

Методы оптимизации

Лекция 1

12 февраля 2021 г.

ВВЕДЕНИЕ В ОПТИМИЗАЦИЮ

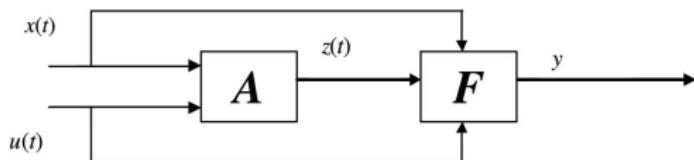
Введение

- Человеческая деятельность связана с постоянным (сознательным или бессознательным) решением оптимизационных задач.
- В технических приложениях под системой обычно понимают совокупность элементов, взаимодействующих в процессе функционирования устройства, выполняющего определённую задачу.
- При разработке или совершенствовании системы естественно стремление сделать её наилучшей в каком-либо смысле.
- Т. е. синтез такой системы, в процессе функционирования которой обеспечивалось бы экстремальное значение некоторого критерия — показателя качества или эффективности работы.
- Этот числовой показатель — *целевая функция*, которую требуется обратить в экстремум. В этом заключается задача оптимизации.

Введение

Имеются задачи двух типов. Задачи первого типа ставятся следующим образом.

Имеется система A с фиксированным входным вектором $x(t)$, управляемым входным вектором $u(t)$, выходным вектором $z(t)$, связанным с векторами $x(t)$ и $u(t)$ некоторым оператором, где t — время.

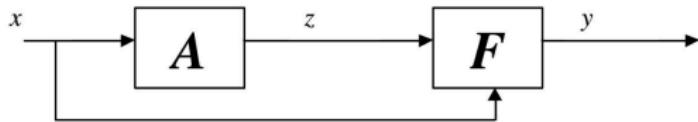


Определён некоторый целевой функционал $y = F(x(t), u(t), z(t))$. Требуется найти такую векторную функцию $u(t)$, которая обеспечивает экстремум y .

Такие задачи рассматриваются обычно в курсе теории оптимального управления. Они относятся к классу задач динамической оптимизации.

Введение

Имеется обширный класс задач исследования и проектирования систем, которые можно отнести ко второму типу. В этих задачах входные и выходные переменные не зависят от времени. При этом говорят, что рассматривается статическая модель системы.



В рамках этой модели определена функция $y = F(x, z)$.

Переменные z могут рассматриваться как функции $z(x)$, и тогда y является функцией только x .

Требуется выбрать значения x , соответствующие экстремуму функции $y = F(x)$.

Такая задача называется задачей статической оптимизации.

Введение

- Оптимизация должна быть выполнена при определённых ограничениях.
- Ограничения могут быть разнообразными и определяются многими факторами, например,
 - условиями реализуемости при проектировании системы,
 - условиями эксплуатации, гарантирующими надёжную работу объекта,
 - техническими заданиями на характеристики объекта.

Методы оптимизации

- В математике до середины прошлого века не существовало единой развитой теории оптимизации.
- Задача безусловной оптимизации функции нескольких переменных (Ферма, Ньютон, Лейбница), задачи вариационного исчисления (Эйлер, Лагранж, Якоби, Вейерштрасс), задачи оптимального управления (Понтрягин и др.), задачи на условный экстремум (Лагранж) и др.
- Однако найти с помощью условий экстремума явное решение задачи удаётся лишь в редких случаях для специально подобранных примеров.
- Методы решения задач оптимизации условно делятся на два класса: аналитические и численные (поисковые).
- К середине 60-х годов в рамках вычислительной математики сложилось самостоятельное направление, связанное с *численными методами оптимизации*.

Примеры

- Задача оценки параметров при обработке статистических данных.
- Задачи решения систем алгебраических уравнений.
- Задачи экспериментальной оптимизации.
- Задача поиска экстремума сложных функций при проектировании изделий.
- Задача идентификации, построении модели по имеющимся данным об объекте.
- и т. д.

Статистическая задача оценки параметров

Пусть z^1, \dots, z^m , $z^i \in \mathbb{R}^s$ — независимая выборка с плотностью распределения $p(z, x)$, $x \in \mathbb{R}^n$ — вектор неизвестных параметров. Требуется определить $x^* \in \mathbb{R}^n$.

Метод максимального правдоподобия — x^* выбирают так, чтобы максимизировать вероятность реализации z^1, \dots, z^m :

$$\max_x p(z^1, x) \dots p(z^m, x)$$

или

$$\max_{x \in \mathbb{R}^n} f_m(x), \text{ где } f_m(x) = -\sum_{i=1}^m \ln p(z^i, x).$$

Метод максимального правдоподобия для оценки среднего и дисперсии нормального распределения $N(\mu, \sigma^2)$ приводит к максимизации

$$f_m(x) = -\frac{m}{2} \ln (2\pi\sigma^2) - \sum_{i=1}^m \frac{(z^i - \mu)^2}{2\sigma^2},$$

Задача регрессии

Пусть

$$y_i = \Phi(u^i, x^*) + \xi_i, \quad i = 1, \dots, m,$$

y_i — результаты измерений, $y_i \in \mathbb{R}^1$, $u_i \in \mathbb{R}^s$ — управления, $\xi_i \in \mathbb{R}^1$ — случайные ошибки. Требуется по $y_i, u^i, i = 1, \dots, m$ оценить x^* .

Пусть ξ_i — случайные ошибки, независимы с плотностью вероятности $p(z)$.

Метод наименьших квадратов: $f_m(x) = -\sum_{i=1}^m (y_i - \Phi(u^i, x))^2$.

$$x^* = \arg \min f_m(x)$$

Для нормальных ошибок метод наименьших квадратов тождественен более общему статистическому методу оценивания параметров статистических распределений — методу максимального правдоподобия.

Задача анализа данных

Задачи описания объектов при отсутствии каких-либо предположений об их вероятностной природе относят к области анализа данных. В таких задачах требуется описать имеющиеся данные с помощью какой-либо модели так, чтобы рассогласование между моделью и объектом было достаточно мало.

Пусть $u = (u^1, \dots, u^S) \in \mathbb{R}^s$, $y = (y_1, \dots, y_m) \in \mathbb{R}^1$ — вход и выход объекта, $\Phi(u, x)$ — модель объекта, $x \in \mathbb{R}^n$ — подбираемые параметры.

$$\min_{x \in \mathbb{R}^n} f(x),$$
$$f(x) = \sum_{i=1}^m F(y_i - \Phi(u^i, x))$$

$F(z)$ — мера рассогласования объекта и модели (критерий идентичности).

Решение системы уравнений

Требуется найти решение системы n нелинейных уравнений

$$f_i(x) = 0, \quad i = 1, \dots, n,$$

$$x \in \mathbb{R}^n$$

Задача сводится к

$$x^* = \arg \min_{x \in \mathbb{R}^n} f(x),$$

$$f(x) = \sum_{i=1}^n F(z_i), \quad z_i = f_i(x)$$

$$F(z) = z^2, \quad F(z) = |z|, \quad F(z) = \max_{1 \leqslant i \leqslant m} |y_i - \Phi(u^i, x)|.$$

Оптимальное проектирование

Задачи, связанные с наилучшим выбором параметров отдельного изделия (например, оптимизация характеристик электрических машин, электронных приборов, строительных конструкций), технологического процесса (например, оптимизация химического реактора), предприятия или отрасли (например, задачи оптимального размещения производства, оптимизация трассы газопровода или электрической сети).

Поскольку характеристики объекта обычно описываются достаточно сложными зависимостями, их часто аппроксимируют полиномиальными выражениями.

$$\begin{aligned} \min f_0(x), \quad & x \in \mathbb{R}^n, \\ f_l(x) \leq 1, \quad & l = 1, \dots, m \\ 0 < a_i \leq x_i \leq b_i, \quad & i = 1, \dots, n, \end{aligned}$$

$$f_l(x) = \sum_{j=1}^{N_l} a_{jl} x_1^{b_{1jl}} x_2^{b_{2jl}} \dots x_n^{b_{njl}}$$

Оптимальное распределение ресурсов

Ряд задач поиска неисправностей, обнаружения цели, планирования эксперимента укладывается в следующую простую схему.

Имеется некоторый ресурс. Как наилучшим образом распределить его по n объектам, если эффективность использования на i -м объекте задаётся функцией $\varphi_i(x)$?

Математически:

$$\begin{aligned} & \min [\varphi_1(x_1) + \dots + \varphi_n(x_n)], \\ & x_1 + \dots + x_n = 1, \quad x_i \geq 0, i = 1, \dots, n. \end{aligned}$$

Задача планирования производства

Пусть n — число различных видов производимых товаров; m — число различных видов исходных материалов, используемых предприятием; x_1, x_2, \dots, x_n — количество произведённых товаров каждого вида; c_1, c_2, \dots, c_n — прибыль от продажи товаров каждого вида; a_{ij} — количество единиц i -го материала, необходимого для производства j -го товара; b_1, b_2, \dots, b_m — запасы исходных продуктов.

Цель задачи — получение максимальной прибыли от продажи товаров.

$$\begin{aligned}c_1x_1 + c_2x_2 + \dots + c_nx_n &\rightarrow \max \\a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &\leq b_1 \\a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &\leq b_2 \\&\dots\dots\dots \\a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &\leq b_m \\x_1 \geq 0, \dots, x_n \geq 0\end{aligned}$$

Поиск наилучшего приближения

Пусть имеется переопределённая система линейных уравнений

$$(a^i, x) = b_i, \quad i = 1, \dots, m, \quad x \in \mathbb{R}^n$$

Такая система может и не иметь решения (особенно при $m > n$), поэтому естественно ставить задачу о минимизации какой-либо нормы от невязки. В частности,

$$\min \sum_{i=1}^m |(a^i, x) - b_i|$$

$$\min \sum_{i=1}^m ((a^i, x) - b_i)^2$$

$$\min \max_{1 \leqslant i \leqslant m} |(a^i, x) - b_i|$$

являются соответственно задачами наилучшего в норме l_1 , наилучшего среднеквадратичного и наилучшего равномерного (или чебышёвского) приближения.

Будем рассматривать численные методы решения конечномерных задач условной оптимизации, называемых также *задачами нелинейного математического программирования*, постановка которых сводится к виду

$$\min_{x \in X} f(x)$$

$$X = \left\{ x \mid x \in \mathbb{R}^n, g_i(x) \geqslant 0, i = 1, \dots, m_1, \right. \\ \left. g_i(x) = 0, i = m_1 + 1, \dots, m_2 \right\},$$

где $f(x)$ — целевая функция и функции ограничений $g_i(x)$ — вещественные скалярные функции.

Классификация оптимизационных задач

Типы $f(x)$	Типы $g_i(x)$
Функция одной переменной	Ограничения отсутствуют
Функция многих переменных	Простые ограничения
Линейная функция	Линейные функции
Квадратичная функция	Квадратичные функции
Гладкая нелинейная функция	Выпуклые гладкие функции
Выпуклая функция	Гладкие нелинейные функции
Выпуклая гладкая функция	Выпуклые функции
Негладкая нелинейная функция	Негладкие нелинейные функции
Стохастическая функция	Непрерывные функции
Непрерывная функция	Стохастические функции
Разрывная функция	Нечёткие функции
Нечёткая функция	

Классификация оптимизационных задач

Сочетание конкретных видов целевых функций и ограничений порождает определённые классы задач оптимизации:

- при отсутствии ограничений — задачу *безусловной оптимизации*;
- при линейных ограничениях и целевой функции — задачу *линейного программирования*;
- при линейных ограничениях и квадратичной целевой функции — задачу *квадратичного программирования*;
- при выпуклости множества X и функции $f(x)$ — задачу *выпуклого программирования*;
- при негладкой целевой функции — задачу *недифференцируемой оптимизации*;

Классификация оптимизационных задач

Сочетание конкретных видов целевых функций и ограничений порождает определённые классы задач оптимизации:

- при нескольких целевых функциях — задачу *многокритериальной оптимизации*;
- при стохастических ограничениях или целевой функции — задачу *стохастического программирования*;
- при нечётких ограничениях или целевой функции — задачу *нечёткой оптимизации*.

- Курс целиком посвящён конечномерным задачам
 - ⇒ задачи оптимизации функции конечного числа непрерывно изменяющихся переменных.
- Будут рассматриваться статические задачи
 - ⇒ задачи, в которых значения целевой функции не зависят от времени.
- Не включены дискретные задачи оптимизации
 - ⇒ не рассматриваются задачи, в которых переменные принимают только дискретные значения, например, целочисленные.
- Для определённости будет рассматриваться задача минимизации.

1. Безусловная оптимизация

- Методы одномерной оптимизации
- Методы первого порядка
- Методы второго порядка
- Прямые методы (нулевого порядка)

2. Условная оптимизация

- Нелинейные задачи
- Линейные задачи

Рекомендуемая литература

1. Поляк Б. Т. Введение в оптимизацию, 3-е изд. — М.: Наука, 2019.
2. Сухарев А. Г., Тимохов А. В., Федоров В. В. Методы оптимизации, 3-е изд. — М.: Издательство Юрайт, 2019.
3. Рыков А. С. Системный анализ. Модели и методы принятия решений и поисковой оптимизации. — М.: МИСиС, 2009.
4. Базара М., Шетти К. Нелинейное программирование. Теория и алгоритмы. Пер. с англ. — М.: 1982.
5. Калиткин Н. Н., Альшина Е. А., Корякин П. В. Численные методы. М.: Издательский центр «Академия», 2013.
6. Акулич И. Л. Математическое программирование в примерах и задачах. — М.: Лань, 2011.
7. Смирнов А. П. Методы оптимизации. Алгоритмические основы задач оптимизации: курс лекций. — М.: МИСиС, 2014.
8. Федунец Н. И., Куприянов В. В. Теория принятия решений. — М.: МГГУ, 2005.

МЕТОДЫ ОДНОМЕРНОЙ МИНИМИЗАЦИИ

Методы одномерной минимизации

Задача одномерной оптимизации:

$$\min_{x \in [a,b]} f(x) \quad [a,b] \subseteq \mathbb{R}^1$$

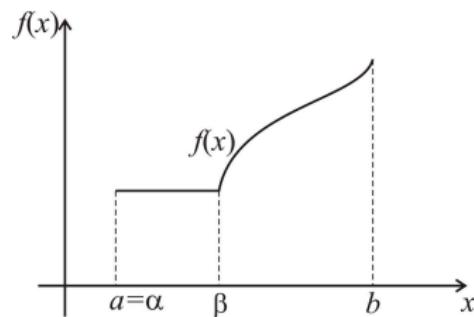
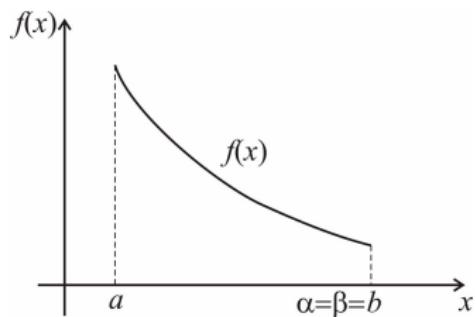
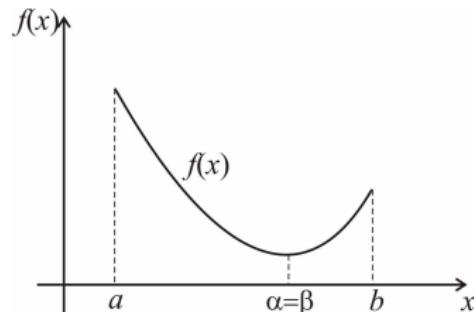
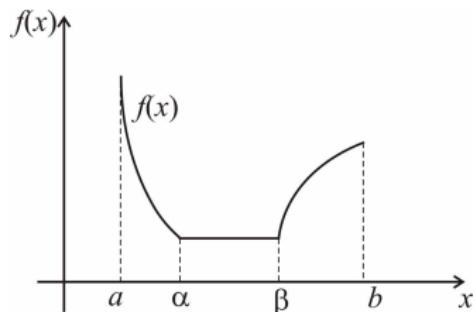
Важность умения решать одномерную задачу оптимизации связана с тем, что во многих методах многомерной оптимизации используются вспомогательные задачи одномерной оптимизации.

Определение.

Функция $f(x)$ называется *унимодальной на отрезке* $[a,b]$, если существуют $\alpha, \beta \in [a,b]$, $\alpha \leq \beta$, такие, что:

1. функция $f(x)$ строго монотонно убывает на отрезке $[a, \alpha]$
2. функция $f(x)$ строго монотонно возрастает на отрезке $[\beta, b]$
3. $f(x') = \min_{x \in [a,b]} f(x)$ при $x' \in [\alpha, \beta]$.

Примеры унимодальных функций



Унимодальные функции не обязаны быть непрерывными. От них требуется только вычисление значения функции.

Понятие о численных методах оптимизации

Любой численный метод (алгоритм) решения задачи оптимизации основан на точном или приближённом вычислении её характеристик (значений целевой функции, функций, задающих допустимое множество, а также их производных).

На основании полученной информации строится приближение к решению задачи — искомой точке минимума x^* или, если такая точка не единственна, — к множеству точек минимума.

Иногда, если только это и требуется, строится приближение к минимальному значению целевой функции $f^* = \min_{x \in X} f(x)$.

Алгоритмы, использующие лишь информацию о значениях минимизируемой функции, называются *алгоритмами нулевого порядка*; алгоритмы, использующие также информацию о значениях первых производных, — *алгоритмами первого порядка*; алгоритмы, использующие, кроме того, информацию о вторых производных, — *алгоритмами второго порядка*.

Типы алгоритмов оптимизации

Работа алгоритма состоит из двух этапов.

На первом этапе вычисляются предусмотренные алгоритмом характеристики задачи. На втором этапе по полученной информации о свойствах функции строится приближение к решению.

Поэтому чтобы задать алгоритм достаточно указать способ выбора точек вычисления (при условии, что уже решён вопрос о том, какие именно характеристики функции следует вычислять).

Если все точки выбираются одновременно до начала вычислений (т. е. положение точек вычисления не зависит от минимизируемой функции), то алгоритм минимизации называется *пассивным*.

Необходимость применения пассивных алгоритмов возникает, например, в связи с использованием параллельных и распределённых вычислений, в связи с условиями постановки и проведения физических экспериментов, результатом которых являются значения минимизируемой функции, и т. д.

Типы алгоритмов оптимизации

Однако для решения большинства задач точки вычисления выбираются поочерёдно, т. е. точка x^{i+1} выбирается, когда уже выбраны точки предыдущих вычислений x^1, \dots, x^i и в каждой из них произведены предусмотренные алгоритмом вычисления функции. Такие алгоритмы называются *последовательными*.

В практике обычно используются наиболее простые виды зависимости, например выбор точки очередного вычисления зависит лишь от точки предыдущего вычисления и полученного результата. Такие методы называются *одношаговыми*.

Методы, в которых новое приближение зависит от s предыдущих:

$$x^{k+1} = \varphi_k(x^k, \dots, x^{k-s+1}),$$

называются *s-шаговыми*, или общим словом *многошаговыми* (при $s > 1$).

Метод равномерного поиска

Метод относится к пассивным стратегиям поиска точки экстремума.

Задаётся начальный интервал неопределённости $L = [a, b]$
и погрешность вычисления ε .

Количество вычислений функции определяется по формуле

$$N = \left\lceil \frac{b - a}{\varepsilon} \right\rceil.$$

Вычисления производятся в N равноотстоящих друг от друга точках

$$x_i = a + \frac{i(b - a)}{N}, \quad i = 0, \dots, N.$$

При этом отрезок делится на $N - 1$ равных интервалов.

Метод равномерного поиска

Путем сравнения величин $f(x_i)$, $i = \overline{0, N}$ находится номер точки k , в которой значение функции наименьшее $f(x_k) = \min_{0 \leq i \leq N} f(x_i)$, $x^* = x_k$, $f^* = f(x^*) = f(x_k)$.

Искомая точка x^* считается заключённой на отрезке $[x_{k-1}, x_{k+1}]$, который называют отрезком локализации минимума.

Центр этого отрезка принимается за аппроксимацию точки x^* .

Погрешность оценки точки минимума x^* не превышает ε , при этом параметр алгоритма $\varepsilon > 0$ может быть выбран сколь угодно малым.

Метод поразрядного поиска

Для уменьшения количества вычислений функции $f(x)$ разумно сначала определить отрезок, содержащий x^* , грубо, то есть с небольшой точностью, а затем искать его на этом отрезке с меньшим шагом h (обычно в 4 раза) дискретизации, повышая точность.

Кроме того, если оказывается, что $f(x_i) < f(x_{i+1})$, то отпадает необходимость вычислять $f(x)$ в точках x_{i+2}, x_{i+3} и т. д. Вычислительный процесс заканчивается при выполнении условия $h \leq \varepsilon$.

Метод поразрядного поиска

Приведём описание алгоритма метода поразрядного поиска

1. Выбрать начальный шаг $\Delta := (b - a)/4$. Положить $x_0 := a$.
Вычислить $f(x_0)$.
2. Положить $x_1 := x_0 + \Delta$. Вычислить $f(x_1)$.
3. Сравнить $f(x_0)$ и $f(x_1)$. Если $f(x_0) > f(x_1)$, то перейти к п. 4,
иначе — к п. 5.
4. Положить $x_0 := x_1$ и $f(x_0) := f(x_1)$. Проверить условие
 $a < x_0 < b$. Если $a < x_0 < b$, то перейти к п. 2, иначе — к п. 5.
5. Проверка на окончание поиска: если $|\Delta| \leq \varepsilon$, то вычисления
завершить, полагая $x^* \approx x_0$, $f^* \approx f(x_0)$, иначе — перейти к п. 6.
6. Изменение направления и шага поиска: положить $x_0 := x_1$,
 $f(x_0) := f(x_1)$, $\Delta := -\Delta/4$. Перейти к п. 2.

Сокращение интервала неопределённости

Свойство унимодальности функции используется в методах оптимизации для сокращения отрезка, на котором располагается минимум функции. Эти методы основаны на исключении отрезков, на которых минимума унимодальной функции гарантированно нет.

Сущность метода исключения отрезков заключается в следующем. Пусть $a < x_1 < x_2 < b$. Если оказалось, что $f(x_1) < f(x_2)$, то в силу унимодальности минимум не может лежать правее точки x_2 и можно от отрезка $[a, b]$ перейти к отрезку $[a, x_2]$, сократив интервал неопределённости, на котором располагается точка минимума. Если при тех же условиях $f(x_1) > f(x_2)$, то минимум не может лежать левее точки x_1 и можно перейти к отрезку $[x_1, b]$.

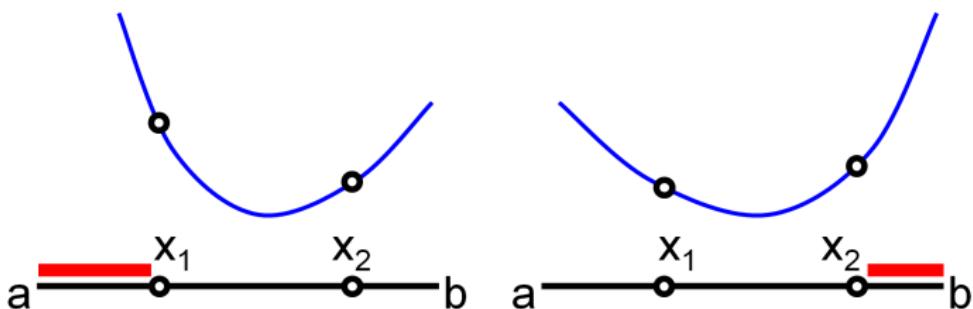
Описанную процедуру можно повторять необходимое число раз последовательно уменьшая отрезок, содержащий точку минимума.

В зависимости от способа выбора пробных точек получаются различные методы исключения отрезков.

Сокращение интервала неопределённости

Лемма.

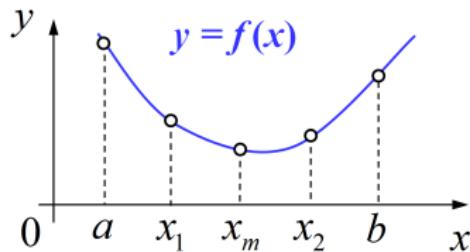
Пусть функция f унимодальна на X , $x_1, x_2 \in X$, $x_1 < x_2$. Тогда, если $f(x_1) \leq f(x_2)$, то $x^* \leq x_2$; если же $f(x_1) \geq f(x_2)$, то $x^* \geq x_1$.



Метод деления отрезка пополам

Метод относится к последовательным стратегиям и позволяет исключать из дальнейшего рассмотрения на каждой итерации в точности половину текущего интервала неопределённости.

Алгоритм уменьшения интервала основан на анализе величин функции в трёх точках, равномерно распределённых на текущем интервале (делящих его на четыре равные части). Иногда этот метод называют трёхточечным поиском на равных интервалах.



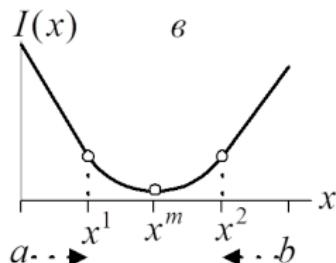
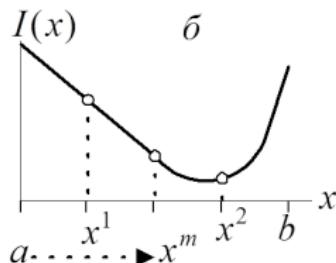
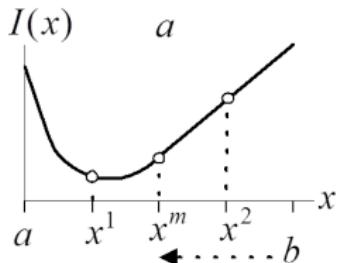
Процесс исключения отрезков продолжается до тех пор, пока длина текущего интервала неопределённости не станет меньше заданной величины.

Метод деления отрезка пополам

Пусть функция $f(x)$ унимодальна на отрезке $[a, b]$, тогда её минимум можно найти с помощью следующего алгоритма. Задаётся начальный интервал неопределённости $[a, b]$ и требуемая точность ε .

1. Выбрать $x_m := \frac{a+b}{2}$, $L := b - a$, вычислить $f(x_m)$.
2. $x_1 := a + L/4$, $x_2 := b - L/4$. Вычислить $f(x_1)$, $f(x_2)$.
3. Сравнить $f(x_1)$ и $f(x_m)$:
 - а) если $f(x_1) < f(x_m)$, то $b := x_m$, $x_m := x_1$. Перейти к п. 5;
 - б) если $f(x_1) \geq f(x_m)$, то перейти к п. 4.
4. Сравнить $f(x_m)$ и $f(x_2)$:
 - а) если $f(x_2) < f(x_m)$, то $a := x_m$, $x_m := x_2$. Перейти к п. 5;
 - б) если $f(x_2) \geq f(x_m)$, то $a := x_1$, $b := x_2$. Перейти к п. 5.
5. $L := b - a$. Проверить условие останова $L \leq \varepsilon$. При выполнении условия останова поиск прекратить, за минимум принять точку с минимальным значением функции $f(x)$, иначе перейти к п. 2.

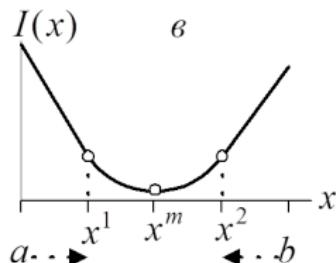
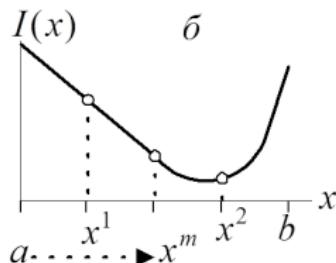
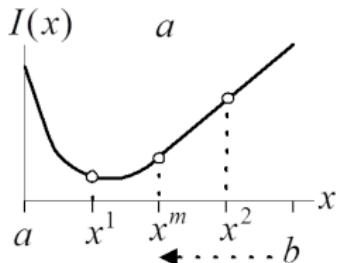
Метод деления отрезка пополам



Особенности метода:

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.

Метод деления отрезка пополам



Особенности метода:

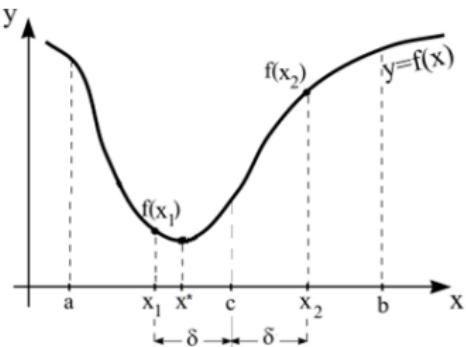
3. Если проведено n вычислений значений функции, то длина полученного интервала составляет $(1/2)^{n/2}$ величины исходного интервала.
4. В литературе показано, что из всех методов поиска на равных интервалах (двуточечный, трёхточечный, четырёхточечный и т. д.) трёхточечный поиск, или метод деления отрезка пополам, отличается наибольшей эффективностью.

Метод дихотомии

Метод относится к последовательным стратегиям. Задаётся начальный интервал неопределённости $[a, b]$ и требуемая точность ε .

Алгоритм опирается на анализ значений функции в двух точках. Для их нахождения текущий интервал неопределённости делится пополам и в обе стороны от середины откладывается по δ , где δ — малое положительное число $\delta \in (0, \varepsilon/2)$.

Процесс исключения отрезков продолжается до тех пор, пока длина текущего интервала неопределённости не станет меньше заданной величины.



Алгоритм.

1. Выбрать допустимую длину отрезка локализации ε и константу различимости $\delta < \varepsilon/2$.
2. Если $a - b < \varepsilon$, то поиск завершается, оптимальное решение содержится в интервале $[a, b]$.
3. Положить $x_1 := \frac{a+b}{2} - \delta$, $x_2 := \frac{a+b}{2} + \delta$. Вычислить $f(x_1)$, $f(x_2)$.
4. Если $f(x_1) < f(x_2)$, то $b := x_2$. В противном случае положить $a := x_1$. Перейти к п. 2.

В результате работы алгоритма на каждой итерации длина отрезка локализации сокращается почти в два раза.

Отметим, что для определения точки минимума надо рассмотреть все исследованные точки итогового отрезка локализации и выбрать ту из них, для которой значение функции минимально.

Определение.

Золотым сечением отрезка называется деление отрезка на две части так, чтобы отношение длины всего отрезка к длине большей части равнялось отношению длины большей части к длине меньшей части отрезка.

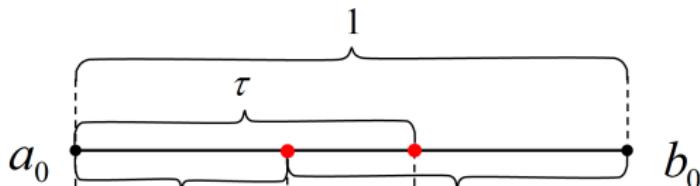
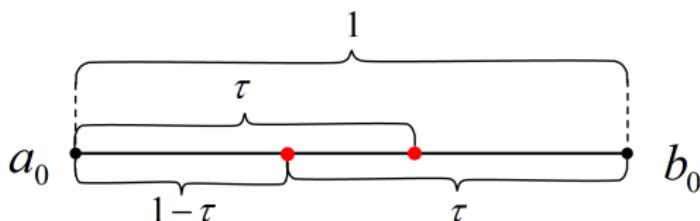
Золотое сечение замечательно тем, что на отрезке есть две точки золотого сечения (τ и $1 - \tau$), и если перейти от исходного отрезка к отрезку $[0, \tau]$, то точка $1 - \tau$ является точкой золотого сечения для этого отрезка. Если перейти к отрезку $[1 - \tau, 1]$, то точка τ будет точкой золотого сечения этого отрезка. Это свойство быть точкой золотого сечения двух отрезков используется в методе золотого сечения и вдвое сокращает число вычислений функции $f(x)$. На каждой итерации метода, кроме первой, производится всего одно вычисление значения функции $f(x)$, используя точку, в которой значение функции уже вычислено.

Метод золотого сечения

Пусть для удобства выбран исходный отрезок $[0, 1]$, тогда из определения золотого сечения имеем:

$$\frac{1}{\tau} = \frac{\tau}{1 - \tau}, \quad 1 - \tau = \tau^2, \quad \tau = \frac{-1 \pm \sqrt{5}}{2}, \quad \tau \approx 0,618,$$

где τ — точка золотого сечения.



Алгоритм.

Задаётся начальный интервал неопределённости $[a, b]$ и требуемая точность ε .

1. $x_1 := b - (b - a)\tau$, $x_2 := a + (b - a)\tau$. Вычислить $f(x_1)$, $f(x_2)$.

2. Сравнить значения $f(x_1)$ и $f(x_2)$:

а) если $f(x_1) > f(x_2)$, то $a := x_1$, $x_1 := x_2$, $x_2 := b + a - x_1$,
вычислить значение $f(x_2)$ и перейти к п. 3;

б) если $f(x_1) \leq f(x_2)$, то $b := x_2$, $x_2 := x_1$, $x_1 := b + a - x_2$,
вычислить значение $f(x_1)$ и перейти к п. 3.

3. $L = b - a$. Проверить условие останова $L \leq \varepsilon$. При выполнении условия останова поиск прекратить, за минимум принять точку с минимальным значением функции $f(x)$, иначе перейти к п. 2.

Метод Фибоначчи

Метод Фибоначчи построен американцем Дж. Кифером (J. Kiefer) в 1953 году.

Метод Фибоначчи обеспечивает максимальное гарантированное сокращение интервала неопределённости при заданном количестве вычислений функции. Эта стратегия опирается на числа Фибоначчи $\{F_k\}$, которые вычисляются рекуррентно по формуле:

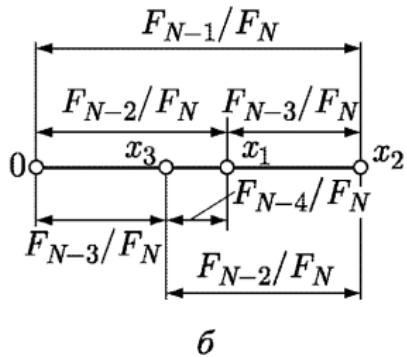
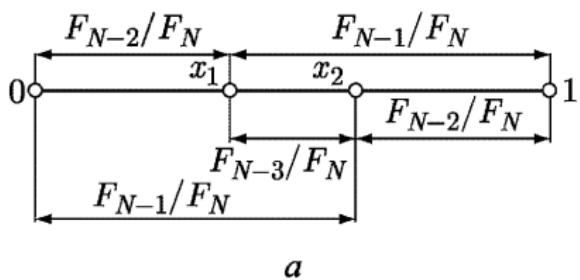
$$F_k = F_{k-1} + F_{k-2}, \quad F_0 = F_1 = 1.$$

Начальными членами последовательности $\{F_k\}$ будут числа 1, 1, 2, 3, 5, 8, 13, ... При фиксированном количестве N обращений к процедуре вычисления значений функции $f(x)$ поиск Фибоначчи состоит в просмотре точек, дробящих интервалы неопределённости в отношениях, заданных числами F_{N-1}, F_{N-2}, \dots

Метод Фибоначчи

Если принять длину исходного интервала за F_N , то длиной k -го интервала будет F_{N-k} , а его оцениваемые внутренние точки будут отстоять от левой границы на F_{N-k-2} и на F_{N-k-1} , причём в одной из них значение функции $f(x)$ всегда известно из предыдущего шага.

Взаимное расположение первых трёх точек вычислений по методу Фибоначчи показано на рисунке



Метод Фибоначчи

Взаимное расположение генерируемых точек поясняет следующая таблица

Число выч.	r_1	r_2	Длина отрезка
1	F_{N-2}/F_N	F_{N-1}/F_N	1
2	F_{N-3}/F_N	F_{N-2}/F_N	F_{N-1}/F_N
3	F_{N-4}/F_N	F_{N-3}/F_N	F_{N-2}/F_N
...
i	F_{N-i-1}/F_N	F_{N-i}/F_N	F_{N-i+1}/F_N
...
$N - 1$	$F_0/F_N = 1/F_N$	$F_1/F_N = 1/F_N$	$F_2/F_N = 2/F_N$
N	δ	$1/F_N$	$1/F_N, 1/F_N + \delta$

После $(N - 1)$ -го шага точка проведённого вычисления оказывается в середине отрезка локализации. Поэтому следующая точка выбирается на расстоянии δ от середины этого отрезка, где δ — заранее фиксированное малое положительное число.

Метод Фибоначчи

Алгоритм.

Задаётся начальный интервал неопределённости $[a, b]$. Выбирается допустимая конечная длина интервала неопределённости ε и константа различимости δ .

1. Определить число N из условия $F_N > \frac{b-a}{\varepsilon}$.
2. Положить $x_1 := a + \frac{F_{N-2}}{F_N}(b - a)$, $x_2 := a + \frac{F_{N-1}}{F_N}(b - a)$. Вычислить значения функции $f_1 := f(x_1)$, $f_2 := f(x_2)$, положить $k = 1$.
3. Если $f_1 > f_2$, то перейти к п. 4, в противном случае — к п. 5.
4. Положить $a := x_1$, $x_1 := x_2$, $f_1 := f_2$, вычислить $x_2 := a + \frac{F_{N-k-1}}{F_{N-k}}(b - a)$. Если $k = N - 2$, то перейти к п. 7, иначе вычислить $f_2 := f(x_2)$ и перейти к п. 6.
5. Положить $b := x_2$, $x_2 := x_1$, $f_2 := f_1$, вычислить $x_1 := a + \frac{F_{N-k-2}}{F_{N-k}}(b - a)$. Если $k = N - 2$, то перейти к п. 7, иначе вычислить $f_1 := f(x_1)$ и перейти к п. 6.

Метод Фибоначчи

Алгоритм.

6. Положить $k := k + 1$ и перейти к п. 3.

7. Положить $x_2 := x_1 + \delta$. Если $f(x_1) > f(x_2)$, то положить $a := x_1$.
В противном случае, положить $b := x_2$. Оптимальное решение
содержится в интервале $[a, b]$.

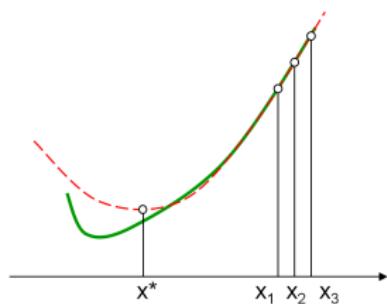
Результатом поиска Фибоначчи с N вычислениями значений
функции $f(x)$ является интервал неопределённости, составляющий
 $(1/F_N)$ -ю часть длины исходного интервала.

К недостаткам метода следует отнести обязательный выбор числа
вычислений N значений функции $f(x)$ до начала процесса
оптимизации.

Методы полиномиальной интерполяции

Метод парабол (квадратичной интерполяции).

Идея метода парабол связана с аппроксимацией минимизируемой функции $f(x)$ параболической моделью. Предполагается, что на ограниченном интервале можно аппроксимировать функцию $f(x)$ квадратичным полиномом, параметры которого однозначно определяются по вычислениям значений функции $f(x)$ в трёх точках. Затем аппроксимирующая функция используется для оценивания координаты точки истинного минимума функции $f(x)$.



Данный метод ориентирован на поиск минимума функций, близких к квадратичным.

Метод полиномиальной интерполяции

Алгоритм.

1. Задаётся x_1 , Δx , ε_1 , ε_2 . Выбрать $x_2 := x_1 + \Delta x$.
2. Вычислить значения функции $f(x_1)$, $f(x_2)$.
3. Если $f(x_1) > f(x_2)$, то $x_3 := x_1 + 2\Delta x$. Если $f(x_1) \leq f(x_2)$, то $x_3 := x_1 - \Delta x$.
4. Вычислить значение функции $f(x_3)$ и найти

$$F_m := \min \{f(x_1), f(x_2), f(x_3)\}, \quad x_m := \arg F_m.$$

5. По точкам x_1 , x_2 , x_3 вычислить \bar{x} .
6. Проверить выполнение условий: $|F_m - f(\bar{x})| \leq \varepsilon_1$, $|x_m - \bar{x}| \leq \varepsilon_2$.
При выполнении поиск прекратить, иначе перейти к п. 7.
7. Выбрать наилучшую точку (x_m или \bar{x}) и две точки по обе стороны от неё. Обозначить точки в естественном порядке (так, чтобы $x_1 < x_2 < x_3$) и перейти к п. 4.

Метод полиномиальной интерполяции

Опишем вычисление точки минимума параболы \bar{x} . Пусть заданы x_1 , x_2 , x_3 и значения функции в этих точках $f(x_1)$, $f(x_2)$, $f(x_3)$.

Уравнение параболы имеет вид:

$$g(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2),$$

где $f(x_1) = a_0$; $f(x_2) = f(x_1) + a_1(x_2 - x_1)$,

$$f(x_3) = f(x_1) + \frac{f(x_2) - f(x_1)}{(x_2 - x_1)(x_3 - x_1)} + a_2(x_3 - x_1)(x_3 - x_2);$$

$$a_1 = \frac{f(x_2) - f(x_1)}{x_2 - x_1};$$

$$a_2 = \frac{1}{x_3 - x_2} \left(\frac{f(x_3) - f(x_1)}{x_3 - x_1} - \frac{f(x_2) - f(x_1)}{x_2 - x_1} \right).$$

Точку минимума параболы можно определить из условия минимума $\frac{dg}{dx} = 0$, получим: $\bar{x} = (x_1 + x_2)/2 - a_1/(2a_2)$. Для вычисления \bar{x} нужно по значениям $f(x_1)$, $f(x_2)$, $f(x_3)$ вычислить значения a_1 , a_2 и подставить их в выражение для \bar{x} .

Вопросы?