МЕТОД ДЕЛЕНИЯ ОТРЕЗКА ПОПОЛАМ

Простейшим методом минимизации функции одной переменной, не требующим вычислений производной является метод деления отрезка пополам. Поиск минимума функции начинается с выбора на отрезке $[a,b]=[a_1,b_1]$ двух точек

$$x_1=\frac{a+b-\delta}{2} \text{ и } x_2=\frac{a+b+\delta}{2}=a+b-x_1,$$

где δ — постоянная величина, являющаяся параметром метода и $0<\delta<b-a$. Величина δ выбирается вычислителем и может характеризовать ошибку измерения величины x. Точки x_1,x_2 расположены симметрично на отрезке $[a,b]$ относительно его середины и при малых значениях δ делят его почти пополам – этим и объясняется название метода.

После выбора точек x_1,x_2 вычисляются значения функции $f(x_1),f(x_2)$ и сравниваются между собой. Если $f(x_1)\le f(x_2)$, то полагают $a_2=a_1,b_2=x_2$, если же $f(x_1)>f(x_2)$, то полагают $a_2=x_1,b_2=b_1$ (рис. 2.2).

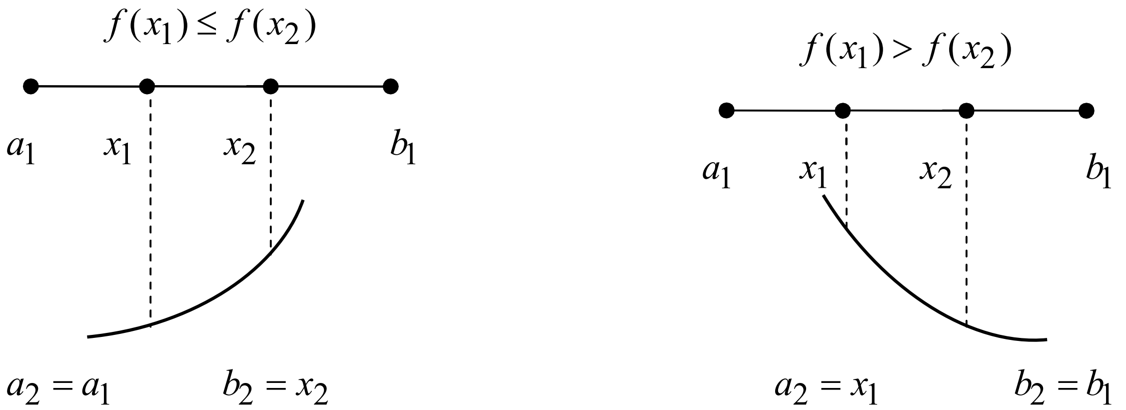

Рисунок 1.2

Так функция $f(x)$ унимодальная на $[a,b]$, то ясно, что отрезок $[a_2,b_2]$ будет содержать точку минимума и его длина будет равна

$$b_2-a_2=\frac{b-a-\delta}{2^1}+\delta. \tag{1.3}$$

Пусть отрезок $[a_k,b_k]$, содержащий точку минимума $\hat{x}$ известен, и имеет длину

$$b_k-a_k=\frac{b-a-\delta}{2^{k-1}}+\delta>\delta,\ k\ge 2. \tag{1.4}$$

Тогда берем точки $x_{2k-1}=\dfrac{a_k+b_k-\delta}{2}$ и $x_{2k}=\dfrac{a_k+b_k+\delta}{2}=a_k+b_k-x_{2k-1}$, расположенные на отрезке $[a_k,b_k]$, симметрично относительно его середины и вычисляем значения $f(x_{2k-1})$ и $f(x_{2k})$.

Если $f(x_{2k-1}) \le f(x_{2k})$, то полагаем $a_{k+1} = a_k, b_{k+1} = x_{2k}$,

если же $f(x_{2k-1}) > f(x_{2k})$, то полагаем $a_{k+1} = x_{2k-1}, b_k + 1 = b_k$.

Длина полученного отрезка $[a_{k+1}, b_{k+1}]$ будет равна

$$b_{k+1} - a_{k+1} = \frac{b-a-\delta}{2^k} + \delta > \delta .$$

Отрезок $[a_{k+1}, b_{k+1}]$ будет содержать точку минимума функции $f(x)$. Если количество вычислений значений целевой функции заранее ничем не ограничено, то описанный процесс деления отрезка пополам можно продолжать до тех пор, пока мы не получим отрезок $[a_{k+1}, b_{k+1}]$, длина которого $b_{k+1} - a_{k+1} < \varepsilon$, где ε — заданная точность и $\varepsilon > \delta$.

Отсюда получаем, что количество итераций $k > \log_2 \frac{b-a-\delta}{\varepsilon-\delta}$. Поскольку каждое деление пополам требует вычисления двух значений функции, то для достижения точности $b_{k+1} - a_{k+1} < \varepsilon$ требуется $n = 2k > 2\log_2 \frac{b-a-\delta}{\varepsilon-\delta}$ таких вычислений.

После определения отрезка $[a_{k+1}, b_{k+1}]$ в качестве точки минимума можно взять точку $\overline{x}_n = x_{2k-1}$, если $f(x_{2k-1}) \le f(x_{2k})$, и $\overline{x}_n = x_{2k}$, если $f(x_{2k-1}) > f(x_{2k})$, а значение $f(\overline{x}_n)$ может служить приближением для значения $f(\hat{x}) = \min\limits_{[a,b]} f(x)$. При таком выборе приближения для точки минимума будет допущена ошибка

$$|x_* - \overline{x}_n| \le \max\{|b_{k+1} - \overline{x}_n|, |a_{k+1} - \overline{x}_n|\} = \frac{b-a-\delta}{2^k}.$$

Если нет необходимости вычислять значение функции именно в точке $\overline{x}_n$, можно взять точку $z_n = \frac{a_{k+1} + b_{k+1}}{2}$ с погрешностью

$$|x_* - z_n| \le \frac{b_{k+1} - a_{k+1}}{2} = \frac{b-a-\delta}{2^{k+1}} + \frac{\delta}{2}. \tag{1.5}$$

Из выше сказанного следует, что методом деления отрезка пополам с помощью $n = 2k$ вычислений можно определить точку минимума унимодальной функции на отрезке $[a, b]$ с точностью $\approx \frac{b-a}{2^{1+n/2}}$. Следует отметить, что метод дихотомии можно применять для минимизации функций не являющихся унимодальными. Однако в этом случае нельзя гарантировать, что найденное решение будет хорошим приближением к глобальному минимуму.

МЕТОД ЗОЛОТОГО СЕЧЕНИЯ

Как известно, ***золотым сечением*** отрезка называется деление отрезка на две неравные части, так чтобы отношение длины всего отрезка к длине большей части равнялось отношению длины большей части к длине меньшей части отрезка.

Непосредственным образом можно убедиться в том, что золотое сечение отрезка $[a,b]$ производится двумя точками

$$x_1 = a + \frac{3-\sqrt{5}}{2}(b-a) = a + 0{,}381966011...(b-a) \tag{1.6}$$

$$x_2 = a + \frac{\sqrt{5}-1}{2}(b-a) = a + 0{,}618033989...(b-a), \tag{1.7}$$

расположенными симметрично относительно середины отрезка, причем $a < x_1 < x_2 < b$ и $x_1 + x_2 = a + b$, $\dfrac{b-a}{b-x_1} = \dfrac{b-x_1}{x_1-a} = \dfrac{b-a}{x_2-a} = \dfrac{x_2-a}{b-x_2} = \dfrac{\sqrt{5}+1}{2} = 1{,}618033989...$ (рис. 1.3)

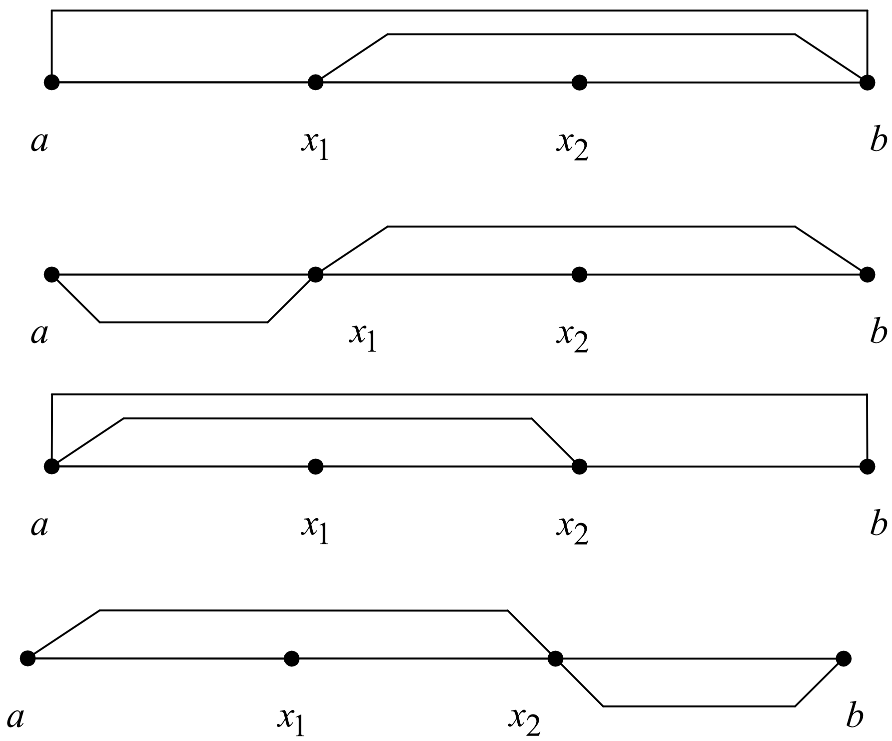

Рис. 1.3

Замечательно, здесь то, что точка x_1 в свою очередь производит золотое сечение отрезка $[a,x_2]$, а точка x_2 производит золотое сечение отрезка $[x_1,b]$. Опираясь на это свойство золотого сечения можно предложить следующий метод минимизации унимодальной функции $f(x)$ на отрезке $[a,b]$.

Положим $a_1 = a, b_1 = b$. На отрезке $[a_1,b_1]$ возьмем точки x_1, x_2, которые осуществляют золотое сечение данного отрезка, то есть определяются по формулам (2.4), (2.5). Вычислим значения целевой функции $f(x_1), f(x_2)$ и сравним между собой. Если

$f(x_1) \le f(x_2)$, то полагают $a_2 = a_1, b_2 = x_2, \bar{x}_2 = x_1$, если же $f(x_1) > f(x_2)$, то полагают $a_2 = x_1, b_2 = b_1, \bar{x}_2 = x_2$ (рис. 1.4). Так функция $f(x)$ строго унимодальная на $[a,b]$, то ясно, что отрезок $[a_2, b_2]$ будет содержать точку минимума x_* и его длина будет равна

$$b_2 - a_2 = \left(\frac{\sqrt{5}-1}{2}\right)^1 (b-a). \qquad (1.8)$$

Весьма важно то, что построенный отрезок $[a_2, b_2]$ будет содержать точку $\bar{x}_2$ с вычисленным значением $f(\bar{x}_2) = \min\{f(x_1), f(x_2)\}$, которая производит золотое сечение отрезка $[a_2, b_2]$ и является приближенным значением точки минимума x_*.

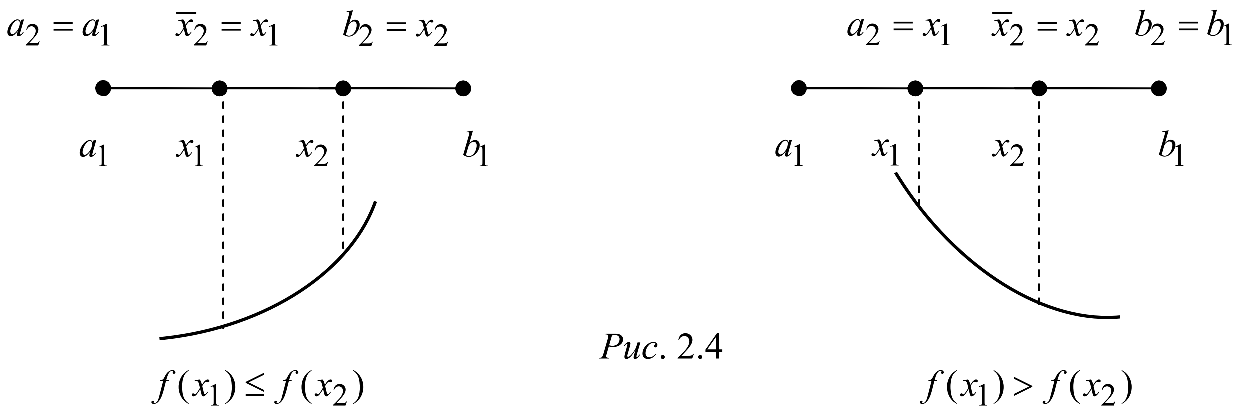

Рис. 2.4

На следующем шаге, на отрезке $[a_2, b_2]$ строим точку $x_3 = a_2 + b_2 - \bar{x}_2$, которая также будет осуществлять золотое сечение отрезка. Однако непосредственное применение этой формулы, приводит к тому, что погрешность, найденная при вычислении точки $\bar{x}_2$ и связанная с приближенным вычислением числа $\sqrt{5}$, переносится с предыдущего шага на этот шаг, что в дальнейшем ведет к накоплению ошибки и даже при достаточно малых n приводит к результатам, далеким от истинных. Поэтому рекомендуется на этом и всех последующих этапах непосредственно проводить золотое сечение отрезка $[a_n, b_n], n \ge 2$ и в качестве точки x_{n+1} брать ту из точек $a_n + \frac{3-\sqrt{5}}{2}(b_n - a_n)$ или $a_n + \frac{\sqrt{5}-1}{2}(b_n - a_n)$, которая наиболее удалена от точки $\bar{x}_n$. Данную формулу будем использовать только для теоретических выкладок.

Для определенности положим, что $a_2 < x_3 < \bar{x}_2 < b_2$ и вычислим значение целевой функции $f(x_3)$. Если

1) $f(x_3) \le f(\bar{x}_2)$, то полагаем $a_3 = a_2, b_3 = \bar{x}_2, \bar{x}_3 = x_3$;

2) $f(x_3) > f(\bar{x}_2)$, то полагают $a_3 = x_3, b_3 = b_2, \bar{x}_3 = \bar{x}_2$.

Длина полученного отрезка будет равна: $b_3 - a_3 = \left(\frac{\sqrt{5}-1}{2}\right)^2 (b-a)$, точка $\overline{x}_3$ осуществляет золотое сечение отрезка $[a_3, b_3]$ и

$$f(\overline{x}_3) = \min\{f(x_3), f(\overline{x}_2)\} = \min\{f(x_3), \min\{f(x_1), f(x_2)\}\} = \min_{1 \le i \le 3}\{f(x_i)\},$$

то есть $\overline{x}_3$ является приближенным значение точки минимума x_*.

Далее, на отрезке $[a_3, b_3]$ строим точку $x_4 = a_3 + b_3 - \overline{x}_3$ и продолжаем процесс.

Предположим, что точки $x_1, ..., x_{n-1}$ найдены и вычислены соответствующие значения $f(x_1), ..., f(x_{n-1})$, найден отрезок $[a_{n-1}, b_{n-1}]$, содержащий точку минимума $\hat{x}$, длина которого равна $b_{n-1} - a_{n-1} = \left(\frac{\sqrt{5}-1}{2}\right)^{n-2} (b-a)$, и известна точка $\overline{x}_{n-1}$, производящая золотое сечение отрезка $[a_{n-1}, b_{n-1}]$ и такая, что $f(\overline{x}_{n-1}) = \min\limits_{1 \le i \le n-1}\{f(x_i)\}$.

Тогда в качестве следующей точки возьмем точку $x_n = a_{n-1} + b_{n-1} - \overline{x}_{n-1}$, которая также осуществляет золотое сечение отрезка $[a_{n-1}, b_{n-1}]$, и вычислим значение целевой функции $f(x_n)$. Пусть для определенности $a_{n-1} < x_n < \overline{x}_{n-1} < b_{n-1}$. Случай, когда $a_{n-1} < \overline{x}_{n-1} < x_n < b_{n-1}$ рассматривается аналогично.

Возможны два варианта (рис. 1.5):

1) если $f(x_n) \le f(\overline{x}_{n-1})$, то полагаем $a_n = a_{n-1}$, $\overline{x}_n = x_n$, $b_n = \overline{x}_{n-1}$.

2) если $f(x_n) > f(\overline{x}_{n-1})$, то полагаем $a_n = x_n$, $\overline{x}_n = \overline{x}_{n-1}$, $b_n = b_{n-1}$.

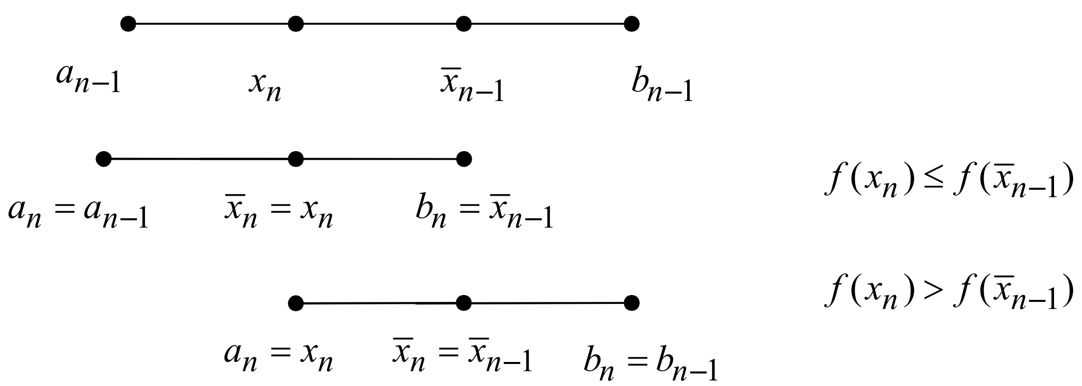

Рис. 1.5

Длина полученного отрезка $[a_n, b_n]$ равна $b_n - a_n = \left(\frac{\sqrt{5}-1}{2}\right)^{n-1} (b-a)$, и точка $\overline{x}_n$ осуществляет золотое сечение отрезка $[a_n, b_n]$ и

$$f(\overline{x}_n) = \min\{f(x_n), f(\overline{x}_{n-1})\} = \min_{1 \le i \le n} f(x_i). \qquad (1.9)$$

Если количество вычислений значений целевой функции $f(x)$ заранее не ограничено, то описанный процесс можно продолжать, например, до тех пор, пока не будет выполняться неравенство $b_n - a_n < \varepsilon$, где $\varepsilon -$ заданная точность. Если число вычислений значений целевой функции $f(x)$ задано и равно n, то процесс заканчивается и в качестве решения берется либо точка $\overline{x}_n$ с погрешностью

$$\left|x_* - \overline{x}_n\right| \le \max\left\{|b_n - \overline{x}_n|, |a_n - \overline{x}_n|\right\} = \frac{\sqrt{5}-1}{2}(b_n - a_n) = \left(\frac{\sqrt{5}-1}{2}\right)^n \left(b-a\right) = A_n, \quad (1.10)$$

либо если не требуется вычислять значения целевой функции $f(x)$ в точке $\overline{x}_n$, точка $z_n = \dfrac{a_n + b_n}{2}$ будет приближенным значением точки минимума с погрешностью

$$\left|x_* - z_n\right| \le \left(\frac{\sqrt{5}-1}{2}\right)^{n-1} \frac{\left(b-a\right)}{2}. \quad (1.11)$$

Вспомним, что с помощью метода деления отрезка пополам путем $n = 2k$ вычислений значений функции $f(x)$ в аналогичном случае мы получили точку $\overline{x}_n$ с погрешностью $\left|x_* - \overline{x}_n\right| \le 2^{-n/2}(b-a) = B_n$. Отсюда имеем $\dfrac{A_n}{B_n} = \left(\dfrac{2\sqrt{2}}{\sqrt{5}+1}\right)^n \approx (0{,}87...)^n$.

Ясно, что уже при небольших значениях n преимущество метода золотого сечения перед методом деления пополам становится ощутимым.

МЕТОД ФИБОНАЧЧИ

Подобно методу золотого сечения, метод Фибоначчи требует двух вычислений целевой функции на первой итерации, а на каждой последующей итерации только по одному вычислению значения целевой функции. Вспомним, что в методе дихотомии на каждой итерации производилось два вычисления значения целевой функции. Однако в отличие от метода дихотомии и метода золотого сечения, метод Фибоначчи требует предварительного задания числа n вычислений значений целевой функции.

Метод Фибоначчи (обозначается через Φ_n) использует известные числа Фибоначчи, которые определяются соотношениями

$$F_{n+2} = F_{n+1} + F_n, \quad n = 1,2,...; \quad F_1 = F_2 = 1. \tag{1.12}$$

При помощи математической индукции можно легко показать, что $n-$тое число Фибоначчи, представимо следующим образом

$$F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right). \tag{1.13}$$

Метод Фибоначчи является последовательным методом и относится к классу симметрических методов. Положим $a_1 = a, b_1 = b$.

Пусть $n = 1$. Получаем метод Φ_1. Выбираем точку $x_1 = \frac{a_1 + b_1}{2}$. В качестве приближения точки минимума x_* берем точку $\bar{x}_1 = x_1$, допуская при этом ошибку

$$\left|\hat{x} - \bar{x}_1\right| \le \frac{b_1 - a_1}{2} = \frac{b-a}{F_3}.$$

Пусть $n = 2$. Тогда метод Φ_2 начинается с выбора двух точек

$$x_1 = a_1 + \frac{F_n}{F_{n+2}}(b_1 - a_1); \quad x_2 = a_1 + \frac{F_{n+1}}{F_{n+2}}(b_1 - a_1) = a_1 + b_1 - x_1,$$

симметричных друг другу относительно середины отрезка $[a_1, b_1]$.

Если $f(x_1) \le f(x_2)$, то полагаем $a_2 = a_1, b_2 = x_2, \bar{x}_2 = x_1$. Если же $f(x_1) > f(x_2)$, то полагаем $a_2 = x_1, b_2 = b_1, \bar{x}_2 = x_2$. В результате получаем отрезок $[a_2, b_2]$, содержащий точку минимума функции $f(x)$ на отрезке $[a_1, b_1]$ и имеющий длину

$$b_2 - a_2 = b_1 - x_1 = x_2 - a_1 = \frac{F_{n+1}}{F_{n+2}}(b_1 - a_1).$$

Точка $\bar{x}_2$ является приближенным значением точки минимума x_* и обладает такими свойствами:

1) $a_2 < \bar{x}_2 < b_2$ по способу выбора точки $\bar{x}_2$; 2) $f(\bar{x}_2) = \min\{f(x_1), f(x_2)\}$;

3) $\bar{x}_2$ совпадает с одной из точек $x_1 \vee x_2$ на отрезке $[a_1, b_1]$;

4) $\bar{x}_2$ совпадает с одной из точек $x_2' \vee x_2''$ на отрезке $[a_2, b_2]$, симметричных друг другу относительно середины отрезка $[a_2, b_2]$, то есть $x_2'' = a_2 + b_2 - x_1'$ и

$$x_2' = a_2 + \frac{F_n}{F_{n+1}}(b_2 - a_2) = a_2 + \frac{F_n}{F_{n+2}}(b_1 - a_1), \text{ если } f(x_1) \le f(x_2)$$

$$x_2'' = a_2 + \frac{F_{n-1}}{F_{n+1}}(b_2 - a_2) = a_2 + \frac{F_{n-1}}{F_{n+2}}(b_1 - a_1), \text{ если } f(x_1) > f(x_2).$$

Потом на отрезке $[a_2, b_2]$ выбираем точку $x_3 = a_2 + b_2 - \bar{x}_2$, которая совпадает с одной из точек $x_1' \vee x_2''$, отличной от точки $\bar{x}_2$ (см. рис. 1.6).

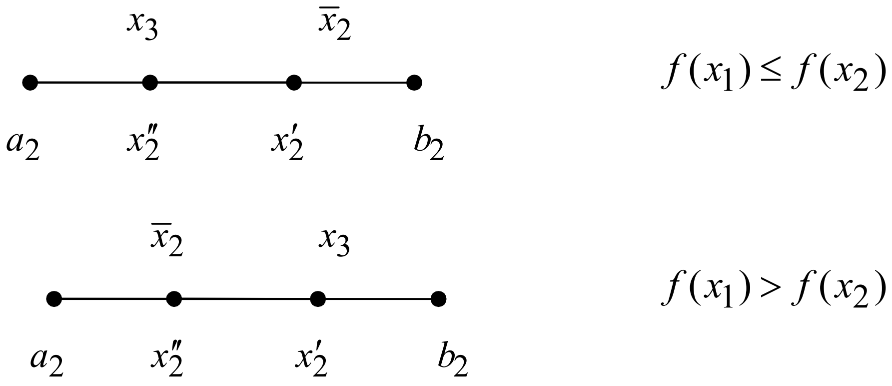

Рис. 1.6

Сравниваем значения $f(x_3)$ и $f(\bar{x}_2)$, находим новый отрезок $[a_3, b_3]$ и т.д.

В общем случае, будем считать, что нам известны точки $x_1, x_2, ..., x_k$ $(2 \le k \le n)$ и найден отрезок $[a_k, b_k]$, имеющий длину

$$b_k - a_k = \frac{F_{n-k+3}}{F_{n+2}}(b_1 - a_1)$$

и содержащий точку минимума функции $f(x)$ на отрезке $[a, b]$ и точку $\bar{x}_k$, которая имеет следующие свойства:

1) $a_k < \bar{x}_k < b_k$

2) $f(\bar{x}_k) = \min\{f(x_k), f(\bar{x}_{k-1})\} = \min\limits_{1 \le i \le k} f(x_i)$;

3) $\bar{x}_k$ совпадает с одной из точек $x_1 \vee x_2 \vee x_3 \vee ... \vee x_k$ на отрезке $[a_1, b_1]$;

4) – $\bar{x}_k$ совпадает сточкой

$$x_k' = a_k + \frac{F_{n-k+2}}{F_{n-k+3}}(b_k - a_k) = a_k + \frac{F_{n-k+2}}{F_{n+2}}(b_1 - a_1), \qquad (1.14)$$

если $a_k = a_{k-1}$, то есть отрезок $[a_k, b_k]$ смещен на отрезке $[a_{k-1}, b_{k-1}]$в сторону левого конца;

– $\bar{x}_k$ совпадает сточкой

$$x_k'' = a_k + \frac{F_{n-k+1}}{F_{n-k+3}}(b_k - a_k) = a_k + \frac{F_{n-k+1}}{F_{n+2}}(b_1 - a_1), \qquad (1.15)$$

если $b_k = b_{k-1}$, то есть отрезок $[a_k, b_k]$ смещен на отрезке $[a_{k-1}, b_{k-1}]$в сторону правого конца.

Точки x_k', x_k'' симметричны друг другу относительно середины отрезка $[a_k, b_k]$, то есть $x_k'' = a_k + b_k - x_k'$.

Если $k < n$ на отрезке $[a_k, b_k]$ выбираем точку $x_{k+1} = a_k + b_k - \overline{x}_k$, которая будет симметрична точке $\overline{x}_k$ и совпадать с одной из точек x_k', x_k'', отличной от $\overline{x}_k$. Далее продолжаем процесс до тех пор, пока количество итераций не будет исчерпано.

Если $k = n$. Тогда заканчиваем процесс и получаем точку

$$x_n' = x_n'' = \overline{x}_n = a_n + \frac{b-a}{F_{n+2}},$$

которая совпадает с серединой отрезка $[a_n, b_n]$, длина которого равна

$$b_n - a_n = \frac{2(b-a)}{F_{n+2}}.$$

В качестве решения берем точку $\overline{x}_n$ с погрешностью

$$\left| x_* - \overline{x}_n \right| < \varepsilon \le \frac{b-a}{F_{n+2}}. \tag{1.16}$$

Отметим следующий факт.

Теорема. При всех $n > 1$ оптимального последовательного метода на классе унимодальных функций не существует. Наилучшая гарантированная точность последовательных методов на этом классе равна $\dfrac{b-a}{F_{n+2}}$. В качестве $\varepsilon-$оптимального метода можно взять метод Фибоначчи Φ_n.

Рекомендации к численной реализации метода Фибоначчи на ЭВМ

Отметим, что число $\dfrac{F_n}{F_{n+2}}$ является бесконечной периодической дробью. Поэтому первая точка x_1 метода Фибоначчи будет задаваться приближенно. Ошибка при задании первой точки метода Φ_n приводит к быстрому росту ошибки на последующих этапах и уже при не очень больших n результаты будут отличаться от верных. Поэтому на практике на каждом отрезке $[a_k, b_k]$, который содержит точку $\overline{x}_k$ из предыдущего шага, необходимо остерегаться применять формулу $x_{k+1} = a_k + b_k - \overline{x}_k$, так как ошибка с предыдущего шага переносится на этот шаг. Вместо этого рекомендуется непосредственно вычислять точки x_k', x_k'' и в качестве точки x_{k+1} взять ту точку, которая наиболее удалена от точки $\overline{x}_k$.

Чаще всего на практике известна длина отрезка, на котором вычисляется точка минимума и точность вычислений $\varepsilon > 0$. Количество итераций, необходимых для достижения заданной точности, равно наименьшему целому из чисел n, удовлетворяющих неравенству

$$\frac{b-a}{F_{n+2}} \le \varepsilon \le \frac{b-a}{F_{n+1}} \quad \text{или} \quad F_{n+1} \le \frac{b-a}{\varepsilon} \le F_{n+2}.$$

Отметим, что длина отрезка $[a,b]$, на котором можно найти точку $\hat{x}$ с заданной точностью $\varepsilon > 0$, выполнив при этом n вычислений значений целевой функции $f(x)$, не должна превосходить $\varepsilon \cdot F_{n+2}$.

Сравнение методов линейного поиска

Пусть длина исходного отрезка равна $b_1 - a_1$. При заданной длине последнего отрезка $b_n - a_n$, которая удовлетворяет указанной точности $\varepsilon > 0$, необходимое количество вычислений значений целевой функции n может быть определено как наименьшее целое положительной число, удовлетворяющее следующим соотношениям:

1) метод деления отрезка пополам $\left(\frac{1}{2}\right)^{\frac{n}{2}} \le \frac{b_n - a_n}{b_1 - a_1}$,

2) метод золотого сечения $\left(\frac{\sqrt{5}-1}{2}\right)^{n-1} \le \frac{b_n - a_n}{b_1 - a_1}$,

3) метод Фибоначчи $F_n \ge \frac{b_1 - a_1}{b_n - a_n}$.

Для фиксированного значения отношения $\frac{b_n - a_n}{b_1 - a_1}$ наименьшее число вычислений целевой функции отвечает наиболее эффективному алгоритму. Поэтому, с данной точки зрения, наиболее эффективным алгоритмом является метод Фибоначчи, потом метод золотого сечения, и наконец, метод деления пополам. Следует отметить, что для достаточно больших n значение $\frac{1}{F_n} \to \left(\frac{\sqrt{5}-1}{2}\right)^{n-1}$, так что методы Фибоначчи и золотого сечения практически равноценны.