Численные методы решения нелинейных и трансцендентных уравнений

Постановка задачи

Пусть дано уравнение

$$F(x) = 0, \qquad (1)$$

где функция $F(x)$ определена и непрерывна в конечном или бесконечном интервале $a < x < b$.

Всякое значение ξ, обращающее функцию $F(x)$ в нуль, то есть такое, что $F(\xi) = 0$, называется корнем уравнения (1) или нулем функции $F(x)$. Предположим, что уравнение (1) имеет лишь изолированные корни, то есть для каждого корня существует окрестность, не содержащая других корней этого уравнения.

Методы решения нелинейных уравнений делятся на аналитические, графические и численные (приближенные). Приближенное нахождение изолированных действительных корней уравнения (1) складывается обычно из двух этапов:

1. **Отделение корней**, то есть установление возможно тесных промежутков $[\alpha, \beta]$, в которых содержится один и только один корень исходного уравнения (1).
2. **Уточнение приближенных корней**, то есть доведение их до заданной степени точности.

Графическое решение уравнений

Действительные корни уравнения $F(x) = 0$ приближенно можно определить как абсциссы точек пересечения графика функции $y = F(x)$ с осью OX (см. рис. 1, *а*) . На практике часто бывает удобнее уравнение (1) заменить равносильным ему уравнением

$$\varphi(x) = \psi(x), \qquad (2)$$

где функции $\varphi(x)$ и $\psi(x)$ более простые, чем функция $F(x)$. Тогда, построив графики этих функций, искомые корни получим как абсциссы точек пересечения этих графиков (см. рис. 1, *б*).

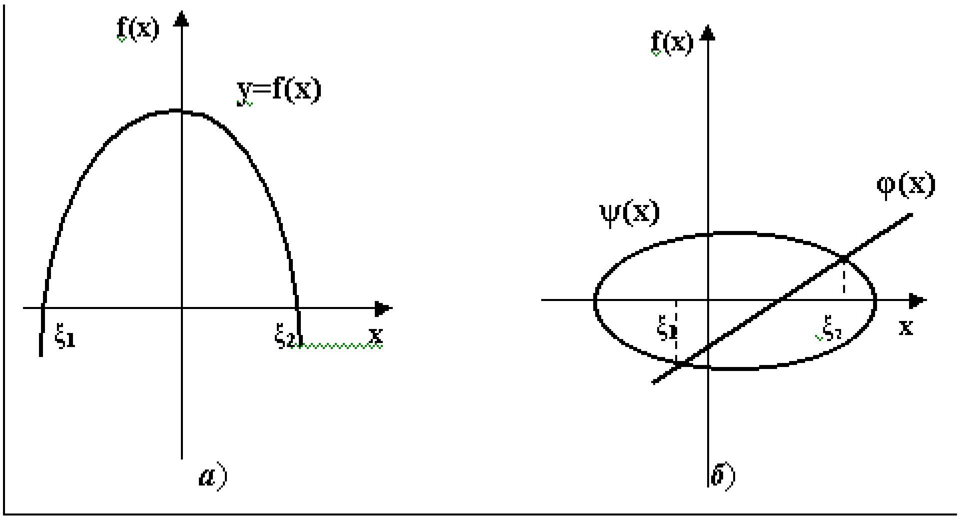

Рис. 1 . Графический метод нахождения корней уравнения.

Численные методы

Метод (дихотомии) деления отрезка пополам

Сформулируем без доказательства очень важную для рассмотрения дальнейших вопросов теорему.

Т е о р е м а: *Если непрерывная функция* $f(x)$ *принимает значения разных знаков на концах отрезка* $[\alpha, \beta]$, *то есть* $f(\alpha)\cdot f(\beta) < 0$, *то внутри этого отрезка содержится по меньшей мере один корень уравнения* $f(x) = 0$, *а именно: найдётся хотя бы одно число* $\xi \in [\alpha, \beta]$ *такое, что* $f(\xi) = 0$.

Пусть дано уравнение:

$$f(x) = 0, \tag{3}$$

где функция $f(x)$ определена и непрерывна на интервале $[a, b]$ и $f(a)\cdot f(b) < 0$. Для нахождения корня уравнения делим отрезок $[a, b]$ пополам:

- если $f((a + b)/2) = 0$, то $\xi = (a + b)/2$ является корнем уравнения (3);
- если $f((a+b)/2) \neq 0$, то выбираем ту половину отрезка $[a, (a + b)/2]$ или $[(a + b)/2, b]$, на концах которого функция $f(x)$ имеет противоположные знаки. Новый суженный отрезок $[a_1, b_1]$ снова делим пополам и проводим тот же анализ и т.д.

Очевидно, что закончить уточнение значения корня можно при достижении условия $|a_j - b_j| < \varepsilon$, где $\varepsilon > 0$ - сколь угодно малое число. Второй способ закончить вычисления - задать максимальное значение невязки: $f((a_j + b_j)/2) < \varepsilon$.

Замечания

- Метод деления отрезка пополам очень прост, здесь нет вычислительной формулы и можно обеспечить практически любую точность.
- Как недостаток метода можно отметить его медленную сходимость (за один шаг интервал, где находится корень, сужается всего в два раза).

Метод хорд

Пусть дано уравнение (3) где функция $f(x)$ определена и непрерывна на интервале $[a, b]$ и выполняется соотношение $f(a)\cdot f(b) < 0$.

Пусть для определенности $f(a) < 0$, $f(b) > 0$. Тогда вместо того, чтобы делить отрезок $[a, b]$ пополам, более естественно разделить его в отношении $-f(a):f(b)$. При этом новое значение корня определяется из соотношения

$$x_1 = a + h_1, \tag{4}$$

где

$$h_1 = \frac{-f(a)}{-f(a)+f(b)}(b-a) = \frac{-f(a)}{f(b)-f(a)}(b-a) \tag{5}$$

Далее этот прием применяем к одному из отрезков $[a, x_1]$ или $[x_1, b]$, на концах которого функция $f(x)$ имеет противоположные знаки. Аналогично находим второе приближение x_2 и т.д. (рис. 2.).

Геометрически этот способ эквивалентен замене кривой $y = f(x)$ хордой, проходящей через точки $A\ (a, f(a))$ и $B\ (b, f(b))$.

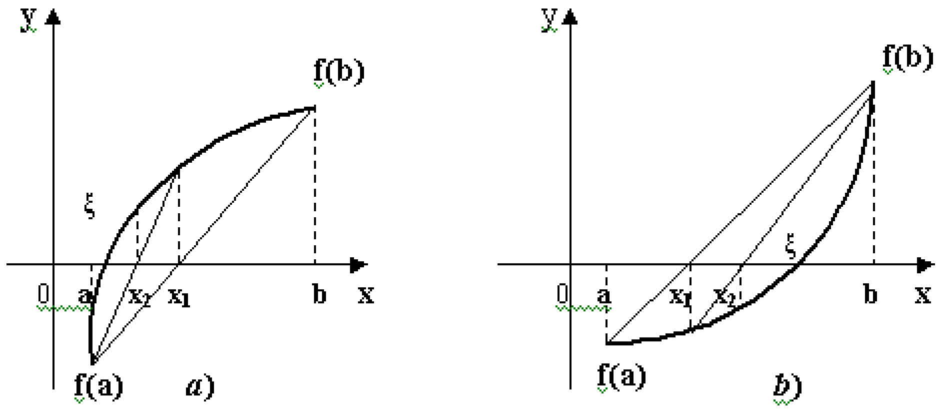

Рис. 2. Уточнение корня уравнения методом хорд

Действительно, уравнение хорды *AB* имеет вид

$$\frac{x-a}{b-a}=\frac{y-f(a)}{f(a)-f(b)}. \tag{6}$$

Учитывая, что при $x = x_1$ => $y = 0$, получим

$$x_1 = a-\frac{f(a)}{f(b)-f(a)}(b-a). \tag{7}$$

Полагая, что на отрезке $[a, b]$ вторая производная $f''(x)$ сохраняет постоянный знак, метод хорд сводится к двум различным вариантам:

1. Из рис. 2, *a* видно, что неподвижна точка *a*, а точка *b* приближается к ξ, то есть

$$X_{n+1} = a-\frac{f(a)}{f(X_n)-f(a)}(X_n-a). \tag{8}$$

   Преобразовав выражение (8), окончательно получим

$$X_{n+1} = X_n-\frac{f(X_n)}{f(X_n)-f(a)}(X_n-a). \tag{9}$$

2. Из рис. 2, *b* видно, что точка *b* остается неподвижной, а точка *a* приближается к ξ, тогда вычислительная формула примет вид

$$X_{n+1} = X_n-\frac{f(X_n)}{f(b)-f(X_n)}(b-X_n). \tag{10}$$

Таким образом, для вычисления корня уравнения имеем две различные вычислительные формулы (9) и (10). *Какую точку брать за неподвижную ?* Рекомендуется в качестве неподвижной выбирать ту точку, в которой выполняется соотношение:

$$f(x)\cdot f''(x) > 0. \tag{11}$$

Метод (касательных) Ньютона

Пусть корень ξ уравнения (3) находится на отрезке [a, b], причем первая и вторая производные $f'(x)$ и $f''(x)$ непрерывны и сохраняют определенные знаки при $a \le x \le b$. Найдя какое-нибудь n-ое приближение корня $x_n \approx \xi \quad (a \le x_n \le b)$, мы можем уточнить его по методу Ньютона следующим образом. Пусть

$$\xi = x_n + h_n, \tag{12}$$

где h_n - величина малая. Отсюда по формуле Тейлора получим (ограничиваясь первым порядком малости относительно h_n)

$$f(x_n + h_n) = f(x_n) + h_n f'(x_n) = 0. \tag{13}$$

Следовательно,

$$h_n = -f(x_n)\,/\,f'(x_n). \tag{14}$$

Подставив полученное выражение в формулу (12), найдем следующее (по порядку) значение корня:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (n = 0, 1, \ldots). \tag{15}$$

Проиллюстрируем графически нахождение корня методом Ньютона (рис. 3.).

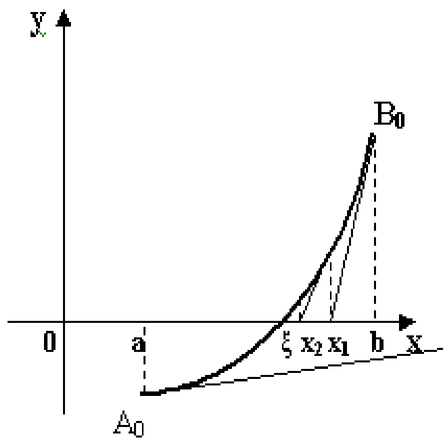

Рис. 3. Уточнение корня методом (касательных) Ньютона

Если в качестве начального приближения выбрать точку $x_0 = B_0$, то процесс быстро сходится. Если же выбрать точку $x_0 = A_0$, то $x_1 \notin [a, b]$, и процесс нахождения корня расходится. Рекомендуется: в качестве x_0 выбрать точку, где $f(x) \cdot f''(x) > 0$.

Комбинированный метод

Пусть $f(a) \cdot f(b) < 0$, а $f'(x)$ и $f''(x)$ сохраняют постоянные знаки на отрезке [a, b]. Соединяя метод хорд и метод касательных, получаем метод, на каждом шаге которого находим значения по недостатку и значения по избытку точного корня ξ уравнения $f(x) = 0$**.** Теоретически здесь возможны четыре случая:

- $f'(x) > 0; f''(x) > 0;$
- $f'(x) > 0; f''(x) < 0;$
- $f'(x) < 0; f''(x) > 0;$
- $f'(x) < 0; f''(x) < 0.$

Рассмотрим только первый случай, так как остальные три ведут себя аналогично и могут быть сведены к первому.

Итак, пусть $f'(x) > 0$ и $f''(x) > 0$ при $a \le x \le b$. Полагаем, что x_0=a (для метода хорд), $\bar{x}_0 = b$ (для метода касательных). Тогда новые значения корня вычисляем по формулам

$$x_{n+1} = x_n - \frac{f(x_n)}{f(\bar{x}_n) - f(x_n)}(\bar{x}_n - x_n);$$
$$\bar{x}_{n+1} = \bar{x}_n - \frac{f(\bar{x}_n)}{f'(\bar{x}_n)} \qquad (n = 0, 1, 2, \ldots). \tag{16}$$

Рис. 4 наглядно иллюстрирует суть комбинированного метода.

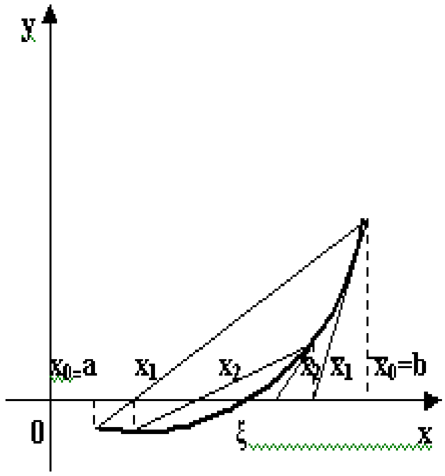

Рис. 4. Уточнение корня комбинированным методом

Доказано, что $x_n < \xi < \bar{x}_n$. Следует обратить внимание на то, что на каждом шаге метод хорд применяется к новому отрезку $[x_n, \bar{x}_n]$. Если задать максимальное значение погрешности $\varepsilon > 0$ **,** процесс уточнения значения корня продолжаем до тех пор, пока не выполнится условие

$$f((x_n + \bar{x}_n)/2) \le \varepsilon. \tag{17}$$

П р и м е р. Вычислить с точностью до 0.0005 положительный корень уравнения

$$f(x) = x^5 - x - 0.2 = 0.$$

На первом этапе отделения корней выбрали интервал [1.0, 1.1], на концах которого функция имеет противоположные знаки. Действительно,

$$f(1) = -0.2 < 0, f(1.1) = 0.31051 > 0.$$

В выбранном нами интервале $f''(x) > 0$, $f''(x) > 0$, то есть знаки производных сохраняются.

Применим комбинированный метод, приняв $x_0 = 1.0, \quad \bar{x}_0 = 1.1$**.** По формулам (16) вычислим

$x_1 = 1.03917651; \quad \bar{x}_1 = 1.050872558;$

$f((x_1 + \bar{x}_1)/2) = 0.0013037 > 0.0005.$

Так как точность недостаточная (погрешность велика), вычислим следующие значения:

$$x_2 = 1.044683244, \quad \bar{x}_2 = 1.044846218,$$

$$f((x_2 + \bar{x}_2)/2) = 0.0000150248 < 0.0005.$$

Таким образом, за два шага мы обеспечили требуемую точность.

Замечания

- Комбинированный метод наиболее трудоемок.
- Метод, как и метод Ньютона не всегда сходится (почему?).
- Комбинированный метод сходится быстрее всех ранее рассмотренных, (если он сходится).

Вопросы для самоконтроля

- Какие точные методы решения нелинейных уравнений вы знаете?
- Для чего нужен первый этап - отделение корней?
- Сформулируйте условия существования решения уравнения. Являются ли эти требования необходимыми и достаточными?
- Что можно сказать о точности методов половинного деления, хорд, касательных и комбинированного? По каким параметрам еще можно сравнить эти методы?
- В соответствии с известной теоремой на отрезке **[***a, b***]** существует решение. Всегда ли его можно найти методом половинного деления, методом хорд, и т.п.?