Тема 06 МЕТОДЫ ПОИСКА ОДНОМЕРНОГО ЭКСТРЕМУМА В E^n

6.1. Метод дихотомии

Одной из возможных (и достаточно часто применяемых на практике) процедур выбора величины σ_k – шага по улучшающему направлению, является правило

$$\sigma_k = \arg\max_{\sigma \geq 0} F(x^k + \sigma \cdot w^k)\,.$$

Рассмотрим эту задачу подробнее.

Пусть в E^n задан функционал $F(x)$ и совокупность элементов $x \in E^n$ таких, что $x = x_0 + \sigma w \quad \forall \sigma \in (0,+\infty)$, а w – некоторый *ненулевой* элемент в E^n. Обозначим

$$f(\sigma) = F(x_0 + \sigma w) \quad \forall \sigma \in (0,+\infty).$$

Задачу отыскания числа $\sigma^* \in E^1$, максимизирующего $f(\sigma)$ принято называть задачей *одномерной оптимизации* $F(x)$ (или *поиска максимума* $F(x)$) *по направлению* w.

В координатной форме эта задача сводится к нахождению максимума функции

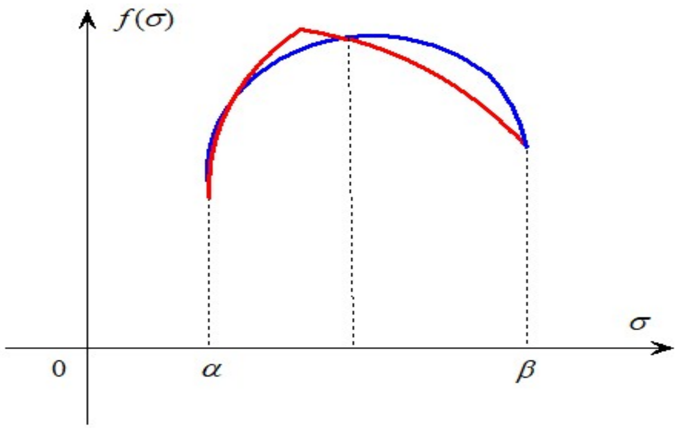
$$f(\sigma) = F(\xi_1^0 + \sigma\,\omega_1,\ \xi_2^0 + \sigma\,\omega_2, \ldots,\ \xi_n^0 + \sigma\,\omega_n),$$

зависящей от *одного* аргумента σ, где $\| w \| = \begin{Vmatrix} \omega_1 \\ \omega_2 \\ \cdots \\ \omega_n \end{Vmatrix}$ – координатное представление элемента w в E^n.

Для решения задач одномерной оптимизации возможно использование необходимых и достаточных условий существования экстремума функции одной переменной. Однако в вычислительной практике также достаточно широко применяются схемы одномерного поиска, не требующие нахождения производных, а использующие только значение $f(\sigma)$.

Определение 6.1	Отрезок $[\alpha,\beta]$ называется *отрезком локализации экстремума* функции $f(\sigma)$, если σ^* – аргумент экстремального значения $f(\sigma)$ – принадлежит $[\alpha,\beta]$, при этом само значение σ^* может быть не известно.

Определение 6.2	Функция $f(\sigma)$ называется *унимодальной*, если она имеет единственный экстремум на отрезке $[\alpha,\beta]$.



Будет справедлива

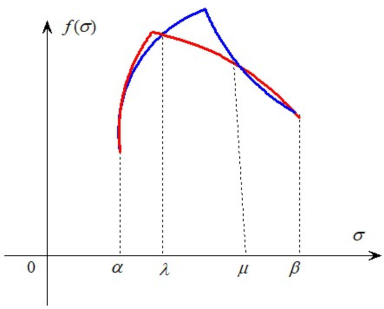
Теорема 6.1 **Пусть функция $f(\sigma)$ определена и унимодальна (на максимум) для $\sigma \in [\alpha, \beta]$. Тогда для фиксированных $\lambda, \mu \in [\alpha, \beta]$ и λ таких, что $\lambda < \mu$:**

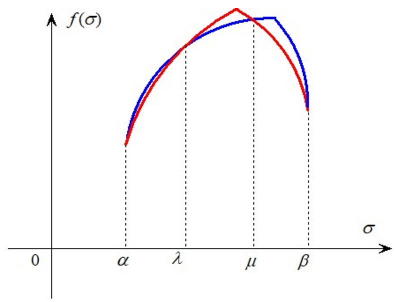
из условия $f(\lambda) > f(\mu)$ следует, что

$$f(\sigma) \geq f(\beta) \qquad \forall \sigma \in [\alpha, \mu],$$

а из условия $f(\lambda) < f(\mu)$ следует, что

$$f(\alpha) \leq f(\sigma) \qquad \forall \sigma \in [\lambda, \beta].$$





Рассмотрим наиболее часто применяемые на практике вычислительные схемы, основанные на идее последовательного сужения отрезка локализации за счет вычисления значения $f(\sigma)$ в некоторых дополнительных точках отрезка $[\alpha,\beta]$.

В дальнейшем будем считать, что значения минимизируемой функции $f(\sigma)$ найдены в двух новых точках λ и μ, принадлежащих $[\alpha,\beta]$, причем таких, что $\alpha < \lambda < \mu < \beta$.

Одной из возможных схем "одномерной оптимизации" является "метод дихотомии", основанный на следующих рассуждениях.

Ясно, что искомое значение σ^* принадлежит либо $[\alpha,\mu]$, либо $[\lambda,\beta]$ и длина нового отрезка локализации будет зависеть от выбора λ и μ.

Организуем этот выбор таким образом, чтобы длина *максимального* из двух отрезков $[\alpha,\mu]$ и $[\lambda,\beta]$ оказалась *как можно меньшей*.

Иначе говоря, необходимо исследовать в E^2 на экстремум по $\{\lambda,\mu\}$ функции вида

$$\Phi(\lambda,\mu) = \min_{\{\lambda,\mu\}} \max \{\mu-\alpha, \beta-\lambda\},$$

вид изолиний которой показан на рис. 6.1.

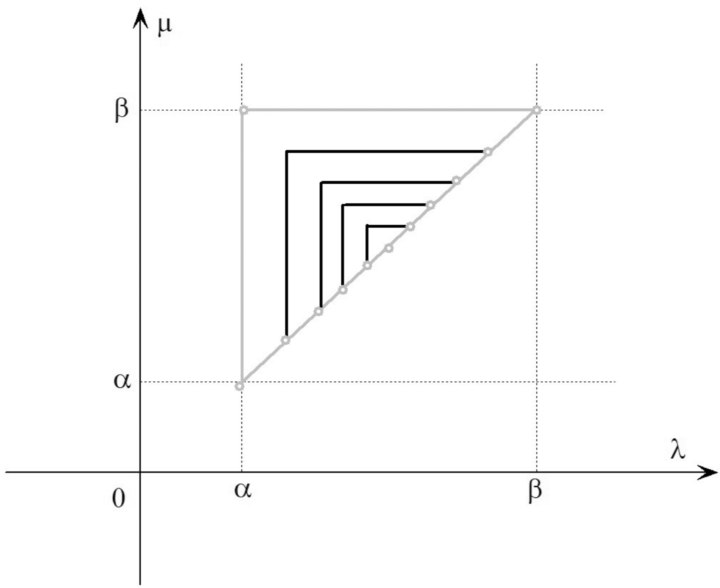


Рис. 6.1

Очевидно, что минимальное значение $\Phi(\lambda, \mu)$ достигается при $\lambda = \mu = \dfrac{\alpha + \beta}{2}$, однако это запрещено условием $\lambda < \mu$.

Действительно, при $\lambda = \mu$ дополнительное значение $f(\sigma)$ находится только для *одной* точки из $[\alpha, \beta]$, что не позволяет сократить отрезок неопределенности.

На практике рекомендуется обходить эту проблему, выбирая
\

$$\lambda = \frac{\alpha+\beta}{2} - \frac{\varepsilon}{2} \quad \text{и} \quad \mu = \frac{\alpha+\beta}{2} + \frac{\varepsilon}{2} \ ,$$

где $\varepsilon > 0$ число достаточно малое, но гарантирующее различие значений $f(\lambda)$ и $f(\mu)$ при заданной допустимой погрешности вычислений.

Процедуру уменьшения длины отрезка локализации можно проводить итеративно, приняв $\alpha_0 = \alpha$, $\beta_0 = \beta$ и используя рекуррентные соотношения

$$\lambda_k = \frac{\alpha_k + \beta_k}{2} - \frac{\varepsilon}{2} \quad \text{и} \quad \mu_k = \frac{\alpha_k + \beta_k}{2} + \frac{\varepsilon}{2}, \; k = 0,1,2,\dots.$$

При этом из теоремы 6.1 следует, что если на *k*-м шаге

$$f(\lambda_k) > f(\mu_k),$$

то нужно выбирать $\alpha_{k+1} = \lambda_k$ и $\beta_{k+1} = \beta_k$, иначе полагать $\alpha_{k+1} = \alpha_k$ и $\beta_{k+1} = \mu_k$.

Отметим, что длина отрезка локализации после k-ой итерации будет равна

$$\beta_{k+1} - \alpha_{k+1} = \frac{\beta - \alpha}{2^k} + \varepsilon\left(1 - \frac{1}{2^k}\right),$$

а число N – количество вычислений функций, необходимых в методе дихотомии для получения длины отрезка локализации меньшей чем Δ, определяется из условий (с учетом $\varepsilon << \Delta$)

$$\left(\frac{1}{2}\right)^{\frac{N}{2}} < \frac{\Delta}{\beta - \alpha} \quad \text{или же} \quad N > 2\log_2 \frac{\beta - \alpha}{\Delta}.$$

Сжатие (в расчете на одно вычисление функции) составит при этом $k < \frac{1}{\sqrt{2}} \approx 0.707$.

6.2. Метод "золотого сечения"

Метод дихотомии требует вычисления на каждом шаге *двух* новых значений $f(\sigma)$ на отрезке локализации.

\В тех случаях, когда для этого не требуются затраты значительных вычислительных ресурсов, алгоритмическая простота данного метода является основным аргументом его использования.

Однако на практике достаточно часто возникает ситуация, когда нахождение значения $f(\sigma)$ само по себе является сложной и/или ресурсоемкой задачей.

В этих случаях рекомендуется применение методов одномерной оптимизации, требующих на каждом шаге, начиная со второго, вычисления *одного* значения $f(\sigma)$ на отрезке локализации. Одной из таких схем является метод "золотого сечения".

Геометрическая задача нахождения "золотого сечения" – разбиения данного отрезка на две неравные части так, чтобы

отношение длин меньшей и большей частей равнялось отношению длины большей части к длине исходного отрезка

– рассматривалась еще Евклидом, а ее алгебраическое решение сводится к решению квадратного уравнения вида

$$\rho^2 - 3\omega\rho + \omega^2 = 0,$$

где ρ – длина меньшей части исходного отрезка длины ω.

Идея использования метода "золотого сечения" в одномерном поиске экстремума заключается в следующем.

Поскольку исходный отрезок локализации $[\alpha_k, \beta_k]$ разбивается точками λ_k и μ_k на три части, а новый промежуток неопределенности $[\alpha_{k+1}, \beta_{k+1}]$ согласно теореме 6.2 выбирается по правилу:

для $f(\lambda_k) > f(\mu_k)$ $[\alpha_{k+1}, \beta_{k+1}] = [\lambda_k, \beta_k]$,

иначе $[\alpha_{k+1}, \beta_{k+1}] = [\alpha_k, \mu_k]$,

то для уменьшения числа новых значений $f(\sigma)$, необходимых для сокращения отрезка локализации, можно потребовать, чтобы для каждой новой итерации

$$\text{либо} \quad \lambda_{k+1} = \mu_k, \quad \text{либо} \quad \mu_{k+1} = \lambda_k .$$

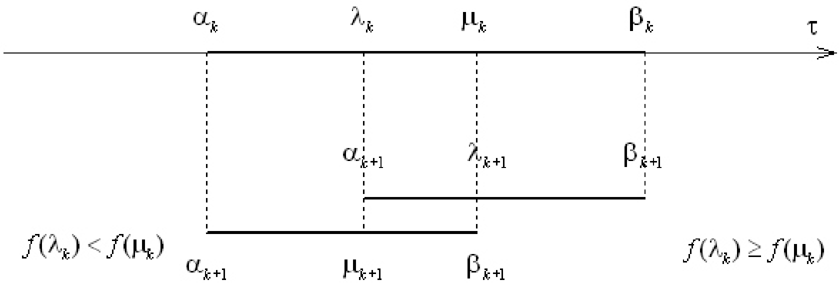


Рис. 6.2

Из рисунка 6.2 очевидно, что добиться этого можно, если

длина $[\alpha_{k+1},\beta_{k+1}]$ не зависит от того, будет ли

$f(\lambda_k) \ge f(\mu_k)$ или же $f(\lambda_k) \le f(\mu_k)$;

то есть, будет справедливо равенство $\beta_k - \lambda_k = \mu_k - \alpha_k$.

Введем в рассмотрение параметр $0 < \rho < 1$, такой что:

1°. $\lambda_k = \alpha_k + \rho(\beta_k - \alpha_k)$.

2°. $\mu_k = \alpha_k + (1-\rho)(\beta_k - \alpha_k)$ и

3°. $(\beta_{k+1} - \alpha_{k+1}) = (1-\rho)(\beta_k - \alpha_k)$.

Если оказывается, что $f(\lambda_k) < f(\mu_k)$, то полагаем:

4°. $\alpha_{k+1} = \alpha_k$ и 5°. $\beta_{k+1} = \mu_k$.

И учтем, наконец, что:

6°. $\mu_{k+1} = \lambda_k$.

Из равенства 2°, верного для любого k, получаем

$$\mu_{k+1} = \alpha_{k+1} + (1-\rho)(\beta_{k+1} - \alpha_{k+1}),$$

что в силу 4°, 5° и 6° означает

$$\lambda_k = \alpha_k + (1-\rho)(\mu_k - \alpha_k).$$

Подставляя в это равенство выражения 1° и 2°, получаем

$$\alpha_k + \rho(\beta_k - \alpha_k) = \alpha_k + (1-\rho)((\alpha_k + (1-\rho)(\beta_k - \alpha_k)) - \alpha_k),$$

а после упрощений

$$\rho(\beta_k - \alpha_k) = (1-\rho)^2 (\beta_k - \alpha_k).$$

Наконец, поскольку это соотношение должно быть верным для любого k, то коэффициент ρ является корнем квадратного уравнения

$$\rho^2 - 3\rho + 1 = 0\ ,$$

являющегося уравнением задачи нахождения "золотого сечения", и, кроме того, $\rho < 1$, то

$$\rho = \frac{3-\sqrt{5}}{2} \qquad \text{и} \qquad 1-\rho = \frac{\sqrt{5}-1}{2} \approx 0.618 < \frac{1}{\sqrt{2}} \approx 0.707\ .$$

Случай $f(\lambda_k) \ge f(\mu_k)$ приводит к точно такому же результату (проверьте это самостоятельно).

Для метода "золотого сечения" число N – вычислений значения функции, необходимых для получения длины отрезка локализации меньшей, чем Δ, в силу соотношения 3°, определяется неравенством

$$\frac{\beta-\alpha}{\tau^{N-1}} < \Delta \qquad \text{или же} \qquad N > \log_\tau \frac{\tau(\beta-\alpha)}{\Delta}\text{, где } \tau = \frac{1+\sqrt{5}}{2}.$$

6.3. Метод Фибоначчи

Метод "золотого сечения" является более эффективным по сравнению с методом дихотомии потому, что отношение $\dfrac{\beta_{k+1}-\alpha_{k+1}}{\beta_k-\alpha_k}$ – коэффициент сжатия отрезка локализации – для "золотого сечения" лучше, чем для дихотомии.

Для данных методов не требуется, вообще говоря, определять число шагов, обеспечивающих достижение требуемой точности. При их использовании достаточно на каждом шаге проверять выполнение условия $\beta_k-\alpha_k \le \Delta$, где Δ – требуемая точность.

Существуют, однако, алгоритмы одномерной оптимизации, для которых необходимо заранее знать значение N – максимального числа шагов при его реализации. К таким методам относится алгоритм, предложенный и обоснованный в 1953 году Дж. Кифером, и получивший название "*метода Фибоначчи*".

Теоретическая оценка коэффициента сжатия в методе Фибоначчи несколько лучше, чем для метода "золотого сечения", в то время как реализация метода Фибоначчи сложнее и потому в вычислительной практике этот метод используется реже.

Однако отметим, что при $N \to +\infty$ эффективность обеих схем асимптотически сближается.

Алгоритм метода Фибоначчи основан на использовании "чисел Фибоначчи", являющихся членами числовой последовательности $\{F_k\}$, где

$$F_0 = F_1 = 1\,;$$
$$F_k = F_{k-1} + F_{k-2}\ ;\ k = 2,3,\ldots$$

или же в явном виде

$$F_k = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{k+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{k+1}\right]$$

,

описывается следующими рекуррентными соотношениями

$$\lambda_k = \alpha_k + \frac{F_{N-k-1}}{F_{N-k+1}}(\beta_k - \alpha_k),$$

$$\mu_k = \alpha_k + \frac{F_{N-k}}{F_{N-k+1}}(\beta_k - \alpha_k).$$

Можно показать, что в методе Фибоначчи либо $\lambda_{k+1} = \mu_k$, либо $\mu_{k+1} = \lambda_k$, то есть на каждой новой итерации требуется вычисление лишь *одного* нового значения $f(\sigma)$.

Полное число шагов метода Фибоначчи N фиксировано, что позволяет получить несколько лучшую (по сравнению с методом "золотого сечения") оценку эффективности. Непосредственная проверка показывает, что

$$\beta_{N-1} - \alpha_{N-1} = \frac{2(\beta - \alpha)}{F_N} .$$

Тогда, приняв за последнее приближение точку, близкую к середине отрезка $[\alpha_{N-1}, \beta_{N-1}]$, приходим к оценке

$$\beta_N - \alpha_N \approx \frac{\beta - \alpha}{F_N}, \text{ где } F_N \approx \frac{\tau^{N+1}}{\sqrt{5}} .$$

Алгоритм вычислений в методе Фибоначчи практически аналогичен схеме "золотого сечения", показанной на рис. 6.1, за двумя исключениями:

1) коэффициент сжатия в нем равен $\dfrac{F_{N-k}}{F_{N-k+1}}$, то есть меняется от итерации к итерации, стремясь, однако, асимптотически к значению $\dfrac{\sqrt{5}-1}{2} \approx 0.618$.

2) для нахождения значений λ_k и μ_k необходимо *знать заранее* N – полное число шагов метода Фибоначчи, которое удовлетворяет оценкам:
$$\frac{\beta-\alpha}{F_N} < \Delta \quad \text{или же} \quad N > \log_\tau \frac{(\beta-\alpha)\sqrt{5}}{\tau\Delta} .$$