Метод деления отрезка пополам

После того как установлены границы интервала, содержащего точку оптимума, можно применить более сложную процедуру уменьшения интервала поиска с целью получения уточненных оценок координат оптимума. Величина подынтервала, исключаемого на каждом шаге, зависит от расположения пробных точек x_1 и x_2 внутри интервала поиска. Поскольку местонахождение точки оптимума априори неизвестно, целесообразно предположить, что размещение пробных точек должно обеспечивать уменьшение интервала в одном и том же отношении. Кроме того, в целях повышения эффективности алгоритма необходимо потребовать, чтобы указанное отношение было максимальным. Подобную стратегию иногда называют минимаксной стратегией поиска.

Метод деления интервала пополам. Рассматриваемый метод позволяет исключать *в точности* половину интервала на каждой итерации. Иногда этот метод называют *трехточечным поиском на равных интервалах*, поскольку его реализация основана на выборе трех пробных точек, равномерно распределенных в интервале поиска. Ниже приводится описание основных шагов поисковой процедуры, ориентированной на нахождение точки минимума функции $f(x)$ в интервале (a, b).

Шаг 1. Положить $x_m=(a+b)/2$ и $L=b-a$. Вычислить значение $f(x_m)$.

Шаг 2. Положить $x_1=a+L/4$ и $x_2=b-L/4$. Заметим, что точки x_1, x_m и x_2 делят интервал (a, b) на четыре равные части. Вычислить значения $f(x_1)$ и $f(x_2)$.

Шаг 3. Сравнить $f(x_1)$ и $f(x_m)$.

(1) Если $f(x_1)<f(x_m)$, исключить интервал (x_m, b), положив $b=x_m$.

Средней точкой нового интервала поиска становится точка x_1. Следовательно, необходимо положить $x_m = x_1$.

Перейти к шагу 5.

(2) Если $f(x_1) \geq f(x_m)$, перейти к шагу 4.

Шаг 4. Сравнить $f(x_2)$ и $f(x_m)$.

(1) Если $f(x_2) < f(x_m)$, исключить интервал (a, x_m), положив $a = x_m$. Так как средней точкой нового интервала становится точка x_2, положить $x_m = x_2$. Перейти к шагу 5.

(2) Если $f(x_2) \geqslant f(x_m)$, исключить интервалы (a, x_1) и (x_2, b). Положить $a = x_1$ и $b = x_2$. Заметим, что x_m продолжает оставаться средней точкой нового интервала. Перейти к шагу 5.

Шаг 5. Вычислить $L = b - a$. Если величина $|L|$ мала, закончить поиск. В противном случае вернуться к шагу 2.

Замечания

1. На каждой итерации алгоритма исключается в точности половина интервала поиска.

2. Средняя точка последовательно получаемых интервалов всегда совпадает с одной из пробных точек x_1, x_2 или x_m, найденных на предыдущей итерации. Следовательно, на каждой итерации требуется не более двух вычислений значения функции.

3. Если проведено n вычислений значения функции, то длина полученного интервала составляет $(1/2)^{n/2}$ величины исходного интервала.

4. В работе [2] показано, что из всех методов поиска на равных интервалах (двухточечный, трехточечный, четырехточечный и т. д.) трехточечный поиск, или метод деления интервала пополам, отличается наибольшей эффективностью.

Пример 2.6. Метод деления интервала пополам

Минимизировать $f(x)=(100-x)^2$ в интервале $60 \leqslant x \leqslant 150$. Здесь $a=60$, $b=150$ и $L=150-60=90$.

$$x_m=(60+150)/2=105.$$

Итерация 1

$$x_1=a+(L/4)=60+(90/4)=82,5,$$
$$x_2=b-(L/4)=150-(90/4)=127,5,$$
$$f(82,5)=306,25>f(105)=25,$$
$$f(127,5)=756,25>f(105).$$

Таким образом, исключаются интервалы (60, 82,5) и (127,5, 150). Длина интервала поиска уменьшается с 90 до 45.

Итерация 2

$$a=82{,}5, \quad b=127{,}5, \quad x_m=105,$$
$$L=127{,}5-82{,}5=45,$$
$$x_1=82{,}5+(45/4)=93{,}75,$$
$$x_2=127{,}5-(45/4)=116{,}25,$$
$$f(93{,}75)=39{,}06>f(105)=25,$$
$$f(116{,}25)=264{,}06>f(105).$$

Таким образом, интервал неопределенности равен (93,75, 116,25).

Итерация 3

$$a=93{,}75 \quad b=116{,}25, \quad x_m=105,$$

$$L=116{,}25-93{,}75=22{,}5,$$

$$x_1=99{,}375, \quad x_2=110{,}625,$$

$$f(x_1)=0{,}39<f(105)=25.$$

Таким образом, исключается интервал (105, 116,25). Новый интервал неопределенности равен (93,75, 105), его средняя точка есть 99,375 (точка x_1 на итерации 3). Отметим, что за три итерации (шесть вычислений значения функции) исходный интервал поиска длины 90 уменьшился до величины $(90)(1/2)^3=11{,}25$.

П о и с к   с   п о м о щ ь ю   м е т о д а   з о л о т о г о   с е ч е н и я. Из проведенного выше обсуждения методов исключения интервалов и минимаксных стратегий поиска можно сделать следующие выводы.

1. Если количество пробных точек принимается равным двум, то их следует размещать на одинаковых расстояниях от середины интервала.

2. В соответствии с общей минимаксной стратегией пробные точки должны размещаться в интервале по симметричной схеме таким образом, чтобы отношение длины исключаемого подынтервала к величине интервала поиска оставалось постоянным.

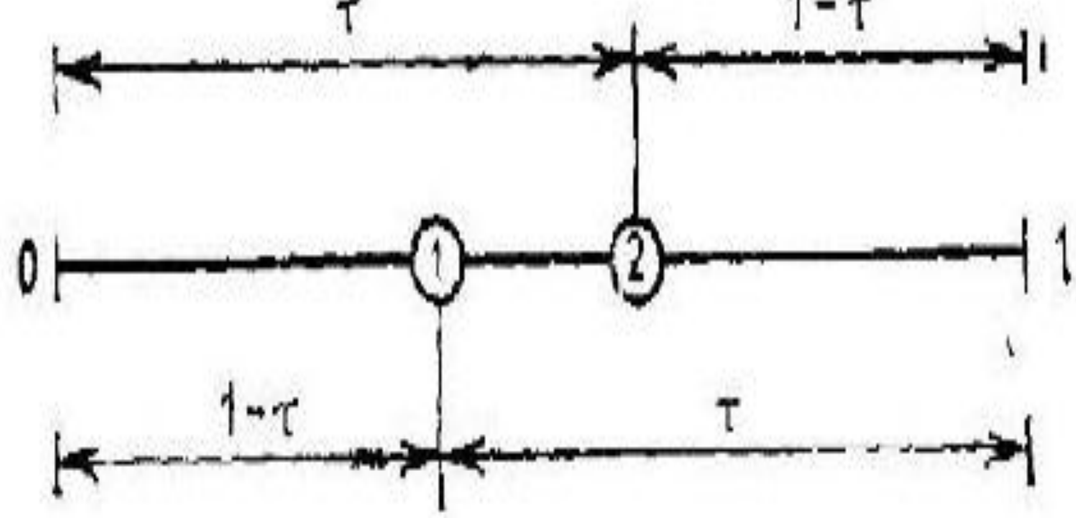

Рис. 2.11. Поиск с помощью метода золотого сечения.

3. На каждой итерации процедуры поиска должно вычисляться *только одно* значение функции в получаемой точке.

Руководствуясь этими выводами, рассмотрим симметричное расположение двух пробных точек на исходном интервале единичной длины, которое показано на рис. 2.11. (Выбор единичного интервала обусловлен соображениями удобства.) Пробные точки отстоят от граничных точек интервала на расстоянии τ. При таком симметричном расположении точек длина остающегося после исключения интервала всегда равна τ независимо от того, какое из значений функции в пробных точках оказывается меньшим. Предположим, что исключается правый подынтервал. На рис. 2.12 показано, что оставшийся подынтервал длины τ содержит одну пробную точку, расположенную на расстоянии $(1-\tau)$ от левой граничной точки.

Для того чтобы симметрия поискового образца сохранялась, расстояние $(1-\tau)$ должно составлять τ-ю часть длины интервала (которая равна τ). При таком выборе τ следующая пробная точка размещается на расстоянии, равном τ-й части длины интервала, от правой граничной точки интервала (рис. 2.13).

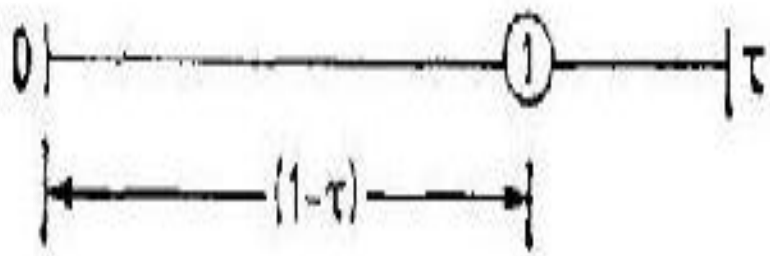

Рис. 2.12. Интервалы, полученные методом золотого сечения.

Отсюда следует, что при выборе τ в соответствии с условием $1-\tau=\tau^2$ симметрия поискового образца, показанного на рис. 2.11, сохраняется при переходе к уменьшенному интервалу, который изображен на рис. 2.13. Решая это квадратное уравнение, получаем

$$\tau = (-1 \pm \sqrt{5})/2,$$

откуда положительное решение $\tau=0{,}61803\ldots$. Схема поиска, при которой пробные точки делят интервал в этом отношении, известна

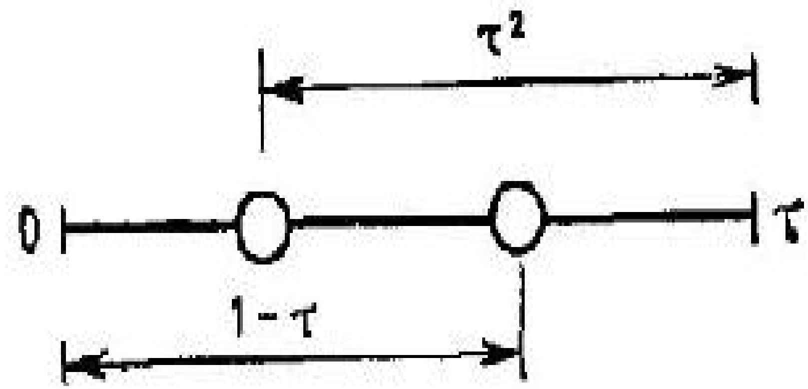

Рис. 2.13. Симметрия золотого сечения интервала.

под названием *поиска с помощью метода золотого сечения*. Заметим, что после первых двух вычислений значений функции каждое последующее вычисление позволяет исключить подынтервал, величина которого составляет $(1-\tau)$-ю долю от длины интервала поиска. Следовательно, если исходный интервал имеет единичную длину, то величина интервала, полученного в результате N вычислений значений функции, равна τ^{N-1}. Можно показать, что поиск с помощью метода золотого сечения является асимптотически наиболее эффективным способом реализации минимаксной стратегии поиска.

Пример 2.7. Метод золотого сечения

Опять рассмотрим задачу из примера 2.6, в которой требуется минимизировать $f(x)=(100-x)^2$ в интервале $60\leqslant x\leqslant 150$.

Для того чтобы перейти к интервалу единичной длины, проведем замену переменной, положив $w=(x-60)/90$. Таким образом, задача принимает следующий вид:

$$\text{минимизировать } f(w)=(40-90w)^2$$

$$\text{при ограничении } 0\leqslant w\leqslant 1.$$

Итерация 1. $I_1=(0, 1)$; $L_1=1$. Проведем два первых вычисления значений функции:

$$w_1=\tau=0{,}618, \quad f(w_1)=244{,}0,$$
$$w_2=1-\tau=\tau^2=0{,}382, \quad f(w_2)=31{,}6.$$

Так как $f(w_2)<f(w_1)$ и $w_2<w_1$, интервал $w\geqslant w_1$ исключается.

Итерация 2. $I_2=(0, 0{,}618)$; $L_2=0{,}618=\tau$. Следующее вычисление значения функции проводится в точке

$$w_3=\tau-\tau^2=\tau(1-\tau)=\tau^3=0{,}236, \quad f(w_3)=352.$$

Так как $f(w_3)>f(w_2)$ и $w_3<w_2$, интервал $w\leqslant w_3$ исключается.

Итерация 3. $I_3=(0{,}236,\ 0{,}618)$, $L_3=0{,}382=\tau^2$. Следующее вычисление значения функции проводится в точке, расположенной на расстоянии $\tau\times$ (длина полученного интервала) от левой граничной точки интервала, или на расстоянии $(1-\tau)\times$ (длина интервала) от правой граничной точки. Таким образом,

$$w_4=0{,}618-(1-\tau)\,L_3=0{,}618-\tau^2L_3=$$
$$=0{,}618-\tau^2(\tau^2)=0{,}618-\tau^4=0{,}472,$$
$$f(w_4)=6{,}15.$$

Так как $f(w_4)<f(w_3)$ и $w_4>w_3$, интервал $w\leqslant w_3$ исключается.

В результате получен следующий интервал неопределенности: $0{,}382\leqslant w\leqslant 0{,}618$ для переменной w, или $94{,}4\leqslant x\leqslant 115{,}6$ для переменной x.

Если в процессе поиска проведено шесть вычислений значений функции, то длина результирующего интервала для переменной w равна

$$\tau^{N-1}=\tau^5=0{,}09,$$

что соответствует интервалу длины 8,1 для переменной x. Для сравнения напомним, что в аналогичной ситуации метод деления интервала пополам привел к получению интервала длины 11,25.

В общем случае если правая и левая граничные точки интервала неопределенности (обозначим их через XR и XL) известны, то координаты всех последующих пробных точек, получаемых в соответствии с методом золотого сечения, можно вычислить по формулам

$$w=\mathrm{XR}-\tau^n \text{ или } w=\mathrm{XL}+\tau^n$$

в зависимости от того, какой подынтервал был исключен на предыдущей итерации — левый или правый. В приведенных выше формулах через τ^n обозначена n-я степень τ, где n — количество вычислений значений функции.

Поиск с помощью метода золотого сечения может быть окончен либо исходя из заданного количества вычислений значений функции (и, следовательно, величины интервала неопределенности), либо по достижении относительной точности искомого значения функции. Наиболее предпочтительным является использование обоих критериев одновременно.

2.3.3. Сравнение методов исключения интервалов

Ниже проводится сравнение относительных эффективностей рассмотренных методов исключения интервалов. Обозначим длину исходного интервала неопределенности через L_1, а длину интервала, получаемого в результате N вычислений значений функции,— через L_N. В качестве показателя эффективности того или иного метода исключения интервалов введем в рассмотрение характеристику *относительного уменьшения* исходного интервала $\mathrm{FR}(N)=L_N/L_1$.

Напомним, что при использовании метода деления интервала пополам и метода золотого сечения длина получаемого интервала составляет $L_1(0,5)^{N/2}$ и $L_1(0,618)^{N-1}$ соответственно. Следовательно, относительное уменьшение интервала после N вычислений значений функции равно

$$FR(N)=\begin{cases}(0{,}5)^{N/2} & \text{для метода деления интервала пополам;}\\ (0{,}618)^{N-1} & \text{для метода золотого сечения.}\end{cases}$$

Для сравнения рассмотрим также метод равномерного поиска, в соответствии с которым оценивание функции проводится в N равноотстоящих друг от друга точках [при этом интервал L_1 делится на $(N+1)$ равных интервалов длины $L_1/(N+1)$]. Пусть x^* — точка, в которой наблюдается минимум функции $f(x)$. Тогда точка истинного минимума $f(x)$ оказывается заключенной в интервале

$$\left[\left(x^*-\frac{L_1}{N+1}\right),\quad \left(x^*+\frac{L_1}{N+1}\right)\right],$$

откуда $L_N=2L_1/(N+1)$. Следовательно, для метода равномерного поиска

$$FR(N)=2/(N+1).$$

В табл. 2.1 представлены значения $FR(N)$, соответствующие выбранным N, для трех методов поиска. Из таблицы следует, что поиск

Таблица 2.1. **Величины относительного уменьшения интервала**

Метод поиска	Количество вычислений значений функции				
	$N=2$	$N=5$	$N=10$	$N=15$	$N=20$
Метод деления интервала пополам	0,5	0,177	0,031	0,006	0,0009
Метод золотого сечения	0,618	0,146	0,013	0,001	0,0001
Метод равномерного поиска	0,667	0,333	0,182	0,125	0,095

с помощью метода золотого сечения обеспечивает наибольшее относительное уменьшение исходного интервала при одном и том же количестве вычислений значений функции. С другой стороны, можно также сравнить количества вычислений значения функции, требуемые для достижения заданной величины относительного уменьшения интервала или заданной степени точности. Если величина $FR(N)=E$ задана, то значение N вычисляется по следующим формулам:

для метода деления интервала пополам

$$N=2\ \ln(E)/\ln(0,5),$$

для метода золотого сечения

$$N=1+[\ln(E)/\ln(0,618)],$$

для метода равномерного поиска

$$N=(2/E)-1.$$

В табл. 2.2 приведены данные о количествах вычислений значений функции, необходимых для определения координаты точки минимума с заданной точностью. Следует еще раз подчеркнуть, что метод

Таблица 2.2. **Требуемые количества вычислений значений функции**

Метод поиска	Заданная точность			
	$E=0,1$	$E=0,05$	$E=0,01$	$E=0,001$
Метод деления интервала пополам	7	9	14	20
Метод золотого сечения	6	8	11	16
Метод равномерного поиска	19	39	199	1999

золотого сечения оказывается более эффективным по сравнению с остальными двумя методами, поскольку он требует наименьшего числа оцениваний значения функции для достижения одной и той же заданной точности.