Итерационные методы поиска корней функции

Волченко Ю.М.

Содержание лекции

Итерационные методы решения уравнений. Отделение и уточнение корней уравнений. Метод простых итераций. Метод хорд. Метод касательных. Комбинированный метод хорд и касательных. Понятие об итерационных методах решения систем уравнений.

Анимация сходимости и расходимости процесса поиска корня. Анимационная демонстрация метода хорд и метода касательных.
Анимация работает только в программе Acrobat Reader!

Поиск корней функции в системе *Mathematica*.

14 апреля 2011 г.

1 Постановка задачи

Часто при расчете параметров какого-нибудь устройства, некоторые его характеристики приходится подбирать способом, который в народе называется «методом тыка», а в научных кругах — методом «неосознанного поиска» или методом «проб и ошибок». Насколько ненаучен или, наоборот, научен этот метод? И, если уж приходится к нему прибегать, то нельзя ли сделать так, чтобы проб и ошибок было поменьше, а степень осознанности была бы побольше?

Перейдем к более точным формулировкам. Пусть требуется на входе аппаратуры подобрать такое напряжение x, чтобы на выходе получилось заданное напряжение $y = a$. Естественно считать, что y есть функция x (известная или неизвестная): $y = g(x)$. Значит, чтобы справиться с поставленной задачей, надо решить уравнение $g(x) = a$. Это уравнение можно переписать в виде

$$f(x) = 0, \tag{1}$$

где $f(x) = g(x) - a$.

Если уравнение (1) — линейное или квадратное, то достаточно применить известные формулы корней соответствующего уравнения и задача будет решена. Для алгебраических уравнений высоких степеней таких формул не существует (и существовать не может, кроме частных случаев), а для неалгебраических даже говорить о каких-то формулах не приходится. Кроме того, функция $f(x)$ может быть вообще неизвестна (то есть неизвестно ее аналитическое выражение), как это чаще всего и бывает при разработке новых устройств. Поэтому точное решение поставленной задачи практически никогда найти не удается, и мы вынуждены обратиться к численным, приближенным методам ее решения.

Задачу решения уравнения (1) будем называть также ***задачей поиска корней функции*** $f(x)$, или ***задачей поиска корней трансцендентного уравнения***, так как в левую часть (1) могут входить синусы, косинусы, логарифмы, экспоненты и т.п., которые носят название трансцендентных функций.

2 Общая схема итерационных методов

Для решения уравнения (1) используют различные методы последовательного приближения к корню функции $f(x)$, которые также называют ***итерационными методами***.

Вначале задается начальное приближение x_0, а следующие приближения $x_1, x_2, \ldots$ получают по формуле

$$x_{k+1} = F_k\left[x_k, f(x_k)\right], \; k = 0, 1, 2, \ldots ,$$

где F_k — функция двух переменных, которая и определяет тот или иной метод последовательного приближения.

Очередное приближение к корню x_k считается решением поставленной задачи, если выполняется неравенство

$$|x^* - x_k| < \varepsilon, \tag{2}$$

где x^* — искомый корень уравнения (1), а $\varepsilon > 0$ — заданная точность вычислений.

Проверить выполнение неравенства (2) удается не всегда, поэтому ***критерий останова*** (2) часто заменяют другими критериями:

$$|f(x_k)| < \varepsilon, \; |x_{k+1} - x_k| < \varepsilon, \; \left|\frac{x_{k+1} - x_k}{x_k}\right| < \varepsilon, \ldots .$$

Для применения большинства методов требуется изолировать корни уравнения (1), то есть найти отрезки, на каждом из которых находится лишь один

корень функции $f(x)$. Поэтому задача поиска корней уравнения (1) разбивается на две: 1) изоляция, или отделение корней, 2) поиск корня на заданном отрезке.

3 Отделение корней уравнения

Задача отделения корней уравнения (1) в общем случае не может быть решена без привлечения нематематических соображений. Обычно, исходя из технических, физических, химических и т.п. особенностей решаемой задачи находят отрезок $[A, B]$, на котором решение имеет смысл, то есть значение величины x при выполнении неравенств $x < A$ или $x > B$ либо не встречается на практике, либо вообще не может существовать.

Далее отрезок $[A, B]$ точками $A = A_0$, A_1, $\ldots$, A_{n-1}, $A_n = B$ разбивают на части и проверяют условия

$$f(A_{i-1})\, f(A_i) < 0,\ i = \overline{1, n}.$$

Если для некоторого i условие выполнено и функция $f(x)$ непрерывна, то на отрезке $[A_{i-1}, A_i]$ находится по крайней мере один корень уравнения (1). Затем, исследуя функцию на этом отрезке, пытаются выяснить, имеются ли на нем еще корни. Иногда это обнаруживается только на втором этапе поиска корней, при применении одного из итерационных методов.

Отметим, что отрезки $[A_{i-1}, A_i]$, $i = \overline{1, n}$, называются ***интервалами изоляции*** корней уравнения.

Если уравнение (1) может быть преобразовано к равносильному уравнению

$$p(x) = q(x), \tag{3}$$

причем, $p(x)$ и $q(x)$ представляют собой функции, графики которых нетрудно построить, то начальное приближение к корню, как и отрезок, на котором он находится, могут быть получены графически. Именно, корнями уравнения (1) будут абсциссы точек пересечения графиков функций $p(x)$ и $q(x)$, так как они определяются уравнением (3), которое равносильно уравнению (1). Заключив каждый графически найденный корень в некоторый отрезок (на котором нет других корней), мы тем самым решим задачу отделения корней (см. рис. 1).

4 Метод простых итераций

Перейдем к изучению второго этапа поиска корней уравнения, к изучению методов последовательного приближения к корню. Предположим, что на отрезке $[a_0, b_0]$, который мы будем называть ***начальным отрезком поиска***, на-

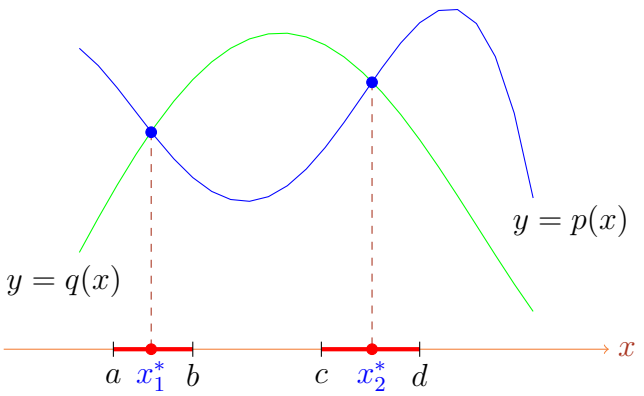

Рис. 1. Графическое отделение корней.

ходится единственный корень непрерывной функции $y = f(x)$. Чтобы с заданной точностью $\varepsilon > 0$ найти его приближенное значение, представим уравнение (1) в виде

$$x = \varphi(x), \tag{4}$$

где $\varphi(x) = x - f(x)$, и зададим ***начальное приближение*** x_0 к корню, взяв любую точку из отрезка $[a_0, b_0]$ (например, середину отрезка). Следующие приближения вычисляются по формуле

$$\boxed{x_{k+1} = \varphi(x_k), \ k = 0, 1, 2, \ldots.} \tag{5}$$

Применение этой формулы составляет суть метода ***простых итераций***, которые названы простыми потому, что при получении очередного приближения в правой части (5), кроме значения функции, ничего не вычисляется.

Если последовательность точек x_0, x_1, x_2, …, x_k, …, получаемая с помощью уравнения (5), неограниченно приближается к корню уравнения x^*, то есть $\lim_{k\to\infty} x_k = x^*$, то говорят, что она ***сходится*** к x^*; в противном случае говорят, что она ***расходится***. При расходимости такой последовательности ее точки могут оставаться вдалеке от точки x^* или даже удаляться от нее. Если же последовательность сходится, то, получив x_k, удовлетворяющее неравенству (2), тем самым решаем поставленную задачу.

Достаточным условием сходимости является выполнение неравенства[1)]

$$|\varphi'(x)| \le M < 1, \tag{6}$$

где $M > 0$ — некоторое число. В качестве критерия останова обычно используется неравенство[2)]

$$|x_{k+1} - x_k| < (1 - M)\,\varepsilon. \tag{7}$$

Рис. 2 и 3 иллюстрируют сходимость и расходимость метода простых итераций. Щелкните мышкой на рисунке или на соответствующей кнопке медиаплеера и вы сможете наблюдать процесс сходимости или расходимости в движении. Как видно из второго рис., расходимость может быть обусловлена как зацикливанием процесса, когда попеременно выполняются равенства $\varphi(\alpha) = \beta$ и $\varphi(\beta) = \alpha$, так и все увеличивающимся удалением текущего приближения к корню от его истинного значения.

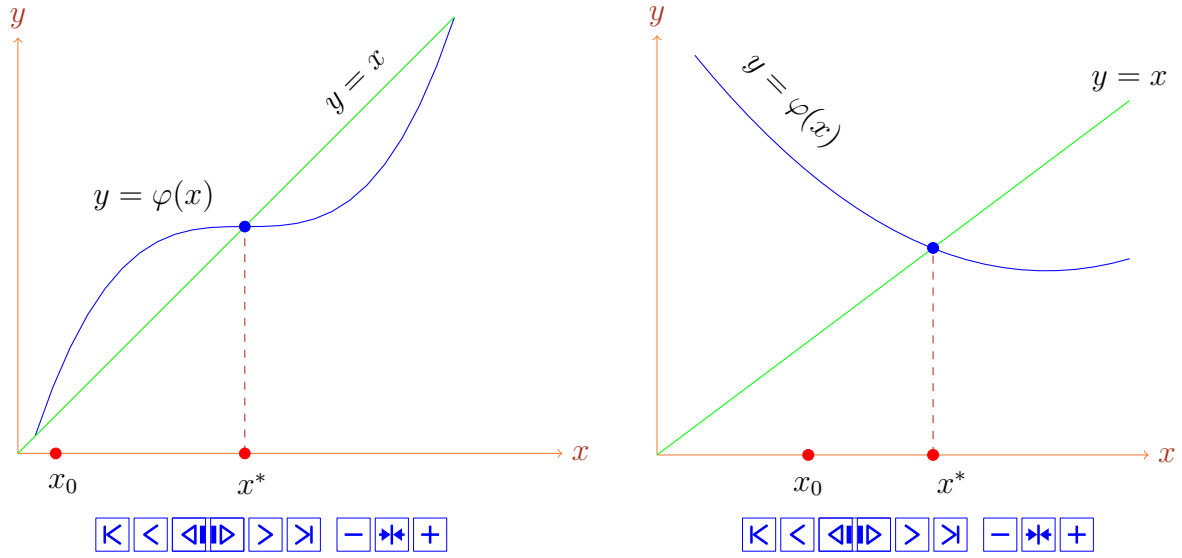

Рис. 2. Сходимость итераций.

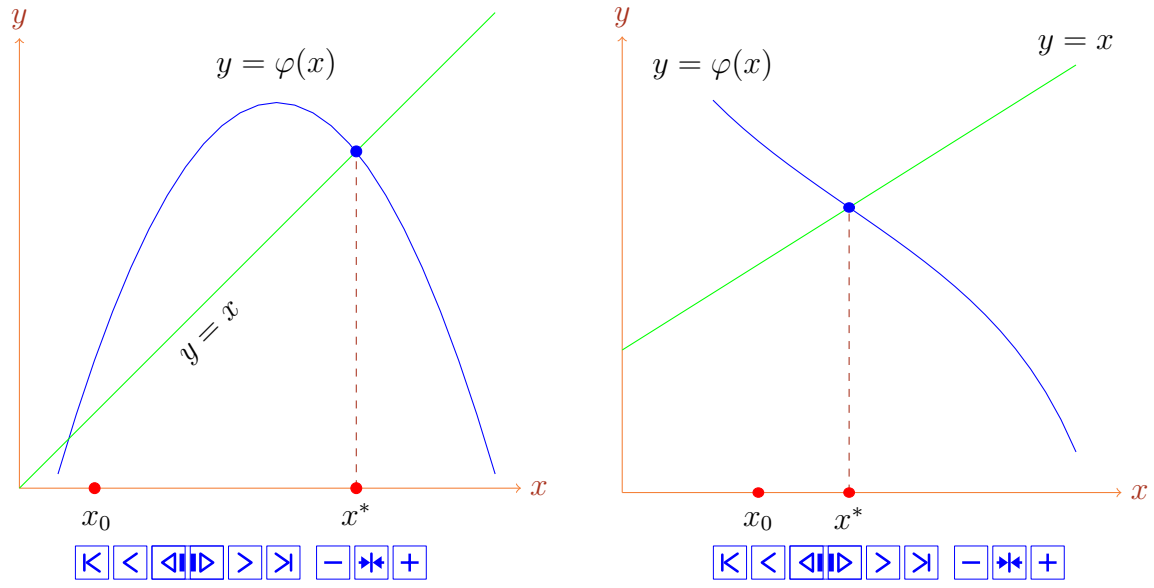

Рис. 3. Расходимость итераций.

Если процесс поиска корня сходится медленно или не выполняется условие (6), то вместо уравнения (4) решают уравнение [3, с. 361]

$$x = \overline{\varphi}\,(x, \lambda)\,,$$

где $\overline{\varphi}(x,\lambda) = x - \lambda f(x)$. Соответственно, последовательность приближений к корню находят по формуле

$$\boxed{x_{k+1} = \overline{\varphi}(x_k, \lambda_k),\ k = 0, 1, 2, \ldots .} \tag{8}$$

Числа λ_k выбирают так, чтобы модуль производной

$$\left|\frac{d\overline{\varphi}(x,\lambda_k)}{dx}\right|$$

был невелик в окрестности точки x_k, полагая, например,

$$\lambda_k = \frac{1}{|f'(x_k)|}.$$

В этом случае в формуле (7) $M = 0$ и критерий останова принимает вид

$$|x_{k+1} - x_k| < \varepsilon. \tag{9}$$

Пример 1. *Вычислить с точностью* $\varepsilon = 0{,}01$ *приближенное значение корня уравнения*

$$x^3 + 5x + 4 = 0.$$

Решение. Представим данное уравнение в виде

$$x^3 = -5x - 4$$

и построим графики функций $y = x^3$ и $y = -5x - 4$ на одном рисунке (рис. 4). Из него мы видим, что корень у заданного уравнения один, и находится он, по-видимому, на отрезке $[-1; 0]$, так как этому отрезку принадлежит абсцисса точки пересечения графиков функций.

Проверим это. Обозначим $f(x) = x^3 + 5x + 4$ и проведем простые вычисления: $f(-1)\,f(0) = -2 \cdot 4 = -8 < 0$. Следовательно, на концах указанного отрезка функция $f(x)$ имеет значения противоположных знаков и поэтому в силу своей непрерывности обязательно имеет на этом отрезке корень. Единственность корня подтверждается тем, что $f'(x) = 3x^2 + 5 > 0$. Это означает, что функция возрастает, и ее график не может пересечь ось абсцисс более одного раза.

В качестве начального приближения к корню возьмем середину отрезка $[-1; 0]$: $x_0 = -0{,}5$. Остальные приближения найдем по формуле (8):

$$x_{k+1} = x_k - \lambda_k\left(x_k^3 + 5x_k + 4\right),\ k = 0, 1, 2, \ldots ,$$

где $\lambda_k = 1/|f'(x_k)| = 1/(3x_k^2 + 5)$, проверяя достижение заданной точности по критерию (9). Вычисления сведем в таблицу:

k	x_k	$\lvert x_{k+1} - x_k\rvert$	$3x_k^2 + 5$	λ_k	$x_k^3 + 5x_k + 4$
0	$-0{,}5$	—	5,750	0,174	1,375
1	$-0{,}739$	0,239	6,638	0,151	$-0{,}099$
2	$-0{,}724$	0,015	6,573	0,152	0
3	$-0{,}724$	0	—	—	—

Таким образом, с точностью до 0,01 корень равен $x_2 = -0{,}72$.

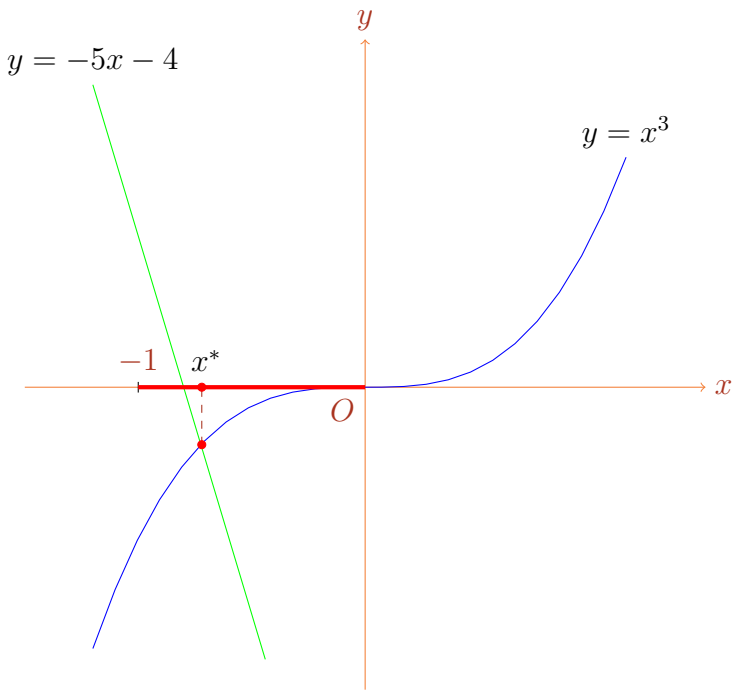

Рис. 4. Графическое решение уравнения.

5 Метод хорд

В методе хорд начальным приближением к корню берется точка пересечения хорды, имеющей концы $(a_0, f(a_0))$ и $(b_0, f(b_0))$ с осью Ox (см. рис. 5). Уравнение хорды имеет вид

$$\frac{y - f(a_0)}{f(b_0) - f(a_0)} = \frac{x - a_0}{b_0 - a_0}.$$

Учитывая, что в точке пересечения с Ox ордината хорды равна нулю, из такого уравнения можно найти абсциссу точки пересечения:

$$x_0 = a_0 - \frac{(b_0 - a_0) f(a_0)}{f(b_0) - f(a_0)}. \tag{10}$$

Если $f(a_0) f(x_0) < 0$, то в качестве следующего отрезка поиска берется отрезок $[a_1, b_1] = [a_0, x_0]$, в противном случае — отрезок $[a_1, b_1] = [x_0, b_0]$. Повторяя построение хорды и поиск абсциссы точки ее пересечения с осью абсцисс на отрезке $[a_1, b_1]$, найдем отрезок $[a_2, b_2]$ и т.д., получая в итоге систему вложенных отрезков $[a_0, b_0] \supset [a_1, b_1] \supset [a_2, b_2] \supset \ldots$ и последовательность приближений к корню $x_0, x_1, x_2, \ldots$. На рис. 5 этот процесс показан в движении.

Критерием останова обычно считают неравенство [2, с. 466]

$$|x_{k+1} - x_k| < \frac{|f(x_k)|}{m}, \tag{11}$$

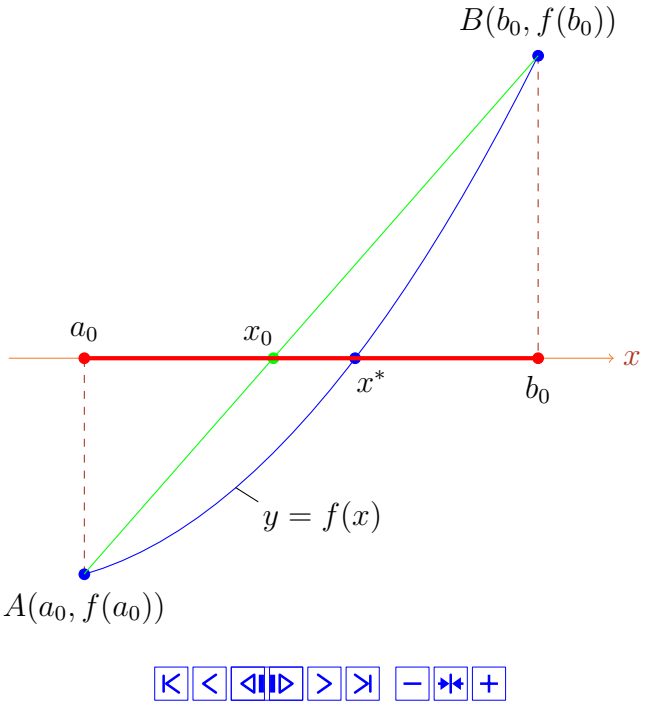

Рис. 5. Метод хорд.

где m определяется условием $0 \leq m \leq |\varphi'(x)|$, $x \in [a_0, b_0]$.

6 Метод касательных

Предположим, что в интервале (a_0, b_0) функция $y = f(x)$ дважды дифференцируема. В этом случае для получения начального приближения к корню можно взять точку пересечения касательной к графику функции $y = f(x)$ с осью Ox (см. рис. 6). Точкой касания берут ту из двух точек, A или B, для которой выполняется неравенство [2, с. 467]

$$f(x) f''(x) > 0. \tag{12}$$

Уравнение касательной имеет вид

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

С учетом того, что в точке пересечения касательной с осью абсцисс ордината равна нулю, получаем абсциссу этой точки:

$$x_0 = b_0 - \frac{f(b_0)}{f'(b_0)}. \tag{13}$$

Новый отрезок поиска корня выбираем, как и в методе хорд; повторяем на нем процедуру построения касательной и т.д. Критерием останова этого процесса

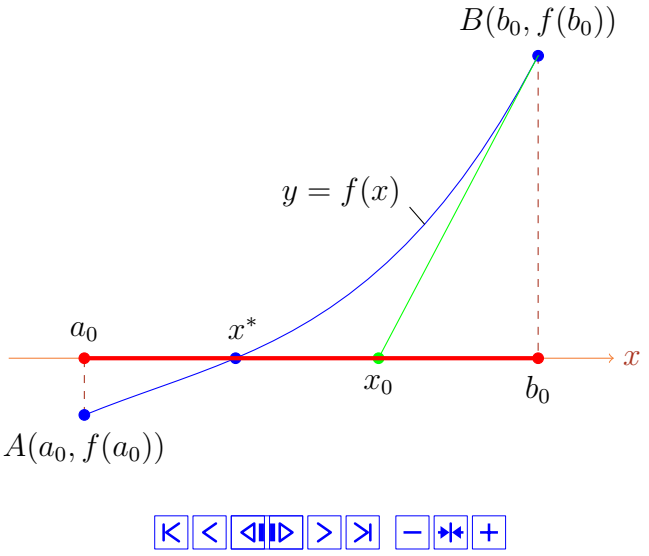

Рис. 6. Метод касательных.

по-прежнему служит неравенство (11). Анимируя рис. 6, можно наблюдать процесс приближения к корню в движении.

7 Комбинированный метод хорд и касательных

В этом методе к отрезку поиска $[a_k, b_k]$ одновременно применяют и метод хорд, и метод касательных. Пусть, например, условие (12) выполняется на правом конце отрезка. Тогда левый конец нового отрезка поиска вычисляют по методу хорд, а именно по формуле (10):

$$\boxed{a_{k+1} = a_k - \frac{(b_k - a_k)\, f(a_k)}{f(b_k) - f(a_k)},}$$

правый конец — по формуле (13) (метод касательных):

$$\boxed{b_{k+1} = b_k - \frac{f(b_k)}{f'(b_k)},}$$

а очередным приближением к корню берут середину отрезка поиска $x_k = (a_k + b_k)/2$, $k = 0, 1, 2, \ldots$. Критерием останова удобно взять неравенство

$$b_k - a_k < \varepsilon.$$

Этот критерий наиболее естественный из всех рассмотренных, а комбинированный метод хорд и касательных — один из самых эффективных.

Пример 2. *Решить предыдущий пример комбинированным методом хорд и касательных.*

Решение. Для начального отрезка поиска $[-1;0]$ неравенство (12) выполняется на его левом конце:

$$f(-1)\,f''(-1) = (-2)(-6) = 12 > 0,$$

и поэтому метод касательных следует применять к левым концам всех (почему?)[3] отрезков поиска:

$$a_{k+1} = a_k - \frac{f(a_k)}{f'(a_k)},$$

а метод хорд — к правым концам

$$b_{k+1} = b_k - \frac{(b_k - a_k)\,f(b_k)}{f(b_k) - f(a_k)}.$$

Вычисления представим в виде таблицы:

k	a_k	b_k	$b_k - a_k$	$f(a_k)$	$f(b_k)$	$f'(a_k)$
0	-1	0	1	-2	4	8
1	$-0{,}75$	$-0{,}667$	0,083	$-0{,}172$	0,368	6,688
2	$-0{,}724$	$-0{,}724$	0	—	—	—

Мы видим, что скорость сходимости в данном случае выше, а ответ, конечно, такой же, как в предыдущем примере.

8 Системы уравнений

Итерационные методы применяют также и при решении систем уравнений. Мы вкратце познакомимся с двумя из них. Запишем систему уравнений в векторной форме:

$$\mathbf{f}(\mathbf{x}) = 0, \tag{14}$$

где $\mathbf{x} = (x_1, ..., x_n)$ — вектор неизвестных, а $\mathbf{f} = (f_1, ..., f_n)$ — вектор правых частей уравнений.

Представим ее в виде

$$\mathbf{x} = \boldsymbol{\varphi}(\mathbf{x}),$$

где $\boldsymbol{\varphi}(\mathbf{x}) = \mathbf{x} - \mathbf{f}(\mathbf{x})$. Пусть вектор $\mathbf{x}^*$ является ***корнем*** безразлично какой из этих двух равносильных систем уравнений, то есть обращает каждое из уравнений системы в числовое тождество. Метод простых итераций для системы уравнений заключается в приближении к корню по формуле

$$\mathbf{x}_{k+1} = \boldsymbol{\varphi}(\mathbf{x}_k),$$

где $\mathbf{x}_0$ — начальное приближение.

Не вдаваясь в подробности, отметим, что при определенных условиях этот ***векторный метод простых итераций*** сходится к $\mathbf{x}^*$ с заданной точностью.

Еще одним итерационным методом является ***метод Ньютона***. Он применяется непосредственно к системе уравнений (14), а его итерационная формула имеет вид

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{f}_{\mathbf{x}}^{-1}(\mathbf{x}_k)\mathbf{f}(\mathbf{x}_k),$$

где $\mathbf{f}_{\mathbf{x}}^{-1}$ — матрица, обратная к матрице

$$\mathbf{f}_{\mathbf{x}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{pmatrix}.$$

Имеются теоремы, которые доказывают существование и невырожденность такой матрицы и сходимость последовательности точек, генерируемых методом Ньютона, к корню уравнения. Нетрудно заметить, что многомерный метод Ньютона является обобщением одномерного метода касательных.

Система *Mathematica* имеет в своем составе операторы, решающие задачу поиска корней как уравнений, так и их систем численными методами[4)].

Приложение

1) Так как $\varphi(x_0) = x_1$ и $x^* = \varphi(x^*)$, то по теореме Лагранжа

$$|x_1 - x^*| = |\varphi(x_0) - \varphi(x^*)| \le |\varphi'(c)|\,|x_0 - x^*| \le M\,|x_0 - x^*|,$$

$c \in [a_0, b_0]$. Продолжая и т.д., получаем

$$|x_n - x^*| < M^n\,|x_0 - x^*| \underset{n\to\infty}{\longrightarrow} 0,$$

а, значит, $\lim_{n\to\infty} |x_n - x^*| = 0$ и поэтому $\lim_{n\to\infty} x_n = x^*$.

2) Используя неравенство треугольника и теорему Лагранжа, приходим к неравенству

$$\begin{aligned}|x^* - x_n| &\le |x_{n+1} - x_n| + |x^* - x_{n+1}| = |x_{n+1} - x_n| + |\varphi(x^*) - \varphi(x_n)| = \\ &= |x_{n+1} - x_n| + |\varphi'(c)|\,|x^* - x_n| < |x_{n+1} - x_n| + M\,|x^* - x_n|,\end{aligned}$$

$c \in [a_0, b_0]$, которое приводит к оценке

$$|x^* - x_n| < \frac{|x_{n+1} - x_n|}{1 - M} < \varepsilon.$$

Из нее и получается условие останова (7).

3) В данном примере $f''(x) = 6x < 0$, $x \in [-1; 0]$, а $f(x) < 0$ для всех $x \in [-1; x^*)$ и поэтому $f(x)\,f''(x) > 0$ для всех $x \in [-1; x^*)$.

4) Прежде всего следует отметить оператор `RootIntervals`, который находит интервалы изоляции корней многочленов с рациональными коэффициентами. Если уравнение имеет вид (1) и $f(x)$ — многочлен, то указанный оператор фактически находит интервалы изоляции корней такого уравнения. Он может даже больше: найти интервалы изоляции сразу для нескольких многочленов и указать к какому многочлену какой интервал относится.

Вот пример:

```
RootIntervals[{x^3 + 2 x^2 - 3, 2 x^2 - 5 x + 2}]
{{{0, 1}, {1, 1}, {1, 3}}, {{2}, {1}, {2}}}
```

Ответ состоит из списка, в котором имеется два подсписка: в первом указаны интервалы изоляции, а во втором — номера многочленов, которые имеют корень в соответствующем интервале. Из этих списков видно, что корни многочленов находятся на отрезках $[0; 1]$, $[1; 1]$, $[1; 3]$, причем, корень первого многочлена находится во втором интервале, а корень второго — в первом и третьем. Второй отрезок состоит из одной точки, $x = 1$, это и есть единственный корень первого многочлена.

Отделим корни уравнения, рассмотренного на лекции:

```
RootIntervals[x^3 + 5 x + 4]
{{{-2, 0}}, {{1}}}
```

Интервал получился шире, чем тот, который был получен графическим методом на лекции. Таким образом, последний в данном случае оказался более эффективным. Поскольку у нас только один многочлен, подсписок с номерами многочленов нам не нужен, поэтому выделим первый подсписок и используем предыдущий оператор в более удобном виде:

```
RootIntervals[x^3 + 5 x + 4][[1]]
{{-2, 0}}
```

Теперь перейдем к оператору, который находит корни уравнений и их систем численно. Это — оператор `FindRoot`. В простейшем случае он требует задания уравнения и пары вида $\{x, x_0\}$, где x — неизвестное, x_0 — начальное приближение к корню. Например, решение трансцендентного уравнения $x - \cos x = 0$ выглядит так:

```
FindRoot[x - Cos[x], {x,-2}]
{x -> 0.739085}
```

Можно заметить, что, уравнения, как такового, под знаком оператора нет: просто `FindRoot` не различает корни функции и корни уравнения, для которого эта функция является левой частью (правая считается равной нулю). Поэтому записывать после функции «== 0» не обязательно.

Решим уравнение, рассмотренное нами на занятии:

```
FindRoot[x^3 + 5 x + 4 == 0, {x, -1}]
{x → -0.724076}
```

А вот корень следующего многочлена (мы уже видели, что корень равен 1) *Mathematica* искать отказывается:

```
FindRoot[x^3 + 2 x^2 - 3, {x, 0}]
FindRoot::jsing: Encountered a singular Jacobian at the point {x} = {0.}.
Try perturbing the initial point(s). »
```

Дело в том, что многочлен $x^3 + 2x^2 - 3$ в начальной точке поиска $x_0 = 0$ имеет минимум, и это препятствует оператору решить задачу. Что ж, возьмем другую начальную точку:

```
FindRoot[x^3 + 2 x^2 - 3, {x, 4}]
{x → 1.}
```

Если уравнение имеет несколько корней, оператор `FindRoot` найдет только один из них, в зависимости от того, какую начальную точку поиска выбрать:

```
FindRoot[2 x^2 - 5 x + 2 == 0, {x, 0}]
{x → 0.5}
FindRoot[2 x^2 - 5 x + 2 == 0, {x, 3}]
{x → 2.}
```

И, наконец, этот оператор может решать системы уравнений, только последние должны быть оформлены списком, а еще один список должны составить пары (неизвестное, начальное значение):

```
FindRoot[{y == Exp[x], x + y == 2}, {{x, 1}, {y, 1}}]
{x → 0.442854, y → 1.55715}
```

Этот пример показывает, что приближенное значение корня (0,442854; 1,55715) было найдено в результате решения системы уравнений

$$\begin{cases} y = e^x, \\ x + y = 2, \end{cases}$$

при начальном приближении $(x_0, y_0) = (1, 1)$.

Литература

[1] Пискунов Н.С. *Дифференциальное и интегральное исчисления для втузов.* — М.: Наука, 1985, — 432 с.

[2] Березин И.С., Жидков Н.П. *Методы вычислений.* — М.: Наука, 1966. — 632 с.

[3] *Задачи и упражнения по математическому анализу* / Под ред. Б.П. Демидовича. — М.: Наука, 1966, 472 с.

[4] Плис А.И., Сливина Н.А. *Лабораторный практикум по высшей математике*. — М.: Высшая школа, 1983, 208 с.

[5] Козин А.С., Лященко Н.Я. *Вычислительная математика*. — Киев, Рад. школа, 1983. — 191 с.