МЕТОДЫ ПОИСКА ОДНОМЕРНОГО ЭКСТРЕМУМА В E^n

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 Ω – совокупность элементов $x \in E^n$ таких, что $x = x_0 + \tau w; \forall \tau \in (-\infty, +\infty)$, а w – некоторый *ненулевой* элемент в E^n. Обозначим

$$f(\tau) = F(x_0 + \tau w); \forall \tau \in (-\infty, +\infty) .$$

Задачу отыскания экстремального для $f(\tau)$ числа $\tau^* \in E^1$ принято называть задачей *одномерной оптимизации* $F(x)$ (или *поиска экстремума* $F(x)$) *по направлению* w.

В вычислительной практике достаточно широко применяются схемы одномерного поиска, не требующие нахождения производных, а использующие только значение $f(\tau)$.

Определение

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

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

Будет справедлива

Теорема

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

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

$$f(\tau)\ge f(\lambda),\ \forall\tau\in[\alpha,\lambda],$$

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

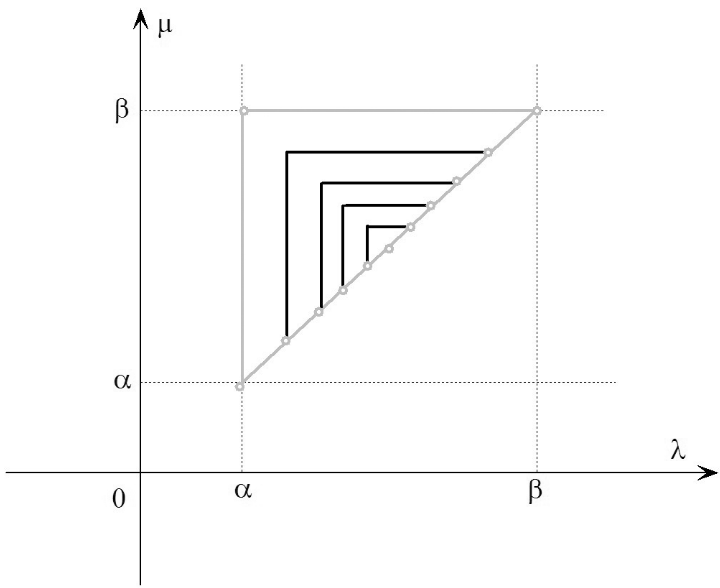
$$f(\mu)\le f(\tau),\ \forall\tau\in[\mu,\beta].$$

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

Метод дихотомии

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

Ясно, что искомое значение τ^* 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 функционал вида $L(\lambda,\mu)=\min\limits_{\lambda}\max\limits_{\mu}\{\mu-\alpha,\beta-\lambda\}$, вид изолиний которого показан на следующем рисунке.



Очевидно, что минимальное значение $L(\lambda, \mu)$ достигается при $\lambda = \mu = \frac{\alpha + \beta}{2}$, однако это запрещено условием $\lambda < \mu$. Действительно, при $\lambda = \mu$ дополнительное значение $f(\tau)$ 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

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

При этом 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 $\varepsilon << \Delta$)

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

Метод "золотого сечения"

Метод дихотомии требует вычисления на каждом шаге *двух* новых значений $f(\tau)$ на отрезке локализации. В тех случаях, когда для этого не требуются затраты значительных вычислительных ресурсов, алгоритмическая простота данного метода является основным аргументом в пользу дихотомии.

Однако на практике достаточно часто возникает ситуация, когда нахождение значения $f(\tau)$ 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 $f(\tau)$ на отрезке локализации. Одной из таких схем является метод "золотого сечения".

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 $[\alpha_{k+1}, \beta_{k+1}]$ выбирается по правилу:

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

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

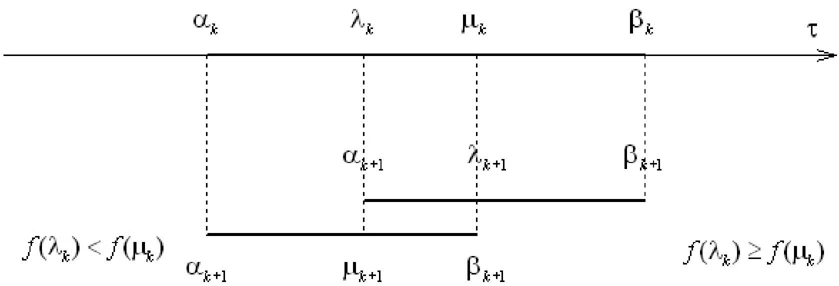
то для уменьшения числа новых значений $f(\tau)$, необходимых для сокращения отрезка локализации, следует потребовать, чтобы для каждой новой итерации либо $\lambda_{k+1} = \mu_k$, либо $\mu_{k+1} = \lambda_k$.

Из следующего рисунка очевидно, что добиться этого можно, если

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

$f(\lambda_k) \geq f(\mu_k)$ или же $f(\lambda_k) \leq f(\mu_k)$;

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



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

$$1^{\circ}. \qquad \lambda_k = \alpha_k + \rho(\beta_k - \alpha_k).$$

$$2^{\circ}. \quad \mu_k = \alpha_k + (1-\rho)(\beta_k - \alpha_k) \quad \text{и}$$

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

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

$$4^{\circ}. \quad \alpha_{k+1} = \alpha_k \quad \text{и} \qquad 5^{\circ}. \quad \beta_{k+1} = \mu_k.$$

И учтем, наконец, что:

$$6^{\circ}. \quad \mu_{k+1} = \lambda_k.$$

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

$$\rho = \frac{3-\sqrt{5}}{2} \quad \text{и} \quad 1-\rho = \frac{\sqrt{5}-1}{2} .$$

Случай $f(\lambda_k) \geq f(\mu_k)$ приводит к такому же результату.

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

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