ldots$$

$$x_{n+1} = x_n - \frac{x_n^2\,e^{x_n} - \pi}{2x_ne^{x_n} + x_n^2e^{x_n}}$$

Выберем x_0 из отрезка, на котором локализован корень (см. п.(1.1)) $x_0 \in [1,2]$. Подберем x_0 так, чтобы выполнялась теорема сходимости метода (21). Для этого вначале найдем $F''(x)$

$$F'(x) = \; 2xe^x + x^2e^x; \;\; F''(x) = \; 2e^x + 4xe^x + x^2e^x;$$

$$F(1) = e - \pi < 0; \rightarrow \; F(1)F''(1) < 0; \; F(2) = 4\,e^2 - \pi > 0 \rightarrow \;\; F(2)F''(2) > 0 \rightarrow x_0 = 2$$

$$x_1 = x_0 - \frac{x_0^2\,e^{x_0} - \pi}{2x_0e^{x_0} + x_0^2e^{x_0}} = 2 - \frac{4e^2 - \pi}{4e^2 + 4e^2} = 2 - 0.4469 = 1.5531$$

следовательно, $x_0 = 2$. Выполним несколько итераций метода Ньютона, результаты поместим в таблицу

n	x_n	$F(x_n)$	$F'(x_n)$	$F(x_n)/F'(x_n)$
0	2.000000	26.414632	59.112449	0.446854
1	1.553146	8.259523	26.082433	0.316670
2	1.236476	2.123017	13.780116	0.154064
3	1.082412	0.316775	9.848478	0.032165
4	1.050247	0.011233	9.156793	0.001227
5	1.049021	0.000016	9.131210	0.000002
6	1.049019			

Таким образом, методом Ньютона корень найден за 6 итераций и равен 1.04902. Заметим, что в 5-й колонке стоит величина разности между соседними итерациями, контроль за которой влияет на окончание итерационного процесса.