

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ
КУРГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

КАФЕДРА «ИНФОРМАТИКА»

ОДНОМЕРНАЯ ОПТИМИЗАЦИЯ

Методические указания
для выполнения лабораторных работ
по курсу «Информатика», «Методы вычислений»
для студентов
по направлениям 190000, 200000, 260000, 280000, 010000
специальностей 140211, 150202, 151001, 150002
192001, 190601, 190603, 190702, 200503, 260601, 280101, 010101

КУРГАН 2008

Кафедра: «Информатика»

Дисциплины: «Информатика», «Методы вычислений»

(по направлениям 190000, 200000, 260000, 280000, 010000
специальностей 140211, 150202, 151001, 150002
192001, 190601, 190603, 190702, 200503, 260601, 280101, 010101)

Составили: доцент В.Я. Котликова

ст.преподаватель Л.Г. Сысолятина

Утверждены на заседании кафедры « 26 » июня 2008 г.

Рекомендованы методическим советом университета

« 9 » июля 2008 г.

1. Общие сведения

Экономическое планирование, проектирование сооружений и многая другая творческая деятельность инженера связана с поиском рационального решения поставленной задачи. При этом оптимальным называют наилучший в определенном смысле вариант из всех возможных.

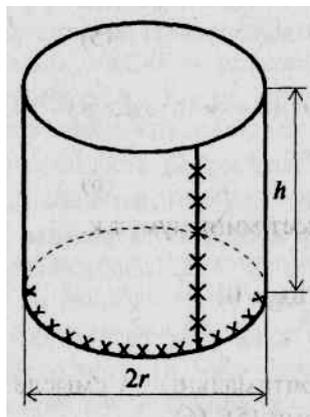


Рис. 1.1

Предположим, что перед нами поставлена задача спроектировать открытый цилиндрический резервуар (рис. 1.1) емкостью V , где V - регламентированный параметр. Очевидно, что при этом принципиально допустимы различные соотношения радиуса r и высоты h изделия - проектных параметров.

Этим обстоятельством можно воспользоваться и наложить какое-либо полезное дополнительное условие задачи. Например, поставить цель сэкономить материал, т.е. получить резервуар, имеющий минимальную поверхность S , т.е. спроектировать изделие, наилучшее в смысле минимума S .

Для этого, прежде всего, выразим S через проектные параметры

$$S = \pi r^2 + 2\pi r h. \quad (1.1)$$

Тем самым, получена целевая функция, минимизация которой позволит получить искомое оптимальное решение.

Заметим, что проектные параметры r, h взаимосвязаны:

$$V = \pi r^2 h. \quad (1.2)$$

Используя соотношение (1.2), исключим в выражении (1.1) один из проектных параметров, например, h :

$$S = \pi r^2 + \frac{2V}{r} \quad (1.3)$$

Таким образом, целевая функция (1.3) фактически зависит от одного проектного параметра и, тем самым, задача оптимизации в нашем примере свелась к поиску минимума функции одного независимого переменного, т.е. к задаче одномерной оптимизации.

Для ее решения воспользуемся классическим методом поиска экстремума непрерывно дифференцируемой функции, т.е. условием

$$\frac{dS}{dr} = 0 \quad (1.4)$$

или

$$\pi r^2 - \frac{V}{r^2} = 0.$$

Отсюда найдем стационарную точку

$$r_1 = \sqrt[3]{\frac{V}{\pi}}, \quad (1.5)$$

Пользуясь соотношением (1.2), получим соответствующее r_1 значение второго проектного параметра:

$$h_1 = r_1. \quad (1.6)$$

Функция S при $r=r_1$ действительно имеет минимум, т.к.

$$\left. \frac{d^2 S}{dr^2} \right|_{r=r_1} = \left(2\pi + \frac{4V}{r^3} \right)_{r=r_1} = 6\pi > 0.$$

Таким образом, параметры резервуара, оптимальные в смысле минимума материала, определяются формулами (1.5), (1.6).

Если при проектировании изделия мы поставим цель сделать минимальным объем сварочных работ, т.е. получить минимальную длину L сварочных швов

$$L = 2\pi r + h, \quad (1.7)$$

то, действуя аналогично предыдущему, получим другое решение

$$r_2 = \sqrt[3]{\frac{V}{\pi^2}} \quad (1.8)$$

$$h_2 = \pi r_2. \quad (1.9)$$

Наконец, можно стремиться и к минимуму себестоимости изделия, принимая ее в виде

$$Q = q_1 S + q_2 L \quad (1.10)$$

или

$$Q = q_1(\pi r^2 + 2\pi r h) + q_2(2\pi r + h), \quad (1.11)$$

где q_1 – стоимость квадратной единицы листового материала, q_2 – стоимость выполнения единицы длины сварного шва. В этом случае оптимальный радиус резервуара можно определить из следующего алгебраического уравнения (при $q_1 \neq 0$):

$$\pi^2 r^3 (r + \alpha) - V(\pi r + \alpha), \quad (1.12)$$

где $\alpha = q_2/q_1$.

Приведенные примеры показывают, что результат оптимизации существенным образом зависит от того, какой смысл вкладывается в само понятие оптимальности.

Следует заметить, что оптимизация всякого реального объекта обычно сопряжена с наличием ряда дополнительных условий, т.е. может оказаться *задачей с ограничениями*. Например, в задании на проектирование открытого резервуара с минимальной поверхностью может быть указано, что под изделие отводится помещение площадью не более, чем $d \times d$, где d – регламентируемый параметр. Очевидно, что в этом случае задача сводится к поиску минимума функции (1.3) с дополнительным условием

$$0 < r < d/2. \quad (1.13)$$

Пусть $V = 27 \text{ м}^3$. Тогда в соответствии с формулой (1.5)

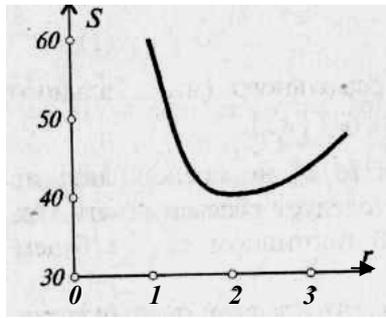


рис. 1.2

$r_1 = \sqrt[3]{\frac{27}{\pi}} \cong 2,048$. При этом заметим, что функция (1.3) (рис.1.2) с

увеличением r при $r < r_1$ – монотонно убывает, а при $r > r_1$ – монотонно возрастает. В связи с этим, в зависимости от величины d решение поставленной задачи будет определяться следующим образом:

$$\begin{aligned} r &= r_1 \text{ при } d/2 \geq r_1, \\ r &= d/2 \text{ при } d/2 < r_1. \end{aligned}$$

В частности, при $d=3$ м будет $r = 1,5$ м, $h \cong 3,820$ м, $S \cong 43,069$ м². В то же время при $r=r_1$ было бы $S \cong 39,454$ м². При решении некоторых сложных проблем не представляется возможным свести задачу к оптимизации одной целевой функции, т.е. имеет место *задача с неопределенностью целей*. Например, при проектировании резервуара можно стремиться к уменьшению как поверхности изделия, так и трудоемкости сварочных работ, т.е. минимизировать по r функции (1.3) и (1.7) одновременно. Однако, как видно из решений (1.5), (1.6) и (1.8), (1.9), их минимумы не будут совпадать.

В этом случае прибегают к методам *системного анализа*, где для разработки стратегии решения даются необходимые рекомендации. Например, в нашей задаче можно пойти по пути построения единой функции цели в виде *линейной свертки* (1.10) функций (1.3) и (1.7), где q_1 и q_2 – некоторые *весовые коэффициенты*. При этом q_1 и q_2 могут быть, например, каким-либо способом прогнозируемые стоимости материала и сварочных работ к моменту начала строительства.

В отличие от рассмотренных выше примеров, реальные инженерные задачи оптимизации сводятся к поиску экстремума целевой функции, зависящей от нескольких независимых переменных, т.е. являются задачами *многомерной оптимизации*.

2. Одномерная оптимизация

Дадим следующую математическую формулировку задачи одномерной оптимизации.

Пусть на отрезке $[a, b]$ задана целевая функция $f(x)$. Требуется найти точку

$$x^* \in [a, b], \quad (2.1)$$

в которой $f(x)$ достигает своего экстремального (минимального или максимального) значения и вычислить $f(x^*)$.

Следует заметить, что если отрезок $[a, b]$ не представляет из себя всю числовую ось, то условие (2.1) следует рассматривать как введенное нами ограничение задачи. В противном случае будем иметь задачу без ограничений.

Для определенности можно считать, что задача оптимизации сводится к поиску минимума $f(x)$, т.к. нахождение максимума всякой функции $\varphi(x)$ всегда можно свести к определению минимума функции $f(x) = -\varphi(x)$.

Заметим, что точку x_* , удовлетворяющую неравенству $f(x_*) < f(x)$ на $[a, b]$ при условии (2.1), называют точкой наименьшего значения функции или точкой *глобального минимума* на отрезке $[a, b]$. В то же время на том же отрезке могут существовать и точки *локального минимума*. Точку x_{**} называют точкой локального минимума, если существует такое число $\varepsilon > 0$, что для всех $x \in (x_{**} - \varepsilon, x_{**} + \varepsilon)$ будет $f(x_{**}) \leq f(x)$.

Очевидно, что в задаче оптимизации отыскивается точка и величина глобального минимума.

Например, при $f(x) = \sin x$ на отрезке $[0; 2\pi]$ в точке $x = 3\pi/2$ глобальный минимум, равный -1 , будет совпадать с локальным минимумом (рис. 2.1).

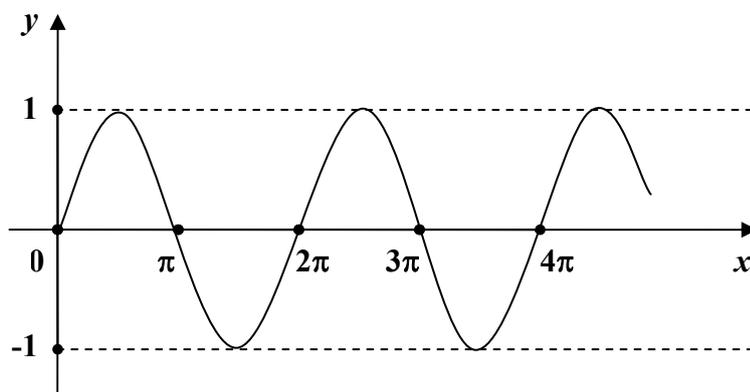


Рис.2.1

Заметим, что на отрезках $[0; \pi]$ и $[0; 4\pi]$ будет по два глобальных минимума в точках $0, \pi$ (на $[0; \pi]$) и $3\pi/2, 7\pi/2$ (на $[0; 4\pi]$). Тем самым в обоих случаях задача оптимизации не имеет однозначного решения.

Задачу одномерной оптимизации можно решить, используя *классический подход*. Действительно, пусть $f(x)$ имеет непрерывную производную $f'(x)$ на $[a, b]$. Тогда каждый корень уравнения

$$f'(x) = 0, \quad (2.2)$$

может быть точкой локального экстремума (минимума или максимума) или точкой перегиба функции $f(x)$. Пусть далее X -множество чисел, состоящее из точек a и b и всех корней уравнения (2.2), лежащих в интервале (a, b) . Тогда решение задачи оптимизации сведется к выбору такого значения x_* из X , для которого $f(x_*)$ будет наименьшим (если решение единственно) из всех других значений $f(x)$ на множестве X .

Рассмотрим пример. Положим, что требуется найти на отрезке $[-2, 2]$ минимум функции

$$f(x) = 3x^4 + 16x^3 + 6x^2 - 72x + 47. \quad (2.3)$$

Приравнявая $f'(x)$ нулю, получим для определения критических точек следующее уравнение:

$$x^3 + 4x^2 + x - 6 = 0. \quad (2.4)$$

Решением уравнения (2.4) будут три корня: $x_1 = -3, x_2 = -2, x_3 = 1$.

Отбрасывая первый корень, лежащий вне отрезка $[-2, 2]$, и учитывая границы отрезка, будем иметь три подозреваемые точки: $-2, 1, 2$. Вычисляя значения $f(x)$ в указанных точках, получим

$$f(-2) = 135, f(1) = 0, f(2) = 103.$$

Следовательно, искомый глобальный минимум функции (2.3) находится в точке $x = 1$ и равен 0.

Заметим, что классический метод можно применять и к функциям, имеющим на отрезке конечное число «переломов» – конечных разрывов производной $f'(x)$. При этом в число подозреваемых следует включать и точки, где $f(x)$ имеет такие переломы.

В то же время в некоторых задачах оптимизации аналитическое представление целевой функции может быть неизвестно. Например, $f(x)$ может быть задана графически, таблично или определять свои значения результатами эксперимента. В таких случаях обычно прибегают к численным методам решения задачи.

Пусть x_{\min} – точка глобального минимума некоторой функции $f(x)$ – точное значение, x^* – приближенное значение той же величины. Тогда будем считать, что решение получено с точностью $\varepsilon > 0$, если

$$|x^* - x_{\min}| \leq \varepsilon. \quad (2.5)$$

Следует заметить, что оценка (2.5) не характеризует отклонения $|f(x^*) - f(x_{\min})|$. Однако, так как основная цель оптимизации состоит в определении значения проектного параметра, т.е. переменной x , то оценка точности в виде (2.5) и будет в указанном смысле предпочтительней.

Ниже рассмотрим некоторые численные методы одномерной оптимизации.

2.1. Метод последовательного перебора

Разобьем отрезок $[a, b]$ на n равных интервалов длиной $h=(b - a)/n$ и положим, что в каждой точке деления $x_k=a+kh$ ($k=0, 1, \dots, n$) может быть однозначно найдено значение целевой функции $f(x_k)$, заданной в каком-либо виде.

Метод последовательного перебора состоит в последовательном определении $f(x_0), f(x_1), \dots, f(x_k), \dots$ и запоминании такой пары величин $x_k, f(x_k)$, для которой $f(x_k)$ меньше ранее вычисленных значений целевой функции. Тогда, в результате перебора всех точек деления отрезка, будет определена пара чисел x^*, y^* , соответствующая некоторой точке $x_i \equiv x^*$ ($i=0, 1, \dots, n$), где $f(x_i) \equiv y^* \leq f(x_k)$ для $k = 0, 1, \dots, n$.

Величины x^*, y^* и принимаются за приближенное решение поставленной задачи. Алгоритм последовательного перебора указан на рис. 2.2.

Заметим, что указанный алгоритм обладает следующей особенностью. Если среди вычисленных значений функции окажется, что

$f(x_i) = f(x_j) = \dots = f(x_l) < f(x_k)$, при $k=0, 1, \dots, n; k \neq i, k \neq j, \dots, k \neq l$, то в результате принятого варианта перебора будет получен один результат, а именно первый из вычисленных x_i, x_j, x_l , т.е. при $x_i < x_j < \dots < x_l$ решением будет x_i . В то же время, если в блоке 5 ввести нестрогое неравенство, то решением станет x_l .

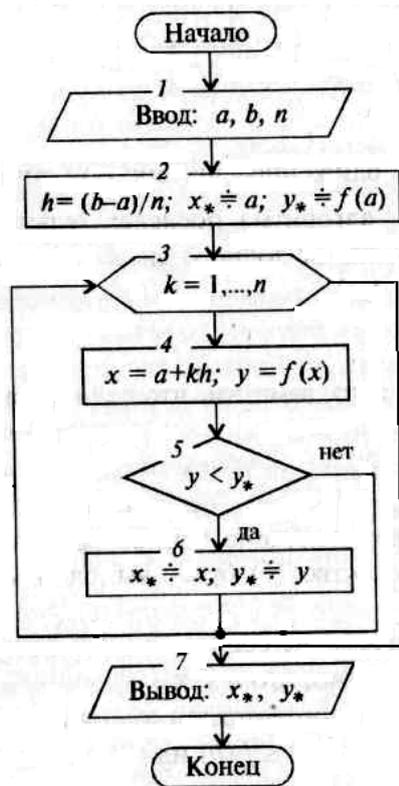


Рис. 2.2. Алгоритм метода последовательного перебора

Точность решения задачи методом последовательного перебора зависит от величины шага h и особенностей целевой функции на отрезке $[a, b]$.

Положим, что непрерывная функция $f(x)$ на $[a, b]$ унимодальна:

- имеет один минимум в точке x_{\min} ($a \leq x_{\min} \leq b$);
- при $a < x_{\min} < b$ на интервале (a, x_{\min}) непрерывно убывает ($f(x_1) > f(x_2) > f(x_{\min})$ при $x_1 < x_2 < x_{\min}$), а на интервале (x_{\min}, b) непрерывно возрастает ($f(x_{\min}) < f(x_1) < f(x_2)$ при $x_{\min} < x_1 < x_2$) (рис. 2.23 а);
- при $x_{\min} = a$ непрерывно возрастает (рис. 2.3 б);
- при $x_{\min} = b$ непрерывно убывает (рис. 2.3 в).

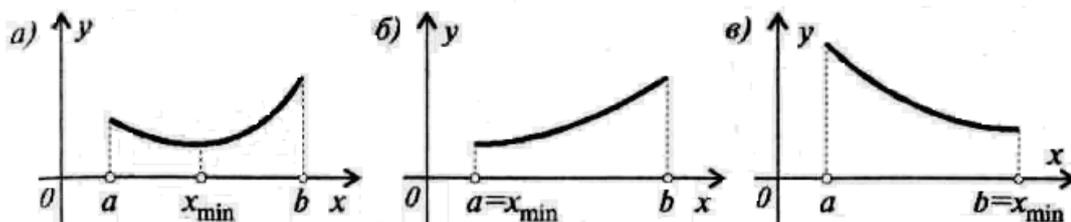


Рис. 2.3

Тогда можно утверждать, что приближенное значение x_* точки минимума, полученное с помощью алгоритма последовательного перебора (см. рис. 2.3), будет отстоять от точного значения x_{\min} менее, чем на шаг $h = (b - a)/n$, $n = 1, 2, \dots$:

$$|x_* - x_{\min}| < h. \quad (2.6)$$

Докажем это. Для чего, прежде всего, заметим, что если какая-либо из точек x_0, x_1, \dots, x_n случайно совпадет с x_{\min} (в частности, это будет иметь место при $x_0 = a = x_{\min}$ для возрастающей функции $f(x)$ (рис. 2.3 б), и при $x_n = b = x_{\min}$ для убывающей функции $f(x)$ (рис. 2.3 в)), то будет получено точное решение. Следовательно, в этом случае неравенство (2.6) будет выполняться при всяком $h > 0$.

Рассмотрим другой случай, когда x_{\min} не совпадает ни с одной из точек x_0, x_1, \dots, x_n . При этом, очевидно, что $a < x_{\min} < b$.

Пусть $n = 1$. Тогда $h = b - a$ и в соответствии с принятым алгоритмом будет получено $x_* = a$ при $f(a) \leq f(b)$ или $x_* = b$ при $f(a) > f(b)$. При этом x_* всегда будет отстоять от x_{\min} ближе, чем на $h = b - a$. Положим теперь, что $n \geq 2$. Тогда

в соответствии с алгоритмом будет найдено такое $x^* = x_i$, что вычисленные значения $f(x)$ будут соотноситься между собой следующим образом:

$$f(x_0) > f(x_1) > \dots > f(x_{i-1}) > f(x_i) \leq f(x_{i+1}) < f(x_{i+2}) < \dots < f(x_n) \quad (2.7)$$

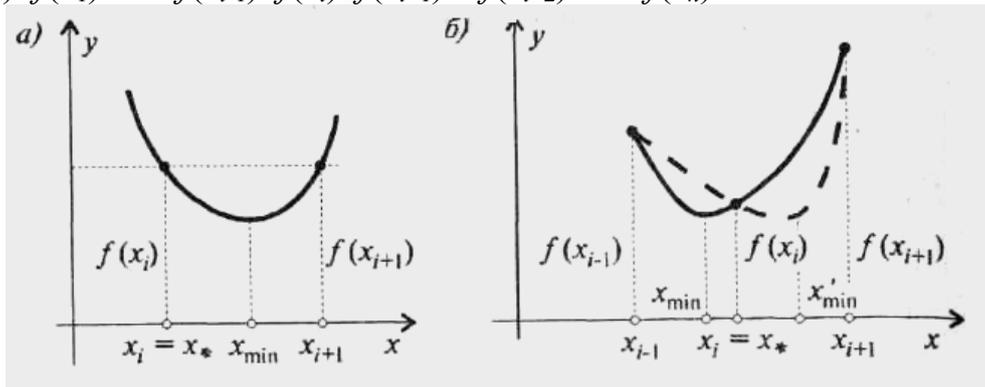


Рис. 2.4

т.е. допустимо одно нестрогое неравенство.

При этом заметим, что для унимодальной функции из равенства $f(x_i) = f(x_{i+1})$, будет следовать, что $x_i < x_{\min} < x_{i+1}$ (рис. 2.4 а), т.е. неравенство (2.6) будет выполняться.

При $f(x_i) < f(x_{i+1})$ (рис. 2.4 б) точка x_{\min} может находиться или на (x_{i-1}, x_i) , или на (x_i, x_{i+1}) . Очевидно, что в обоих случаях x_{\min} будет отстоять от $x^* = x_i$ ближе, чем на величину h .

Таким образом, мы доказали, что при оговоренных выше условиях неравенство (2.6) будет выполняться всегда.

Заметим, что оценка (2.6) будет справедлива и не только тогда, когда унимодальность имеет место на всем отрезке $[a, b]$, т.е. когда $f(x)$ глобально унимодальна на $[a, b]$.

Положим, что заданная на $[a, b]$ непрерывная функция $f(x)$ имеет

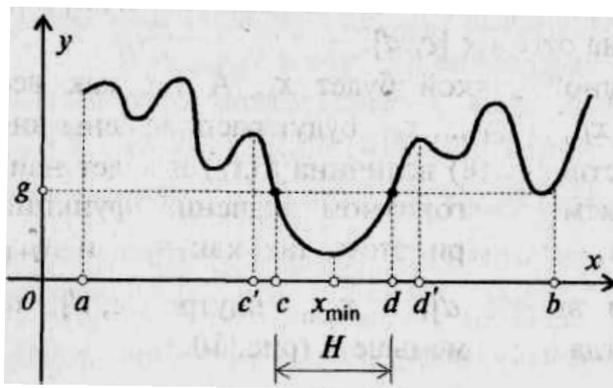


Рис. 2.5

содержащийся в $[a, b]$ участок $[c', d']$ унимодальности, т.е. является локально унимодальной (рис. 2.5). Пусть далее g - минимальное значение $f(x)$ на $[a, b]$ вне отрезка $[c', d']$. Тогда, если из $[c', d']$ можно выделить такой отрезок $[c, d]$ при $d - c \equiv H > 0$, что $f(x) \leq g$ на $[c, d]$ то будем говорить, что на $[a, b]$ $f(x)$ обладает локальной унимодальностью с параметром H . При этом, очевидно, что

$$f(x) \geq g \text{ на } [a, b] \text{ вне } [c, d] \quad (2.8)$$

и x_{\min} на $[c, d]$ будет глобальным минимумом на $[a, b]$.

Заметим, что т.к. $f(x)$ непрерывна и унимодальна на $[c, d]$, то при этом будут возможны следующие случаи:

- $g = f(c) = f(d)$ при $c > a$ и $d < b$ (рис. 2.4),
- $g = f(d) > f(c)$ при $c = a$ и $d < b$ (рис. 2.5 а),
- $g = f(c) > f(d)$ при $c > a$ и $d = b$ (рис. 2.5 б).

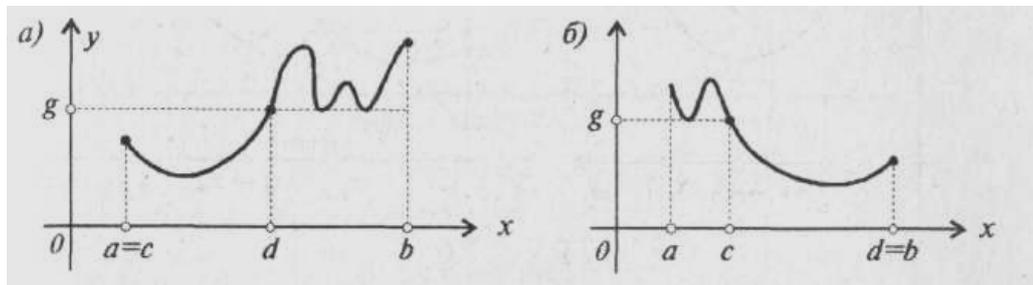


Рис. 2.6

Для введенного нами класса функций будет справедливо следующее утверждение: если на отрезке $[a, b]$ функция $f(x)$ обладает локальной унимодальностью с параметром H , то результат x_* , полученный методом последовательного перебора с шагом $h < H$, будет удовлетворять неравенству (2.6).

Для доказательства, прежде всего, рассмотрим случай, когда $c > a$ и $d < b$ (рис. 2.5). При этом заметим, что при переборе с шагом $h < H$ хотя бы одна точка из последовательности x_0, x_1, \dots, x_n ($h = (b - a)/n$) попадет на отрезок $[c, d]$.

Пусть такой единственной точкой будет x_i . А так как все остальные точки $x_0, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ будут расположены вне отрезка $[c, d]$, то в соответствии с (2.8) величина $f(x_i)$ и будет наименьшей из всех вычисляемых алгоритмом значений функции $f(x)$, т.е. будет принято $x_* = x_i$. При этом, так как x_{i-1} и x_{i+1} окажутся расположенными вне $[c, d]$, а x_{\min} – внутри $[c, d]$, то расстояние $|x_* - x_{\min}|$ всегда будет меньше h (рис. 2.7).

Положим теперь, что на (c, d) попали следующие m точек: $x_{k+1}, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_{k+m}$ ($m = 2, 3, \dots$), причем x_i принята за результат x_* . Добавляя к указанной последовательности x_k и x_{k+m+1} , заметим, что вследствие локальной унимодальности на (c, d) и условия (2.8) будет:

$$f(x_k) > f(x_{k+1}) > \dots > f(x_{i-1}) > f(x_i) \leq f(x_{i+1}) < \dots < f(x_{k+m}) < f(x_{k+m+1}).$$

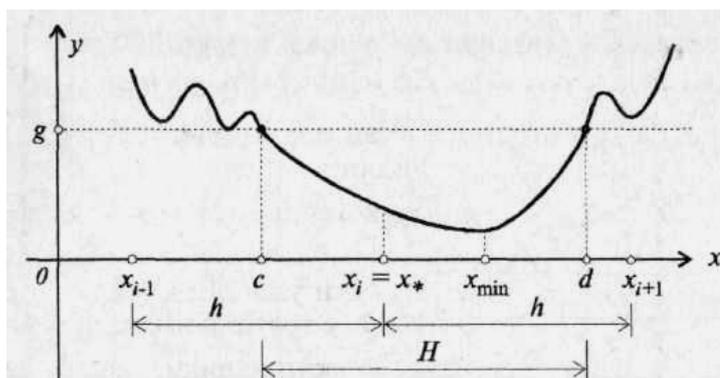


Рис. 2.7

При этом мы получили тот же вид соотношения последовательности вычисленных значений функции $f(x)$, что и при глобальной унимодальности (2.7). Следовательно, применяя те же рассуждения, приходим к выводу, что при любом числе точек, попавших на интервал (c, d) , неравенство (2.6) будет выполняться.

То же совершенно аналогично можно доказать и для $f(x)$. Функция обладает на $[a, b]$ глобальной или локальной унимодальностью, то использование метода последовательного перебора с шагом $h \leq H$ обеспечит результат с погрешностью $\varepsilon < h$. Однако, к сожалению, необходимой

информацией об указанных свойствах целевой функции инженер, как правило, не располагает. В связи с этим можно рекомендовать воспользоваться следующим *необходимым признаком* результатов, полученных с определенной точностью.

Так, если результаты x'_* и x_* получены с погрешностями, соответственно, 2ε и ε , т.е. $|x'_* - x_{\min}| \leq 2\varepsilon$, $|x_* - x_{\min}| \leq \varepsilon$, то, очевидно, должно быть

$$|x'_* - x_*| \leq 3\varepsilon. \quad (2.9)$$

Поэтому, если в результате «грубого» (с шагом $h = 2\varepsilon$) и «точного» (с шагом $h = \varepsilon$) просчетов результаты x'_* и x_* действительно окажутся вычисленными с погрешностями 2ε и ε , соответственно, то они должны удовлетворять неравенству (2.9).

Указанным обстоятельством можно воспользоваться при составлении алгоритма решения задачи одномерной оптимизации методом перебора с проверкой (рис. 2.8). Здесь в качестве исходной информации предусмотрен ввод

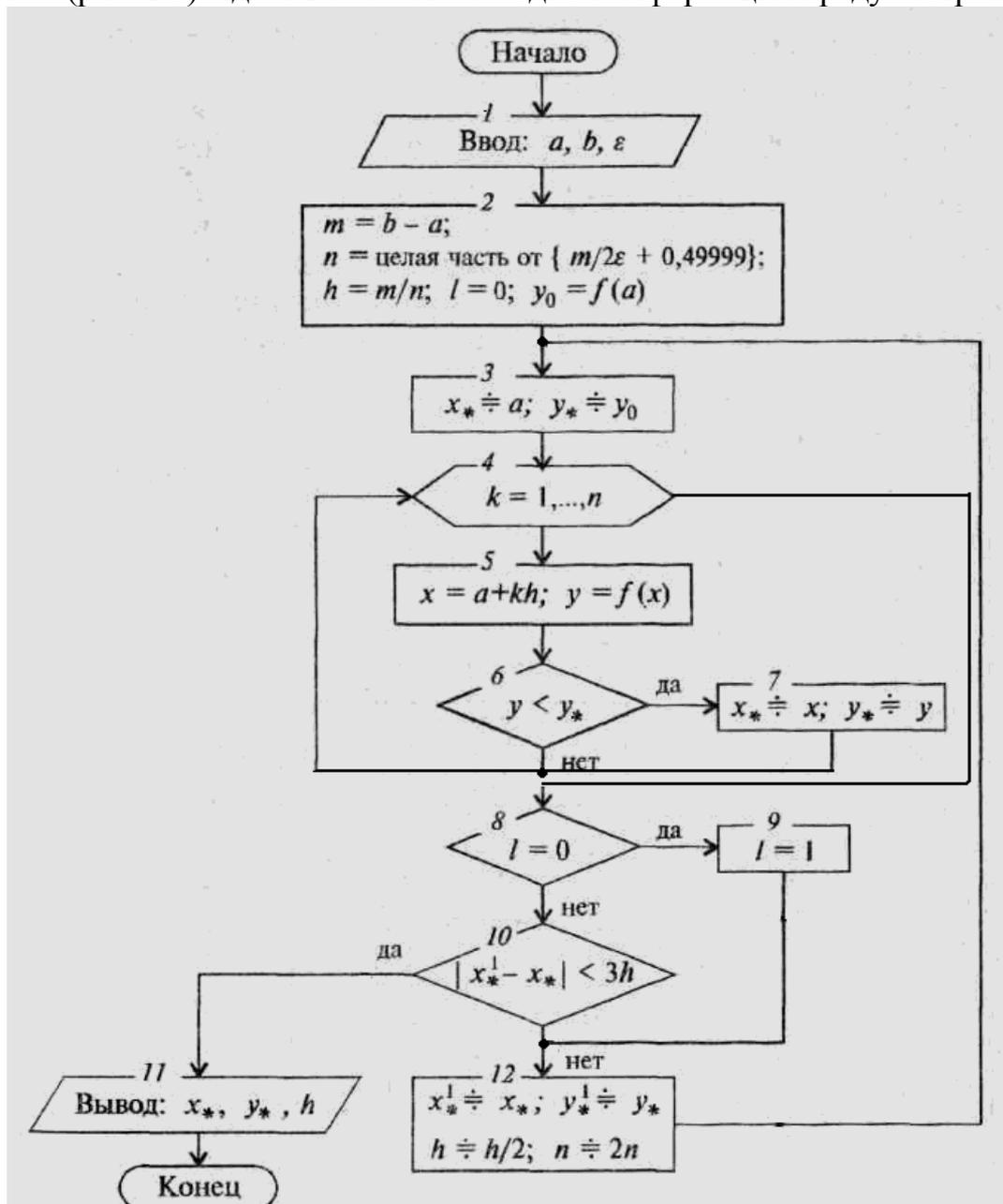


Рис. 2.8

величин a, b – начала и конца отрезка $[a, b]$ и ε – допуская на погрешность вычислений (блок 1). В блоке 2 предусматривается вычисление числа n шагов «грубого» прохода отрезка. Принимается точно $n = (b - a)/2\varepsilon$, если результат – целое число. В противном случае полученный от деления результат округляется до большего целого. По вычисленному значению n определяется длина первичного шага h (при этом будет $h < 2\varepsilon$) и вычисляется $f(x)$ при $x=a$. Заметим, что признаком вычислений с первичным шагом будет нулевое значение некоторого параметра $l(l=0)$. Далее блоки 3–7 выполняют поиск x^*, y^* методом последовательного перебора (рис. 2.2). В дальнейшем осуществляется переобозначение переменных ($x_*^1 := x^*, y_*^1 := y^*$) и уменьшение шага вдвое (блок 12). Вторичный ($l=1$) перебор (блоки 3–7) позволяет получить новые значения x^*, y^* при возросшем в два раза числе шагов, после чего предусматривается (блок 10) проверка необходимого условия (2.9). При положительном результате проверки выдаются на печать x^*, y^* и шаг h , характеризующий точность результата (блок 11).

Если условие (2.9) не выполняется, то производится дальнейшее уменьшение вдвое шага h и вычисление новых x^*, y^* с последующей проверкой и т.д.

Ниже приведем пример, реализующий указанный алгоритм при

$$f(x) = 5,2 + e^x \cdot \sin x^5, \quad (2.10)$$

график которой изображен на рис. 2.9.

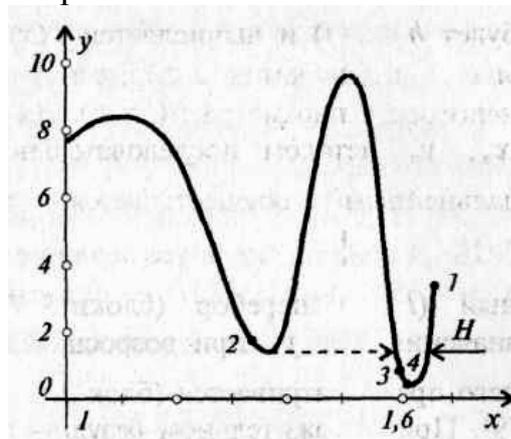


Рис. 2.9

Заметим, что функция (10) на отрезке $[1,05; 1,65]$ имеет глобальный минимум $y_1=0,1686$ в точке $x_1=1,6161$ и локальный минимум $y_2=1,284$ в точке $x_2=1,367$ – все числа указаны с округлением (рис. 2.9). При этом H локальной унимодальности составляет $0,045$.

Очевидно, что при задании $\varepsilon > 0,045$ мы рискуем получить результат x_2, y_2 . Действительно, при $a = 1,05, b = 1,65, \varepsilon = 0,075$ получим результаты

$XZ=1.35 \ YZ=1.442715$ Погрешность 0.075 , соответствующую локальному минимуму (x_2, y_2) .

В то же время при $\varepsilon = 0,001$ будем иметь

$XZ=1.616 \ YZ=0.1686624$ Погрешность $=0.001$.

Мы получили достаточно точный результат. В этом можно убедиться увеличивая на порядок точность ($\varepsilon = 0,0001$):

$XZ=1.616111 \ YZ=0.1686174$ Погрешность $=9.996668E-05$.

Заметим, что эти результаты мы могли бы получить и с помощью программы, составленной по алгоритму без использования проверки (2.9) (рис. 2.2). В то же время приведенный выше алгоритм дает нам больше шансов найти достаточно точное решение даже тогда, когда $\varepsilon > H$. Так, при $\varepsilon=0,1 > H=0,045$ будем иметь $XZ=1,625$ $YZ=0,4045381$. Погрешность 0.025.

Процесс получения данного решения отображается в следующей таблице.

Таблица 2.1

№ просчета	n	h	x^*	y^*
1	3	0,2	1,650	3,4804270
2	6	0,1	1,350	1,4427150
3	12	0,05	1,600	0,8768272
4	24	0,025	1,625	0,4045382

При заданном $\varepsilon = E = 0,1$ вычислено

$$n = (b - a)/2E = (1,65 - 1,05)/0,2 = 3,$$

и в результате «грубого» перебора с шагом $h = 2E = 0,2$ была найдена пара значений $x^* = 1,65$, $y^* = 3,480427$ (точка 1 на рис. 2.9) вблизи глобального минимума (табл. 2.1).

В результате просчета с шагом $h = E = 0,1$ было получено $x^* = 1,35$, $y^* = 1,442715$ – окрестность локального минимума (точка 2). При этом проверка условия (2.9) дала отрицательный результат:

$$|1,35 - 1,65| \equiv 0,3 = 3 \cdot 0,1 = 0,3.$$

Поэтому был произведен третий просчет с шагом $h = 0,05$ (точка 3). Однако и при этом условие (2.9) не выполнилось:

$$|1,60 - 1,35| \equiv 0,25 > 3 \cdot 0,05 \equiv 0,15.$$

И только результат четвертого просчета ($h = 0,025 < H = 0,045$) привел к удовлетворению необходимого условия:

$$|1,625 - 1,60| \equiv 0,025 < 3 \cdot 0,025 \equiv 0,075.$$

2.2. Последовательный перебор с переменным шагом

Недостатком метода последовательного перебора является необходимость поточечного вычисления $f(x)$ на всем интервале неопределенности с одинаковой подробностью (с шагом $h = \varepsilon$). При этом для некоторых задач трудоемкость вычислений может оказаться столь значительной, что получение результата в приемлемое для пользователя время не сможет обеспечить и самая быстродействующая ЭВМ. В этом случае может помочь следующий алгоритм вычислений.

Пусть для унимодальной на отрезке $[a, b]$ функции $f(x)$ требуется найти x_{\min} с погрешностью ε .

Зададимся некоторым шагом $h_0 > \varepsilon$ и методом последовательного перебора найдем такое x_i , что $f(x_i) < f(x_k)$. Для всех $k = 0, 1, 2, \dots$ кроме $k = i$ и может быть $k = i + 1$, т.к. в соответствии с алгоритмом может оказаться $f(x_i) = f(x_{i+1})$. Тем самым будет найден содержащий x_{\min} интервал неопределенности $[x_{i-1}, x_{i+1}]$ длиной $l_0 = x_{i+1} - x_{i-1} = 2h_0$. На это потребуется затратить

$$N_0 = 1 + \frac{b-a}{h_0}$$

вычислений функции $f(x)$.

Далее сделаем 1-е уточнение с помощью того же последовательного перебора, проходя отрезок $[x_{i-1}, x_{i+1}]$ с шагом

$$h_1 = \frac{h_0}{\lambda},$$

где λ – некоторое положительное число – коэффициент дробления, обеспечивающий деление интервала $l_0 = 2H_0$ на целое число частей $j = 2\lambda$. Так при $j = 3, 4, 5, \dots$ соответственно, будет $\lambda = 1, 5; 2; 2, 5; \dots$

Чтобы не усложнять алгоритм, будем считать, что из всех трех вычисленных на $[x_{i-1}, x_{i+1}]$ при первичном проходе значений функции $f(x)$ ($f(x_{i-1})$, $f(x_i)$, $f(x_{i+1})$), при уточнении используется только $f(x_{i-1})$, а все последующие значения $f(x)$, определяемые с шагом h_1 , вычисляются заново. Тогда на такое уточнение затратится

$$N_1 = \frac{l_0}{h_1} = 2\lambda$$

вычислений целевой функции $f(x)$. При этом новый интервал неопределенности будет иметь длину $l_1 = 2h_1$.

Аналогично для k -го уточнения ($k = 1, 2, 3, \dots$), будет

$$h_k = \frac{h_{k-1}}{\lambda} = \frac{h_0}{\lambda^k},$$

$$l_k = 2h_k,$$

$$N_k = \frac{l_{k-1}}{h_k} = 2\lambda.$$

Положим, что в результате m уточнений шаг

$$h_m = \frac{h_0}{\lambda^m} \tag{2.11}$$

окажется равным величине допустимой погрешности

$$h_m = \varepsilon. \tag{2.12}$$

Тогда вычисления можно прекратить, т.к. в соответствии с неравенством (2.6) погрешность результата окажется меньше ε .

Из соотношений (2.11), (2.12) можно найти требуемый для получения решения за m уточнений коэффициент λ дробления шага:

$$\lambda = \alpha^{\frac{1}{m}}, \tag{2.13}$$

где обозначено $\alpha = \frac{h_0}{\varepsilon}$.

При этом общее число вычислений $f(x)$ составит

$$N = N_0 + \sum_{k=1}^m N_k = 1 + \frac{b-a}{h_0} + 2m\lambda = 1 + \frac{b-a}{h_0} + m2\alpha^{\frac{1}{m}}. \tag{2.14}$$

Если условиться считать, что m – вещественная переменная, то можно найти стационарную точку функции N из следующего уравнения:

$$\frac{dN}{dm} \equiv 2\alpha^{\frac{1}{m}} - 2\frac{1}{m}\alpha^{\frac{1}{m}} \ln \alpha \equiv 2\alpha^{\frac{1}{m}} \left(1 - \frac{1}{m} \ln \alpha\right) = 0.$$

Отсюда (при $\alpha > 0$) следует

$$m = \ln \alpha \quad (2.15)$$

Подставляя полученное выражение в соотношение (13), будем иметь

$$\lambda = \alpha^{\frac{1}{\ln \alpha}} \equiv e, \quad (2.16)$$

где $e=2,71828$ – основание натуральных логарифмов.

В связи с тем, что вторая производная функции $N(14)$ в точке $m = \ln \alpha$ положительна

$$\left. \frac{d^2 N}{d^2 m} \right|_{m=\ln \alpha} = \frac{2e}{\ln \alpha} > 0,$$

(т.к. при $h_0 > \varepsilon$ будет $\alpha > 1$ и $\ln \alpha > 0$), то можно утверждать, что полученные результаты (2.15), (2.16) обеспечивают минимум N .

Примечательно, что найденный нами оптимальный коэффициент дробления (2.16) не зависит от h_0 и ε . В то же время воспользоваться иррациональным коэффициентом для реализации принятого алгоритма мы не можем и вынуждены выбрать в качестве λ какое-либо близкое к e число, обеспечивающее деление интервала на целое число частей, а количество уточнений определять из условия выполнения неравенства

$$h_m \leq \varepsilon. \quad (2.17)$$

В случае, если λ задано, то при определении требуемого числа уточнений m_* – целого числа следует воспользоваться величиной m , которую можно найти из соотношения (2.11) и условия (2.12):

$$m = \frac{\ln \alpha}{\ln \lambda}. \quad (2.18)$$

Положим, что вычисленная по формуле (2.18) величина m – целое число. Тогда примем $m_* = m$. При m – дробном числе, следует взять в качестве m_* наименьшее из целых чисел, больших m .

Очевидно, что при дробном m

$$m_* < m + 1.$$

Поэтому, учитывая (2.18), можно дать следующую оценку m_* :

$$\frac{\ln \alpha}{\ln \lambda} \leq m_* < \frac{\ln \alpha}{\ln \lambda} + 1. \quad (2.19)$$

А так как общее число N_* вычислений $f(x)$, необходимых для m_* уточнений, составляет

$$N_* = \sum_{k=1}^m N_k = 2m_* \lambda,$$

то с помощью двойного неравенства (2.19) можно дать двустороннюю оценку N_* :

$$2 \frac{\lambda}{\ln \lambda} \ln \alpha \leq N_* < 2\lambda \left(\frac{1}{\ln \lambda} \ln \alpha + 1 \right).$$

В частности, при $\lambda=3$

$$5,45 \ln \alpha \leq N_* < 5,46 \ln \alpha + 6, \quad (2.20)$$

а при

$$h_0=1, \varepsilon = 0,005 \quad (2.21)$$

будем иметь $28,94 < N_* < 34,94$.

Для иллюстрации выигрыша, получаемого при использовании λ , близких к оптимальному, рассмотрим таблицу 2.2, где для данных (2.21) и $\lambda=1,5; 2; 2,5; 3; \dots; 10$ приведены соответствующие значения j -числа частей разбиения интервала неопределенности, m_* – требующегося количества уточнений, N_* – числа вычислений функции $f(x)$, h_m – длины шага при последнем уточнении, характеризующей достигнутой точности определения x_{\min} ($h_m < \varepsilon = 0,005$).

Таблица 2.2

λ	j	m_*	N_*	N_{**}	h_m
1,5	3	14	42	28	0,0034
2	4	8	32	16	0,0039
2,5	5	6	30	24	0,0041
3	6	5	30	20	0,0041
4	8	4	32	24	0,0039
5	10	4	40	32	0,0016
10	20	3	60	54	0,0010

Как видно из таблицы 2.2, для получения результата при $\lambda = 2,5$ (деление интервала неопределенности на 5 частей) и при $\lambda = 3$ (деление на 6 частей) потребовалось в два раза меньше вычислений $f(x)$, чем при $\lambda = 10$.

Заметим, что алгоритм перебора можно построить таким образом, чтобы при каждом k -м проходе интервала неопределенности использовать в совпадающих точках все значения $f(x)$, вычисленные при $k-1$ -й итерации. Тогда отпадет необходимость при каждом проходе повторно вычислять $f(x)$, конечной точке интервала, а при четных j и значение $f(x)$ в середине отрезка. Поэтому количество N_{**} вычислений целевой функции, требующееся для m_* уточнений при таком алгоритме составит

$$N_{**} = N_* - m_* \text{ для нечетных } j,$$

$$N_{**} = N_* - 2m_* \text{ для четных } j.$$

Как видно из таблицы 2.2, наименьшее значение N_{**} при данных (2.21) имеет место при $\lambda = 2$.

Приведем вариант алгоритма перебора с переменным шагом без повторных вычислений $f(x)$ при $\lambda = 2$.

Пусть для унимодальной функции $f(x)$ в результате последовательного перебора с шагом h_0 получен содержащий x_{\min} интервал неопределенности $[x_{i-1}, x_{i+1}]$ при $f(x_{i-1}) > f(x_i) \leq f(x_{i+1})$. Для упрощения дальнейших рассуждений введем следующие обозначения: $a = x_{i-1}$, $b = x_{i+1}$, $x^* = x_i$, $y^* = f(x_i)$. Тогда последующие действия можно осуществить в соответствии с рис. 2.10 и блок-схемой на рис. 2.11:

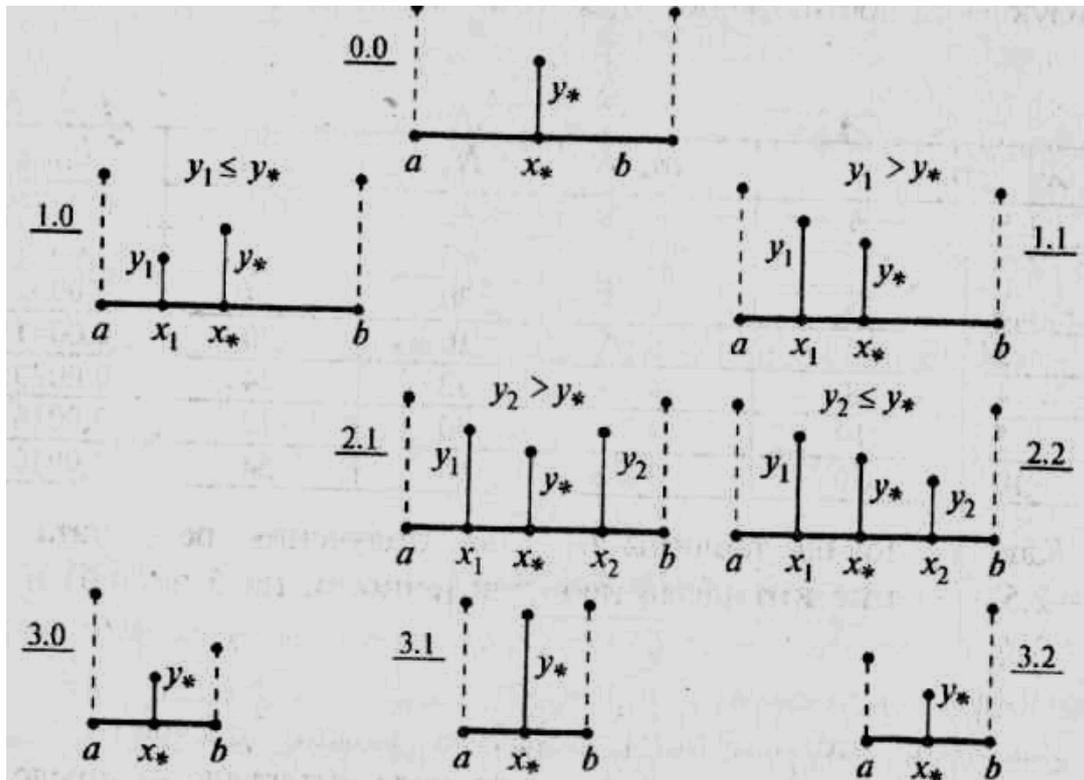


Рис. 2.10

1) вычисляем $y_1=f(x_1)$, $x_1=(a+x_*)/2$ и сравниваем y_1 с y_* (блоки 2, 3 на рис. 2.11). При этом из $y_1 \leq y_*$ из будет следовать, что $f(x)$ имеет x_{\min} на (a, x_*) –позиция 1.0 на рис. 2.10. Поэтому, в результате переобозначений (блок 4), получим уменьшенный интервал неопределенности - позиция 3.0;

2) при $y_1 > y_*$ (позиция 1.1) вычислим $y_2=f(x_2)$ при $x_2=(x_*+b)/2$ (блок 5). При этом из $y_2 \leq y_*$ (позиция 2.2) будет следовать, что x_{\min} принадлежит (x_*, b) . В противном случае (позиция 2.1) – интервалу (x_1, x_2) ;

3) во всех трех вариантах придем к уменьшенному в два раза интервалу неопределенности с известным в центральной точке значением $f(x)$ (позиции 3.0, 3.1, 3.2). При $(b - a)/2 \leq \varepsilon$ точка x_* будет отстоять от x_{\min} не далее, чем на ε – решение получено. В противном случае необходимо произвести следующую итерацию.

Дадим оценку трудоемкости алгоритма по числу вычислений $f(x)$. Для этого первоначально положим, что $x_{\min}, - a \leq \varepsilon$ (рис. 2.12)

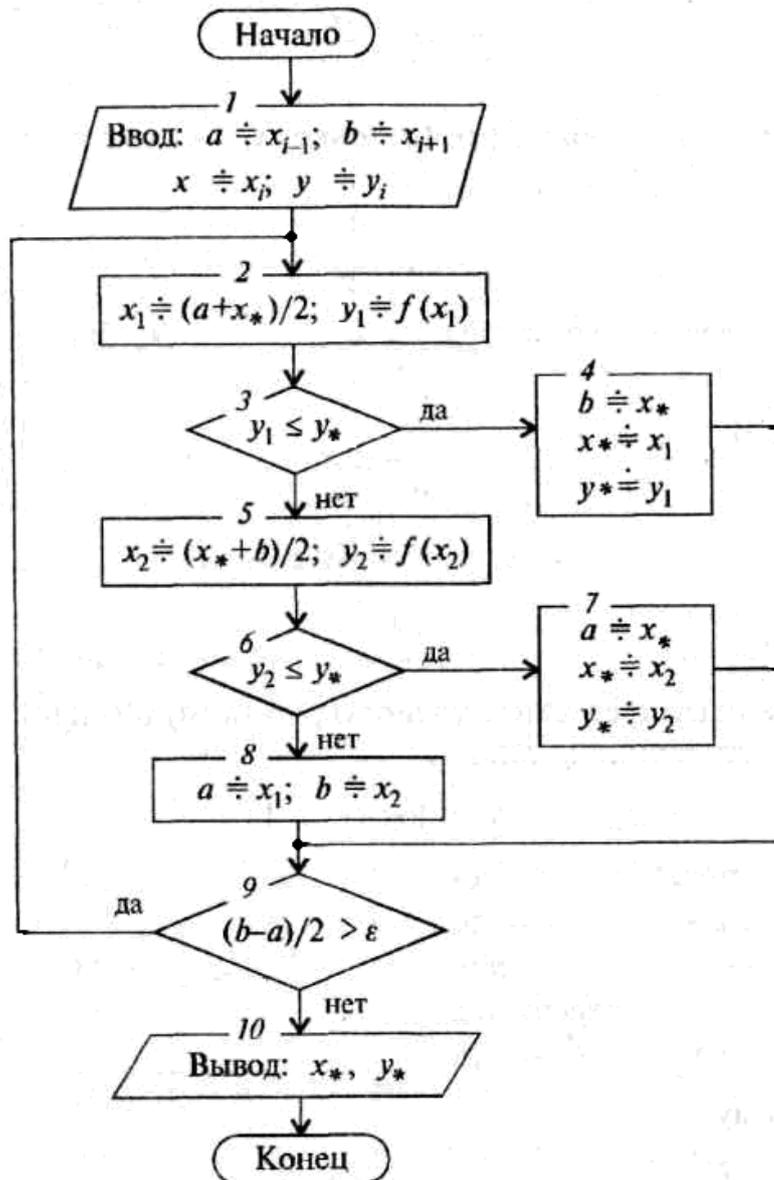


Рис. 2.11

Тогда в результате первого вычисления $f(x)$ в точке $x_1(a+x_*)/2$ будет $y_1=f(x_1)<y_*$, следовательно, будет установлено, что x_{\min} лежит внутри $[a, x_*]$ и если принять x_1 за приближение значение x_{\min} , то погрешность результата не

будет превосходить величины $\Delta_1 = \frac{b-a}{2^2} = \frac{2h_0}{2^2} = \frac{h_0}{2}$.

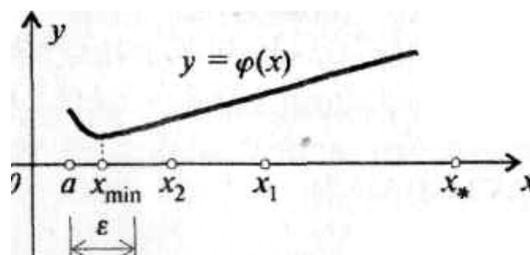


Рис. 2.12

Второе вычисление $f(x)$ (в точке x_2) сократит погрешность вдвое:

$$\Delta_2 = \frac{h_0}{2^2}.$$

И так далее. Результат будет достигнут за N_*^1 вычислений $f(x)$, если

$$\frac{h_0}{2^{N_*^1}} \leq \varepsilon.$$

Отсюда $N_*^1 \geq \log_2 \alpha$,

где $\alpha = h_0 / \varepsilon$.

Так как под N_*^1 будем понимать наименьшее целое число, обеспечивающее требуемую точность, то нетрудно написать и следующую двустороннюю оценку

$$\log_2 \alpha < N_*^1 < \log_2 \alpha + 1. \quad (2.22)$$

Можно показать, что если $b - x_{\min} < \varepsilon$ (рис. 2.13), то для достижения результата той же точности потребуется число $N_*^2 = 2N_*^1$ вычислений $f(x)$. Поэтому

$$2 \log_2 \alpha \leq N_*^2 < 2(\log_2 \alpha + 1). \quad (2.23)$$

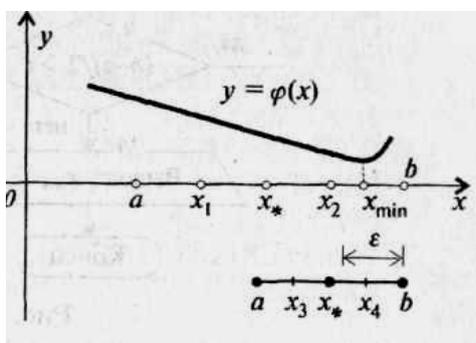


Рис. 2.13

А так как нами рассмотрены два крайних случая, то, используя (2.22), (2.23), получим оценку для числа N_* вычислений $f(x)$ при любом положении x_{\min} на $[a, b]$:

$\log_2 \alpha \leq N_* < 2(\log_2 \alpha + 1)$, или

$$1,44 \ln \alpha \leq N_* < 2,88 \ln \alpha + 2. \quad (2.24)$$

Можно доказать, что если $f(x)$ на отрезке $[a, b]$ обладает локальной унимодальностью с параметром $H > 0$, то алгоритм последовательного перебора с переменным шагом при $0 < h_0 < H$

$1 < \lambda < +\infty$ будет решать задачу определения приближенного значения x_* глобального минимума x_{\min} на отрезке $[a, b]$ с наперед заданной погрешностью $\varepsilon > 0$.

Ниже приведена БЕЙСИК-программа, реализующая метод перебора с переменным шагом для целевой функции $f(x) = 5,2 + e^x \cdot \sin x^5$ (2.10), имеющей на отрезке $[1,05; 1,65]$ локальную унимодальность с параметром $H = 0,045$.

Программа 27

```

PRINT "Одномерная оптимизация" PRINT "Повторный перебор,
лямбда=2" INPUT "A, B, H, E ";A,B,H,E X=A: XZ=A: GOSUB 5: YZ=Y
1 X=X+H: IF X>B THEN 2
GOSUB 5
IF Y<YZ THEN XZ=X: YZ=Y GOTO 1
2A=XZ-H: B=XZ+H

```

```

3 X1=(A+XZ)/2: X=X1: GOSUB 5: Y1=Y
  IF YK=YZ THEN B=XZ: XZ=X1: YZ=Y1: GOTO 4
  X2=(XZ+B)/2: X=X2: GOSUB 5: Y2=Y
  IF Y2<=YZ THEN A=XZ: XZ=X2: YZ=Y2: GOTO 4
  A=X1: B=X2
4 IF (B-A)/2>E THEN 3 PRINT "XZ="; XZ; "YZ="; YZ
  END
5 Y=5.2+EXP(X)*SIN(X^5)
  RETURN

```

Для работы программы следует задать начало (A), конец (B) интервала неопределенности, начальный шаг (H) и допустимую погрешность (E) результата.

Первоначально (строки 4–8) предусматривается проход всего отрезка [A, B] с шагом H и выделение интервала неопределенности длиной 2H. Затем, в соответствии с алгоритмом (рис. 2.11), осуществляется уточнение результата до требуемой точности.

Очевидно, что программа даст правильный результат, если выбрать $H < 0,045$.

При $A=1,05$, $B=1,65$, $H=0,044$, $E=0,001$ была получена следующая выдача:
 $XZ=1.615813$ $YZ=0.1689014$.

При этом было произведено 24 вычисления функции $f(x)$. Заметим, что для решения этой задачи методом перебора за один проход (с шагом 0,001) потребовалось бы 601 вычисление $f(x)$.

В заключение заметим, что для уточнения результатов можно воспользоваться и некоторыми классическими методами одномерной оптимизации.

3. Классические методы поиска безусловного экстремума

3.1. Общие положения

Применение необходимых и достаточных условий безусловного экстремума невозможно в тех случаях, когда целевая функция задана неявно, либо целевая функция не имеет непрерывных производных до второго порядка включительно. Кроме того, использование необходимого условия экстремума первого порядка

$$\nabla f(\mathbf{x}) = 0 \quad (3.1)$$

связано с решением системы \mathbf{n} в общем случае нелинейных алгебраических уравнений. Как правило, такую задачу приходится решать численно, причем трудоемкость решения сопоставима с трудоемкостью численного решения задачи оптимизации.

Численные методы оптимизации обеспечивают нахождение глобального экстремума только для выпуклых (вогнутых) функций, а в общем случае позволяют идентифицировать локальные экстремумы.

В данном разделе рассмотрены некоторые численные методы поиска безусловного локального минимума функции

$$f(\mathbf{x}^*) = \min_{\mathbf{x} \in R^n} f(\mathbf{x}) \quad (3.2)$$

Большинство классических численных методов оптимизации относится к классу итерационных. В соответствии с определенным правилом строится последовательность точек $\mathbf{x}^0, \mathbf{x}^2, \mathbf{x}^3, \dots$, обладающих свойством:

$$f(\mathbf{x}^{k+1}) < f(\mathbf{x}^k) \quad k = 0, 1, 2, \dots \quad (3.3)$$

Преобразование точки \mathbf{x}^k в точку \mathbf{x}^{k+1} представляет собой итерацию. Очередная точка последовательности $\{\mathbf{x}^k\}$ определяется по формуле:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + t_k \mathbf{d}^k \quad k = 0, 1, \dots, \quad (3.4)$$

где \mathbf{d}^k – вектор, определяющий направление перехода из точки \mathbf{x}^k в точку \mathbf{x}^{k+1} .
 t^k – величина шага.

Равенство (3.4) является общим, каждый конкретный алгоритм отличается правилами выбора векторов \mathbf{d}^k и чисел t^k . Кроме того, для реализации итерационного процесса необходимо задать точку \mathbf{x}^0 (начальное приближение) и условие окончания алгоритма.

Начальная точка поиска \mathbf{x}^0 задается, исходя из физического содержания решаемой задачи, в качестве критерия окончания счета, как правило, выбирается одно из условий:

$$|\mathbf{x}^{k+1} - \mathbf{x}^k| \leq \varepsilon_1, \quad (3.5)$$

$$f(\mathbf{x}^{k+1}) - f(\mathbf{x}^k) \leq \varepsilon_2, \quad (3.6)$$

$$\|f'(\mathbf{x}^{k+1})\| \leq \varepsilon_3. \quad (3.7)$$

где ε_i , $i=1,3$ – малое положительное число.

Иногда используются критерии, состоящие в одновременном выполнении двух из условий (3.5–3.7) или всех трех условий.

Важной характеристикой численных методов оптимизации является сходимость. Метод ($\mathbf{x}^{k+1} = \mathbf{x}^k + t_k \mathbf{d}^k$) сходится, если $\mathbf{x}^k \rightarrow \mathbf{x}^*$ при $k \rightarrow \infty$,

где \mathbf{x}^* – решение задачи (3.2).

Сходимость последовательности $\{\mathbf{x}^k\}$ зависит от характера функции $f(\mathbf{x})$ и от выбора начальной точки \mathbf{x}^0 .

В зависимости от наивысшего порядка частных производных функции $f(\mathbf{x})$, используемых в алгоритмах решения задачи оптимизации, численные методы можно разделить на три группы.

1. Методы нулевого порядка или методы прямого поиска, использующие только информацию о значениях функции $f(\mathbf{x})$.

2. Методы первого порядка или градиентные методы, использующие информацию о первых производных функции $f(\mathbf{x})$.

3. Методы второго порядка, в которых требуется знание вторых производных функции $f(\mathbf{x})$.

Для каждой конкретной задачи вопрос о том, каким методом оптимизации воспользоваться, решается в зависимости от свойств минимизируемой функции (например, для минимизации недифференцируемой функции нельзя воспользоваться алгоритмом, предусматривающим возможность вычисления в произвольной точке градиента функции).

В специальной литературе [1], [2], [3], представлены достаточно полные обзоры наиболее эффективных методов численно решения задач безусловной оптимизации.

Ниже рассмотрены лишь некоторые алгоритмы одномерной минимизации.

3.2. Численные методы одномерной минимизации

Задача одномерной минимизации может быть решена с помощью необходимых и достаточных условий безусловного экстремума, однако уравнение $f'(x)=0$ не всегда решается простым способом, поэтому приходится прибегать к численным методам. Ниже рассмотрены несколько простых численных процедур, непосредственного локализирующих минимум функции $f(x)$.

Процесс поиска включает следующие этапы:

1. Выбор границ начального интервала неопределенности, относительно которого точно известно, что этот интервал содержит точку оптимума и внутри этого интервала функция унимодальна, т.е. имеет один минимум в точке x^* .
2. Уменьшение интервала неопределенности таким образом, чтобы не исключить из него точку оптимума.

Эта процедура не дает возможности определить точный оптимум, однако позволяет уменьшить длину интервала, содержащего искомую точку. Длина такого интервала может быть сделана сколь угодно малой.

3. Проверка условия окончания. Поиск заканчивается, когда длина текущего интервала неопределенности оказывается меньше установленной величины.

3.2.1. Метод дихотомии

Пусть требуется найти локальный минимум функции $f(x)$, унимодальной на интервале неопределенности $[a, b]$.

Определим две симметрично расположенные относительно середины отрезка $[a, b]$ точки x_1 и x_2 :

$$x_1 = \frac{a+b}{2} + \frac{\varepsilon}{2},$$

$$x_2 = \frac{a+b}{2} - \frac{\varepsilon}{2},$$

где ε – малое положительное число

Затем вычислим значения $f(x_1)$ и $f(x_2)$.

Возможны следующие варианты:

1. Если $f(x_1) < f(x_2)$, то точка оптимума x^* должна лежать между a и x_2 . (рис. 3.1 а)
2. Если $f(x_1) > f(x_2)$, то $x_1 < x^* < b$ (рис. 3.1 б)

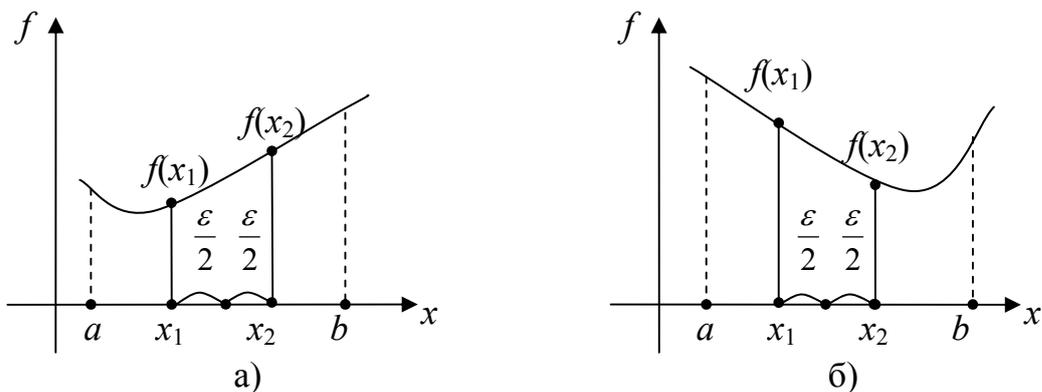


Рис. 3.1

В любом случае интервал, не содержащий точку x^* , исключается из дальнейшего рассмотрения, исходный интервал неопределенности уменьшается.

Полученный интервал можно вновь разделить точками x_1 и x_2 точно так же, как это было сделано с исходным интервалом неопределенности.

Продолжая процесс деления можно уменьшить интервал, содержащий точку локального минимума до величины ε . Блок-схема алгоритма метода дихотомии приведена на рис. 3.2.

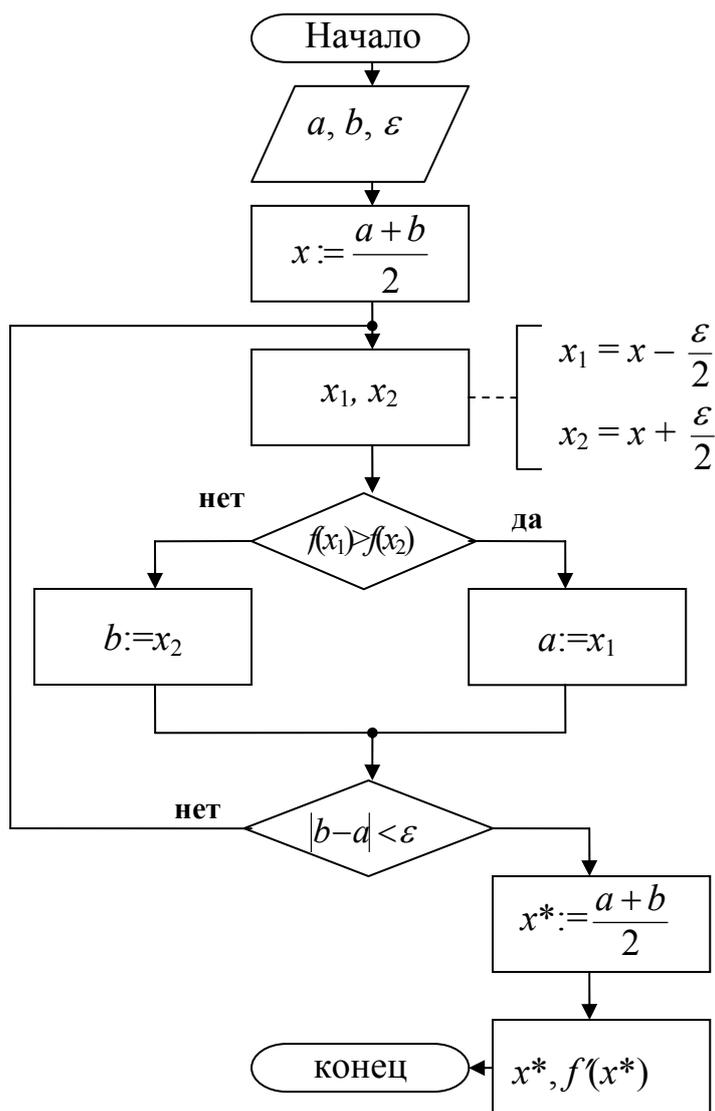


Рис. 3.2. Блок-схема алгоритма метода дихотомии

3.2.2. Метод золотого сечения

В методе дихотомии на каждом шаге следует выбирать две внутренние точки из интервала неопределенности и определять значение функции в этих точках.

Желательно задать такое правило выбора двух внутренних точек, чтобы одна из них всегда использовалась в качестве внутренней и для следующего интервала. Тогда число вычислений функции сокращается в два раза и на каждой итерации придется вычислять только одно новое значение функции.

В методе золотого сечения в качестве двух внутренних точек выбираются точки золотого сечения. Точка производит «золотое сечение» отрезка, если отношение длины всего отрезка к большей части равно отношению большей части к меньшей.

Золотое сечение отрезка $[a, b]$ производят симметрично расположенные точки: $x_1 = a + \alpha(b - a)$, $x_2 = b - \alpha(b - a)$, где $\alpha \approx 0,382$.

Кроме того, точка x_1 производит золотое сечение отрезка $[a, x_2]$, а точка x_2 – отрезка $[x_1, b]$ (рис.3.3).

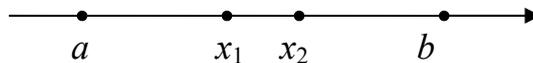


Рис 3.3

Блок-схема алгоритма метода золотого сечения приведена на рис.3.4.

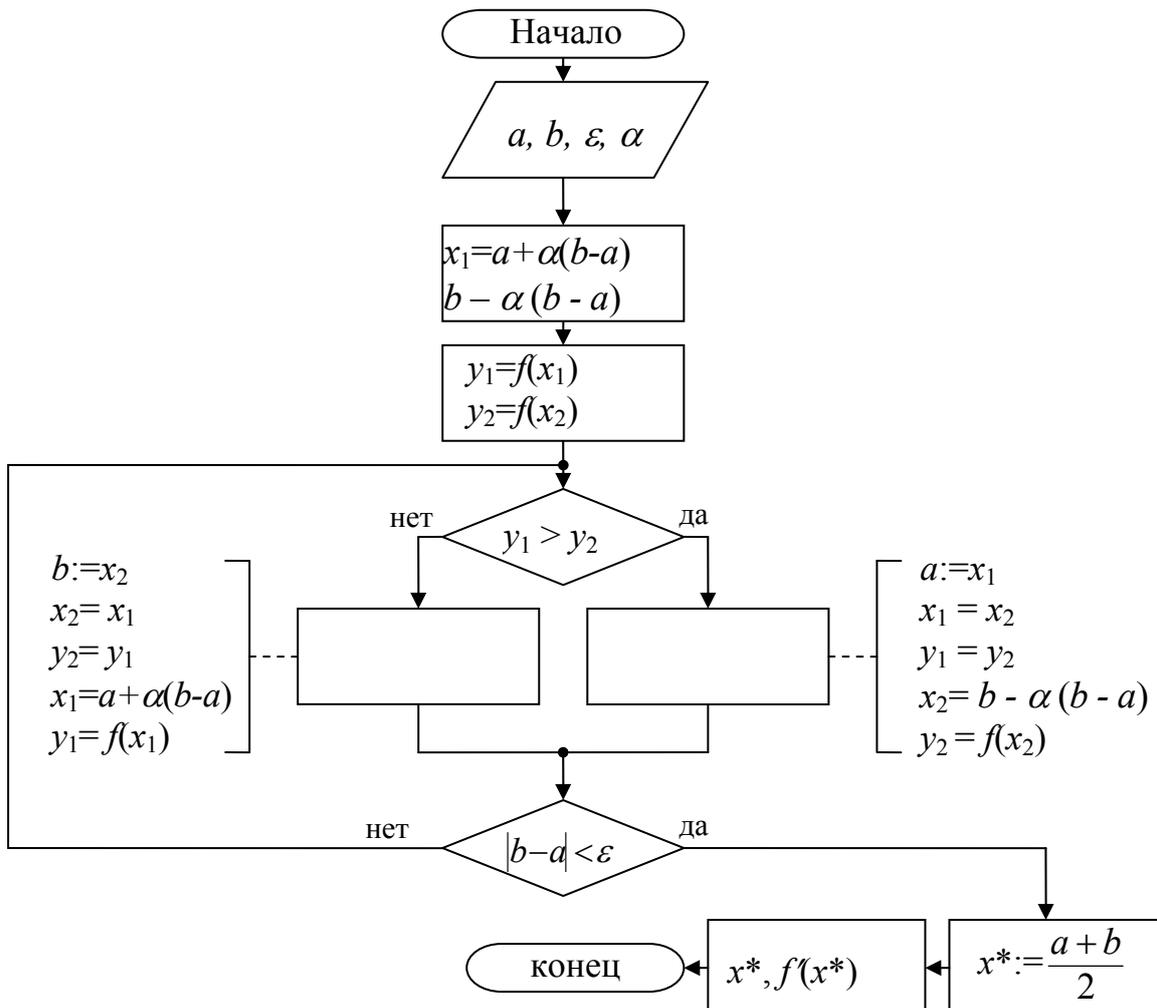


Рис. 3.4. Блок-схема алгоритма метода золотого сечения

3.2.3. Метод Фибоначчи

Последовательность чисел F_n , $n = 0, 1, 2, \dots$, определяемая условиями $F_0 = F_1 = 1$; $F_2 = F_0 + F_1 = 2$; $F_3 = F_1 + F_2 = 3$; \dots ; $F_{n+1} = F_{n-1} + F_n$ называется последовательностью Фибоначчи.

Формируем последовательности чисел Фибоначчи F_0, F_1, \dots до тех пор, пока не найдем F_n , такое, что

$$F_n \geq (b - a) / \varepsilon.$$

Обозначим $\Delta = (b - a) / F_n \leq \varepsilon$.

Полагаем $x_1 = a + F_k \cdot \Delta$; $x_2 = b - F_k \cdot \Delta$.

Вычисляем значения $f(x_1), f(x_2)$ и сравниваем их.

Если $f(x_1) > f(x_2)$, то полагаем $a = x_1$, иначе полагаем $b = x_2$.

Начальное значение $k = N - 2$.

Затем значение k уменьшается на 1.

Вычисления x_1, x_2 и сравнения значений функции повторяются до тех пор, пока длина интервала не станет меньше 2ε , т.е. при $b - a \leq 2\varepsilon$ процесс заканчивается и приближенное решение задачи может быть $\tilde{x} = (a + b) / 2$.

3.2.4. Метод квадратичной интерполяции

В методах дихотомии и золотого сечения определяется малый интервал, в котором находится минимум функции. Возможен и другой подход в определении минимального значения функции одной переменной.

В методе квадратичной интерполяции используются значения функции в трех различных точках для аппроксимации[©] функции полиномом второй степени в небольшой области значений. Затем положение минимума исходной функции аппроксимируется положением минимума квадратичной функции.

Если известны значения функции $f(x)$ в трех различных точках x_1, x_2, x_3 , равные соответственно f_1, f_2, f_3 , то функция $f(x)$ может быть аппроксимирована квадратичной функцией

$$\varphi(x) = Ax^2 + Bx + C, \quad (3.8)$$

где A, B и C определяются из уравнений (правило Крамера)

$$\begin{cases} Ax_1^2 + Bx_1 + C = f_1, \\ Ax_2^2 + Bx_2 + C = f_2, \\ Ax_3^2 + Bx_3 + C = f_3. \end{cases} \quad (3.9)$$

Выразим A, B , и C из уравнений (3.9)

[©] Аппроксимация – приближенное выражение математических величин (чисел, функций и пр.) через другие, более простые. Обычно стараются построить аппроксимацию с любой степенью точности. Аппроксимация непрерывной на отрезке $a \leq x \leq b$ функций $y = f(x)$ алгебраическими многочленами (полиномами) возможна с любой степенью точности, где мерой отклонения многочлена $P(x)$, от функций $f(x)$ служит максимум абсолютной величины разности $f(x)$ и $P(x)$, т.е. $\max_{[a,b]} |f(x) - P(x)| \leq \varepsilon$ (ε – точность).

$$\begin{aligned}
 A &= [(x_3 - x_2)f_1 + (x_1 - x_3)f_2 + (x_2 - x_1)f_3] / \Delta, \\
 B &= [(x_2^2 - x_3^2)f_1 + (x_3^2 - x_1^2)f_2 + (x_1^2 - x_2^2)f_3] / \Delta, \\
 C &= [x_2x_3(x_3 - x_2)f_1 + x_3x_1(x_1 - x_3)f_2 + x_1x_2(x_2 - x_1)f_3] / \Delta,
 \end{aligned}
 \tag{3.10}$$

где $\Delta = (x_1 - x_2)(x_2 - x_3)(x_3 - x_1)$.

Функция $\varphi(x)$ имеет минимум в точке $\bar{x} = -\frac{B}{2A}$, если $A > 0$.

Таким образом, можно аппроксимировать точку минимума функции $f(x)$ значением

$$\bar{x} = \frac{1}{2} \left[\frac{(x_2^2 - x_3^2)f_1 + (x_3^2 - x_1^2)f_2 + (x_1^2 - x_2^2)f_3}{(x_2 - x_3)f_1 + (x_3 - x_1)f_2 + (x_1 - x_2)f_3} \right].
 \tag{3.11}$$

Поиск минимального значения функции происходит в следующем порядке.

В качестве исходных данных необходимо задать:

начальную точку x_1 , величину шага $\Delta x > 0$, ε_1 и ε_2 – малые положительные числа, характеризующие точность.

Далее следует вычислить $x_2 = x_1 + \Delta x$, $f(x_1)$, $f(x_2)$.

Возможны два варианта выбора точки x_3 :

1. ($v = 1$) Если $f(x_1) > f(x_2)$, то $x_3 = x_1 + 2\Delta x$.

В этом случае начало интервала $x_N = x_1$, конец интервала $x_K = x_3$ (рис.3.5 а)

2. ($v = 2$) Если $f(x_1) < f(x_2)$, то $x_3 = x_1 - \Delta x$.

Начало интервала $x_N = x_3$, конец интервала $x_K = x_2$ (рис. 3.5 б)

Затем следует определить $f_3 = f(x_3)$, выбрать минимальное значение функции

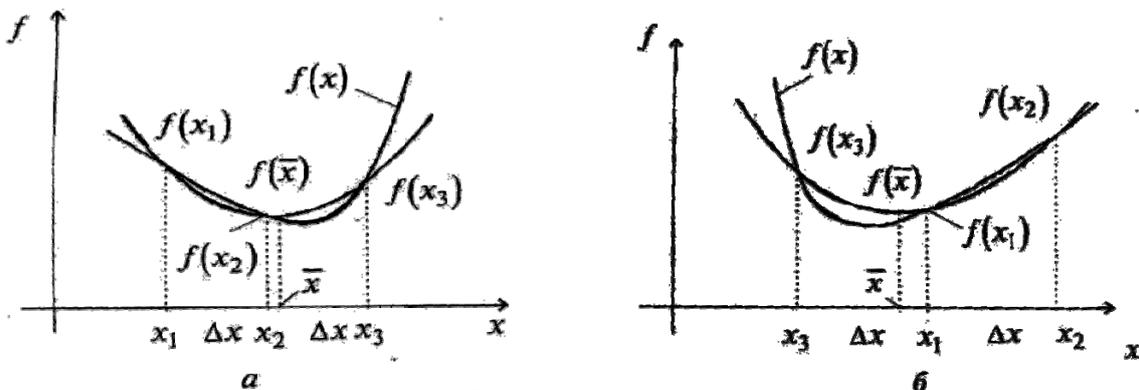


Рис. 3.5

$F_{\min} = \min(f_1, f_2, f_3)$ и соответствующее минимальному значению функции значения аргумента x_{\min} , а также, используя точки x_1, x_2, x_3 , найти значение \bar{x} в соответствии с уравнением (3.11) и вычислить $f(\bar{x})$.

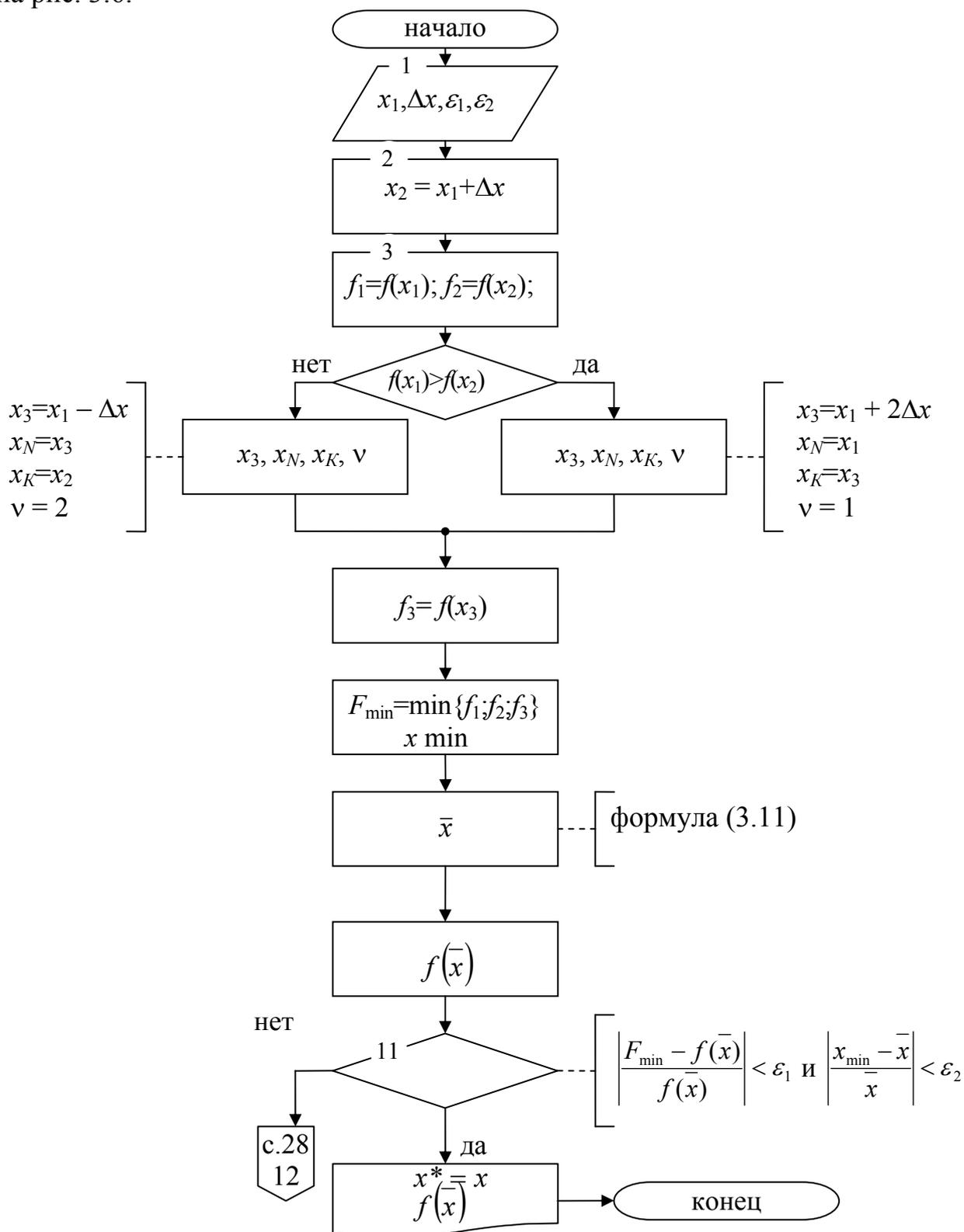
Процесс поиска заканчивается, когда полученная точка отличается от наилучшей из трех опорных точек не более чем на заданную величину, т.е. выполняются условия [1]

$$\left| \frac{F_{\min} - f(\bar{x})}{f(\bar{x})} \right| < \varepsilon_1 \quad \text{и} \quad \left| \frac{x_{\min} - \bar{x}}{\bar{x}} \right| < \varepsilon_2.
 \tag{3.12}$$

Если хотя бы одно из условий (3.12) не выполнено и $\bar{x} \in [x_N, x_K]$, то следует вновь выбрать три опорные точки: взять точку \bar{x} и две ближайшие точки (среди x_1, x_2, x_3) по обе стороны от \bar{x} и, перенумеровав точки в естественном порядке, снова перейти к определению точки минимума интерполяционного полинома.

Если хотя бы одно из условий (3.12) не выполнено и $\bar{x} \notin [x_N, x_K]$, то следует положить $x_1 = \bar{x}$ и повторить все вычисления.

Блок-схема алгоритма метода квадратичной интерполяции представлена на рис. 3.6.



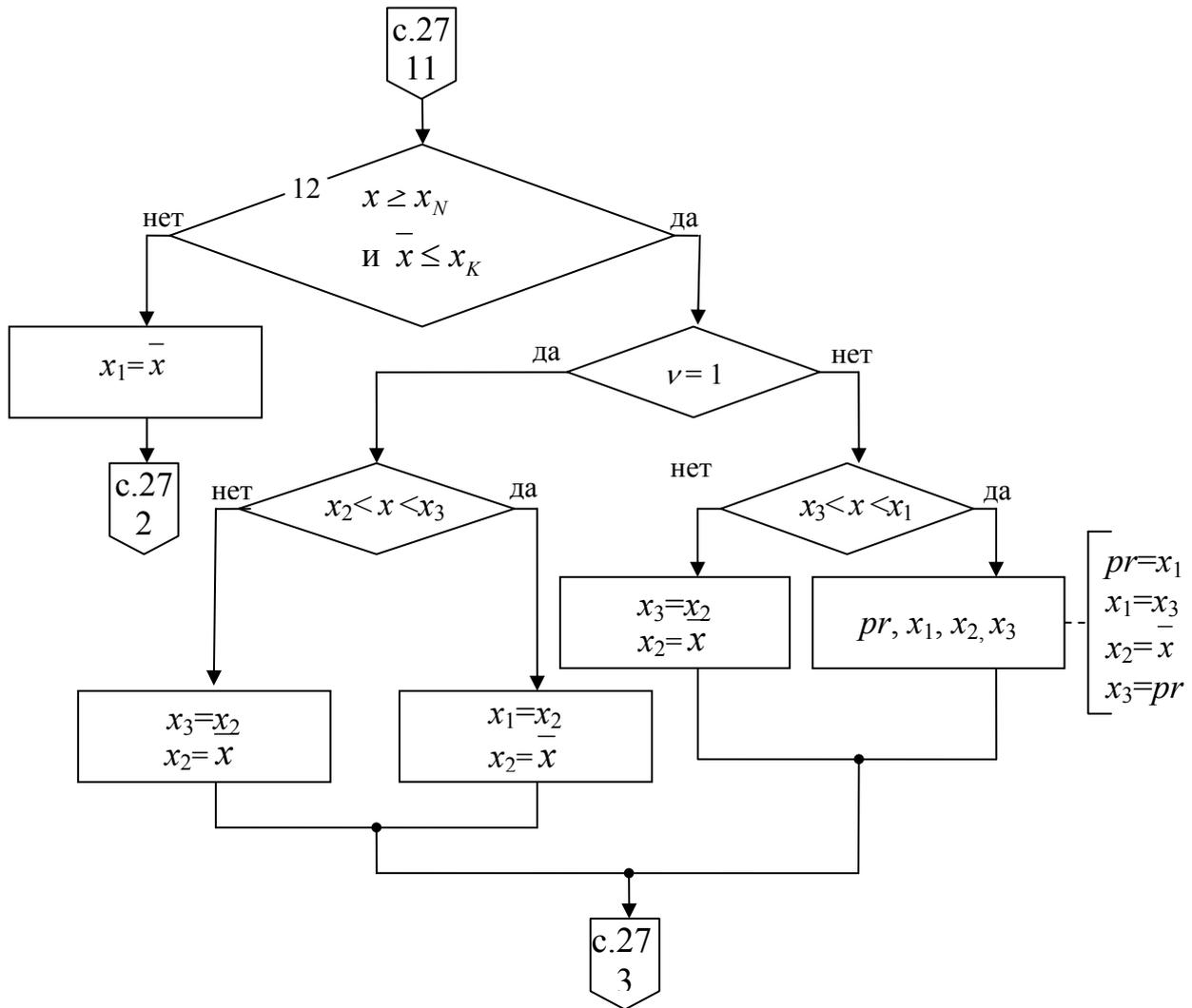


Рис. 3.6 Блок-схема алгоритма квадратичной интерполяции

4. Метод пошагового поиска

Рассмотренные выше методы одномерной оптимизации требуют знания границ a и b отрезка, содержащего x_{\min} . Однако очень часто такая информация отсутствует. В этих случаях можно прибегнуть к *методу пошагового поиска*.

Указанный метод является одной из разновидностей методов перебора, однако в нем, в отличие от рассмотренных выше методов последовательного перебора, пошаговое движение начинается не от начала отрезка неопределенности, а от некоторой точки x_0 , задаваемой по каким-либо соображениям самим вычислителем. Очевидно, что при удачном выборе x_0 и правильном определении направления пошагового вычисления можно существенно сократить трудоемкость решения задачи.

Зададимся некоторым шагом $h_1 > 0$ и положим, что точки $x_0, x_0 + h_1, x_{\min}$ лежат на отрезке, где функция $f(x)$ является унимодальной. Тогда, в зависимости от соотношения значений $f(x)$ в точках x_0 и $x_0 + h_1$, можно решить вопрос о направлении пошагового движения к x_{\min} . Действительно,

$$\text{при } f(x_0) > f(x_0 + h_1) \rightarrow x_0 < x_{\min}, \quad (4.1)$$

$$\text{при } f(x_0) \leq f(x_0 + h_1) \rightarrow x_0 + h_1 > x_{\min}. \quad (4.2)$$

Поэтому для поиска x_{\min} следует двигаться в случае (4.1) от точки x_0 в положительном направлении, а в случае (4.2) от точки $x_0 + h_1$ в отрицательном направлении. После определения направления от x_0 к x_{\min} , можно воспользоваться методом последовательного перебора с некоторым начальным шагом h_2 для получения результата с заданной точностью.

Алгоритм пошагового поиска, использующий для уточнения метод перебора с дроблением шага (при $\lambda = 2$) указан на рис. 4.1, где принято $h_1 = h_2$. Соответствующая БЕЙСИК-программа для целевой функции $f(x) = \sqrt{2x^2 + x + 1}$ приведена ниже.

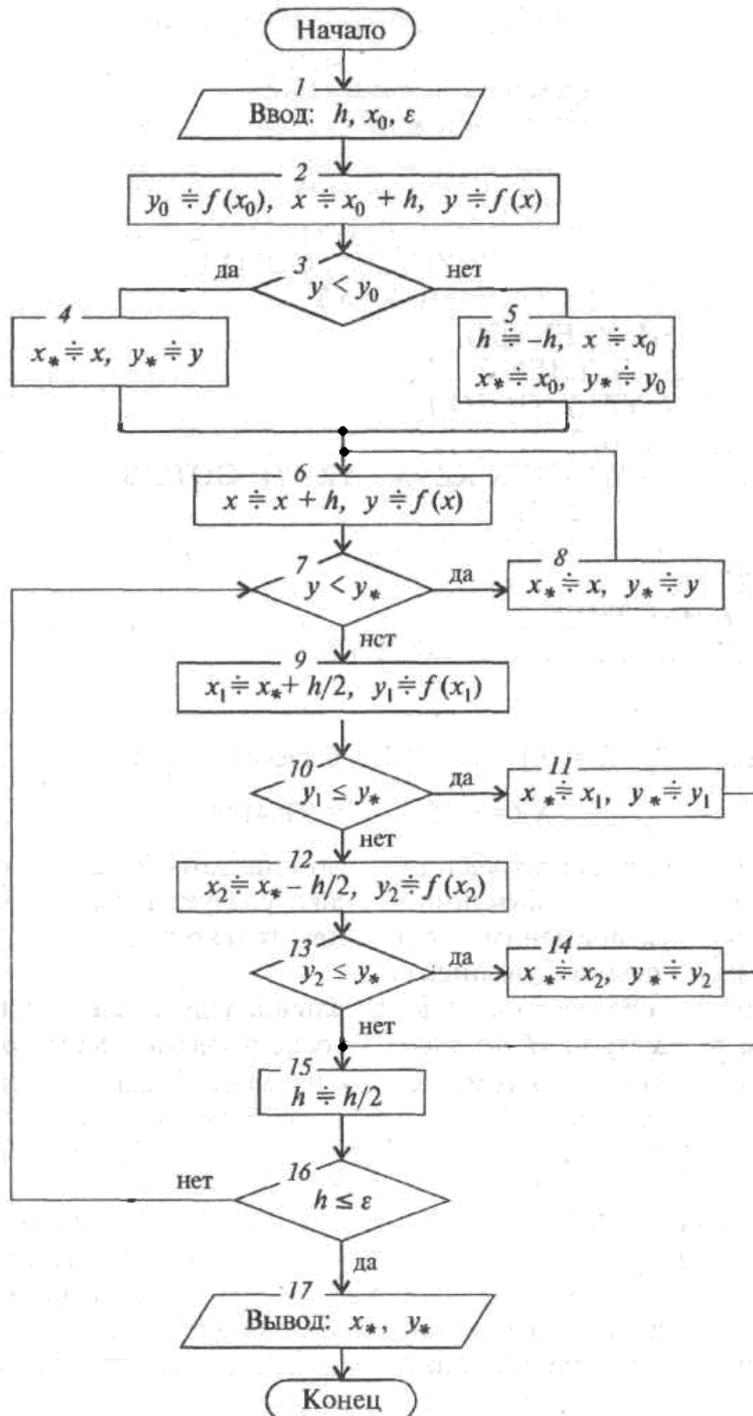


Рис. 4.1

```

PRINT "Одномерная оптимизация"
PRINT "Пошаговый поиск"
INPUT X0, H, E
DEF FNY(X)=SQR(2*XA2+X+1)
Y0=FNY(X0): X=X0+H: Y=FNY(X)
IF Y<Y0 THEN XZ=X: YZ=Y: GOTO 1
XZ=X0: YZ=Y0: H=-H: X=X0
1 X=X+H: Y=FNY(X)
IF Y=>YZ THEN 2
XZ=X: YZ=Y: GOTO 1
2 X1=XZ+H/2: Y1=FNY(X1)
IF YK=YZ THEN XZ=X1: YZ=Y1: GOTO 3
X2=XZ-H/2: Y2=FNY(X2)
IF Y2<=YZ THEN XZ=X2: YZ=Y2
3 H=H/2
IF H>E THEN 2
PRINT "XZ-"; XZ; "YZ="; YZ
END

```

При $x_0 = 0$, $h = 0,1$, $\varepsilon = 0,001$ было получено:

$$XZ=-0.25 \quad YZ=0.9354144.$$

Заметим, что для получения результата потребовалось только 6 вычислений целевой функции. Однако следует помнить, что пошаговый поиск применим, в основном, только к задачам с унимодальными целевыми функциями.

Если $f(x)$ обладает на $[a, b]$ локальной унимодальностью с известным параметром H , то следует воспользоваться методом перебора с переменным шагом, задавая начальный шаг $h_0 < H$. Если величина H неизвестна, то остается прибегнуть к самому трудоемкому методу – последовательному перебору с проверкой. При этом решение может быть получено за приемлемое время при применении современных быстродействующих ЭВМ и умеренных требованиях к точности результата. Однако следует помнить, что при очень грубом просчете и этот метод для некоторых «неспокойных» целевых функций может привести к ошибочному решению. В то же время функции такого вида нетипичны для практических задач.

5. Сравнение методов

Известно, что количество итераций N^* [4] метода последовательного перебора с переменным шагом определяется формулой последовательного перебора:

$$1,44\ln\alpha \leq N^* < 2,88\ln\alpha + 2, \quad (5.1)$$

где $\alpha = h_0/\varepsilon$, h_0 – шаг перебора, метода дихотомии:

$$2,88\ln\alpha \leq N^* < 2,88\ln\alpha + 3, \quad (5.2)$$

где $\alpha = l_0/(2\varepsilon)$, $l_0 = b - a$,

метода золотого сечения:

$$2,08\ln\alpha+2 \leq N_* < 2,08\ln\alpha + 3, \quad (5.3)$$

где $\alpha = \frac{l_0}{2\varepsilon}$, $l_0 = b - a$, $[a, b]$ – отрезок неопределенности.

Опираясь на полученные выше двусторонние оценки вида

$$\varphi_1 < N_* < \varphi_2,$$

сравним методы одномерной оптимизации по следующему усредненному показателю

$$\gamma = \frac{\varphi_1 + \varphi_2}{2},$$

характеризующему трудоемкость их применения.

Тогда получим:

– для метода дихотомии из (5.2)

$$\gamma = 2,88\ln\alpha + 2, \quad (5.4)$$

– для метода золотого сечения из (5.3)

$$\gamma = 2,08\ln\alpha + 2,5, \quad (5.5)$$

– для последовательного перебора при $\lambda = 2$ из (5.1)

$$\gamma = 2,16\ln\alpha + 1. \quad (5.6)$$

В таблице 5.1 приведены показатели γ , вычисленные по формулам (5.4), (5.5), (5.6) при $l_0 = 1$ и различных ε .

Таблица 5.1

ε	Методы		
	дихотомии	золотого сечения	перебора при $\lambda = 2$
0,1	6,64	5,85	4,48
0,01	13,64	10,64	9,45
0,001	19,90	15,43	14,43
0,0001	26,53	20,22	19,40

Как видно из таблицы, в достаточно широком диапазоне изменения ε метод последовательного перебора с переменным шагом при $\lambda = 2$ имеет по введенному нами показателю трудоемкости некоторые преимущества даже по сравнению с методом золотого сечения.

6. Варианты заданий

I. Воспользовавшись классическим подходом – исследование функции на экстремум

для нечетных вариантов: найти оптимальное соотношение между высотой h и радиусом r закрытого цилиндрического резервуара из условия минимума его поверхности емкостью $V=30 \text{ м}^3$;

для четных вариантов: найти оптимальное соотношение между высотой h и радиусом r закрытого конического резервуара из условия минимума его поверхности емкостью $V=30 \text{ м}^3$.

Расчетные формулы : для цилиндра $V_{ц}=\pi r^2 h$, $S_{н.н.ц}=2\pi r^2+2\pi r h$,

для конуса $V_{к}=\pi r^2 h/3$, $S_{н.н.к}=\pi r^2+2\pi r\sqrt{h^2+r^2}$.

II. Найти точку минимума (максимума) функции $f(x)$ на отрезке $[a, b]$ с точностью ε . Составить программы на языке Pascal, реализующие указанные в вариантах заданий алгоритмы, с использованием подпрограмм. Отрезок неопределенности $[a, b]$ найти графически.

В основной программе предусмотреть ввод исходных данных a, b, ε , вывод результата задания, вывод количества итераций, обращение к подпрограммам. Провести сравнительный анализ результатов, полученных различными методами, сделать выводы.

Вид функции $f(x)$ таков:

№ варианта	$F(x)$	№ варианта	$F(x)$
1	$F(x)=4x^2-16x+4$	11	$F(x)=-56x^2-16x+26$
2	$F(x)=-4x^2-16x+4$	12	$F(x)=7x^2+5x+25$
3	$F(x)=6x^2-72x-6$	13	$F(x)=-7x^2-16x+25$
4	$F(x)=8x^2+8x+4$	14	$F(x)=56x^2+120x+56$
5	$F(x)=8x^2+8x+4$	15	$F(x)=56x^2+120x+56$
6	$F(x)=24x^2+40x+25$	16	$F(x)=12x^2-16x+2$
7	$F(x)=24x^2+40x+25$	17	$F(x)=-12x^2-16x+2$
8	$F(x)=38x^2-16x+10$	18	$F(x)=62x^2-16x+1$
9	$F(x)=38x^2-16x+10$	19	$F(x)=-62x^2-16x+1$
10	$F(x)=56x^2-16x+26$	20	$F(x)=7x^2-16x+9$

Исходные данные и расчетные методы для студентов I и IV курсов.

I курс	IV курс
$\varepsilon=0,01$	$\varepsilon=0,0001$
Методы: <i>четные варианты</i>	
последовательного перебора и золотого сечения	последовательного перебора, золотого сечения и квадратичной интерполяции
Методы: <i>нечетные варианты</i>	
дихотомии и Фибоначчи	последовательного перебора с переменным шагом, дихотомии и Фибоначчи

СОДЕРЖАНИЕ

1. Общие сведения.....	3
2. Одномерная оптимизация.....	5
2.1 Метод последовательного перебора.....	7
2.2 Последовательный перебор с переменным шагом.....	13
3. Классические методы поиска безусловного экстремума.....	20
3.1. Общие положения.....	20
3.2 Численные методы одномерной минимизации.....	22
3.2.1 Метод дихотомии.....	22
3.2.2 Метод золотого сечения.....	24
3.2.3 Метод Фибоначчи.....	25
3.2.4 Метод квадратичной интерполяции.....	25
4. Метод пошагового поиска.....	28
5. Сравнение методов.....	30
6. Варианты заданий.....	32
Список литературы.....	33

Список литературы

1. Пантелеев А.В., Летова Т.А. Методы оптимизации в примерах и задачах. – М.: Высшая школа, 2002.
2. Банди Б. Методы оптимизации. Вводный курс. – М.: Радио и связь, 1988.
3. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. – М.: Мир, 1985.
4. Супрун А.Н., Найдаенко В.В. Вычислительная математика для инженеров-экологов: методическое пособие. – М.: Изд-во АСВ, 1996.
5. Плис А.И., Сливина Н.А. Лабораторный практикум по высшей математике. – М.: Высшая школа, 1983.

Котликова Вера Яковлевна

Сысолятина Лидия Геннадьевна

ОДНОМЕРНАЯ ОПТИМИЗАЦИЯ

Методические указания
для выполнения лабораторных работ
по курсу «Информатика», «Методы вычислений»
для студентов
по направлениям 190000, 200000, 260000, 280000, 010000
специальностей 140211, 150202, 151001, 150002
192001, 190601, 190603, 190702, 200503, 260601, 280101, 010101

Редактор Н.М. Кокина

.....
Подписано к печати 11.07.08 Формат 60*84 1/16. Бумага тип. N1
Печать трафаретная Усл. печ. л. 2,25 Уч. – изд. л. 2,25
Заказ Тираж 100 Цена свободная
.....

РИЦ Курганского государственного университета.
640669, г. Курган, ул. Гоголя, 25.
Курганский государственный университет.