Методы оптимизации

5 марта 2024 г.

Содержание

Литература.

1. *Васильев — Численные методы решения экстремальных задач*. Функциональный анализ, сжато, сложно, математично.

1 Основные понятия

Терминология.

1. Целевая функция — функция, у которой мы ищем экстремум.
2. Объект оптимизации — то, что мы оптимизируем.
3. Параметры оптимизации — параметры, от которых зависит целевая функция (аргументы).
4. Ограничения — равенства и неравенства, наложенные на параметры.

Пример 1. Спроектировать бак горючего в виде прямого кругового цилиндра с заданным объёмом V, на изготовление которого будет потрачено наименьшее количество листовой стали, то есть он должен иметь *наименьшую площадь поверхности.*

Решение. Параметры оптимизации: R и H. Целевая функция — функция площади поверхности, $S = 2\pi R^2 + 2\pi RH = 2\pi R(R+H) \to \min$. Ограничения: $\pi R^2 H = V$ $(V > 0)$; R, $H > 0$. По известной формуле

$$H = \frac{V}{\pi R^2}.$$

Тогда

$$S = 2\pi R\left(\frac{V}{\pi R^2} + R\right) = \frac{2\pi RV}{\pi R^2} + 2\pi R^2 = \\ = \frac{2V}{R} + 2\pi R^2 = \frac{V}{R} + \frac{V}{R} + 2\pi R^2 \geqslant 3\sqrt[3]{\frac{V}{R}\frac{V}{R}2\pi R^2} = 3\sqrt[3]{2\pi V^2}. = S_{\min}.$$

Здесь было использовано неравенство между средним арифметическим и средним геометрическим. Есть способ проще (через производные). Тогда $S = S_{\min}$ при $V/R = 2\pi R^2$, то есть $V = 2\pi R^3$ и $R = \sqrt[3]{\frac{V}{2\pi}}$, а

$$H = \frac{V}{\pi R^2} = \frac{V\sqrt[3]{4\pi^2}}{\pi\sqrt[3]{V^2}} = \sqrt[3]{\frac{4V}{\pi}}.$$

Отсюда

$$2R = 2\sqrt[3]{\frac{V}{2\pi}} = \sqrt[3]{\frac{8^4 V}{2\pi}} = H.$$

□

Линейное программирование придумали в России (см. Канторович).

Пример 2 (линейное программирование). Предприятие выпускает m наименований продукции, используя n видов ресурсов. Обозначим a_{ij} — затраты i-го вида ресурса на производство j-го вида продукции; x_j — планируемый объём выпуска продукции; d_j — цена единицы продукции j-го вида; а $(x_1, \ldots, x_n)$ — оптимальный план. Через b_i обозначим общий запас ресурса вида i.

Тогда целевая функция имеет вид

$$S = \sum_{j=1}^{n} d_j x_j \to \max,$$

а ограничения будут следующими: $a_j \leqslant x_j \leqslant A_j$; $a_{i1}x_1 + \ldots + a_{in}x_n \leqslant b_i$. Нишевые продукты тоже должны производиться. Аналогично, в транспортной задаче, задаче коммивояжёра и пр.

2 Интерполяция

1. Имеется набор значений функции где можно считать, что $x_0 < x_1 < \ldots < x_n$ (во всяком случае, все они попарно различны). Тогда существует единственный многочлен f степени $\deg f \leqslant n$, такой что $f(x_i) = y_i$, где $i = 0, \ldots, n$.

Запишем многочлен в виде

$$f = a_0 + a_1 x_+ \ldots + a_n x^n,$$
$$\begin{cases} a_0 + a_1 x_0 + \ldots + a_n x_0^n = y_0, \\ \ldots, \\ a_0 + a_1 x_n + \ldots + a_n x_n^n = y_n. \end{cases}$$

Здесь $(n+1)$ уравнение и столько же переменных. Определитель СЛАУ есть определитель Вандермонда, он не равен нулю, поскольку все точки x_i попарно различны. Поэтому система имеет и единственное решение. Эту систему можно просто решить. Можно поступить хитрее.

Второй способ.

По теореме Безу g_0 $x_1, \ldots, x_n$ — корни, откуда $g \vdots (x - x_1), \ldots, (x - x_n)$. $g_0(x) = c_0(x - x_1) \cdot \ldots \cdot (x - x_n)$, $c_0 \in \mathbb{R}$. $1 = g_0(x_0) = c_0(x_0 - x_1) \cdot \ldots \cdot (x_0 - x_n)$, поэтому

$$g_0(x) = \frac{(x - x_1) \ldots (x - x_n)}{(x_0 - x_1) \ldots (x_0 - x_n)} = \frac{\prod_{i=1}^n (x - x_i)}{\prod_{i=1}^n (x_0 - x_1)}.$$

Тогда

$$f(x) = y_0 g_0(x) + y_1 g_1(x) + \ldots + y_n g_n(x),$$
$$f(x_0) = y_0 g_0(x_0) + y_1 g_1(x_0) + \ldots + y_n g_n(x_0) \ldots,$$

где на второй строчке $g_0(x_0) = 1$, $g_1(x_0) = 0$, $g_n(x_0) = 0$.

Интерполяционный полином Лагранжа. Получили формулу

$$f(x) = \sum_{i=0}^{n} y_i \frac{\prod\limits_{j=0,\ldots,n; j \neq i} (x - x_j)}{\prod\limits_{j=0,\ldots,n; j \neq i} (x_i - x_j)}$$

Третий способ.

$$\begin{aligned} f_0(x) &= y_0, \\ (f_1 - f_0)(x_0) = 0 \implies (x - x_0) \mid f_1 - f_0 &\implies = q_1(x - x_0), \\ f_1 &= f_0 + q_1(x - x_0), \quad y_1 = f_1(x_1) = f_0(y_1) + q_1(x_1 - x_0), \\ q_1 &= \frac{y_1 - y_0}{x_1 - x_0}, \qquad f_1 = y_0 + \frac{y_1 - y_1}{x_1 - x_0}(x - x_0), \\ f_i - f_{i-1} &= q_i(x - x_0) \ldots (x - x_{i-1}), \\ y_i &= f_i(x_i) = f_{i-1}(x_i) + q_i(x_i - x_0) \ldots (x_i - x_{i-1}), \\ a_i &= \frac{y_i - f_{i-1}(x_i)}{(x_i - x_0) \ldots (x_i - x_{i-1})}. \end{aligned}$$

Интерполяция помогает и удобный инструмент, когда точек мало. Когда точек много, точная интерполяция сложна и бессмысленна.

2. Приближённая интерполяция. Когда точек много и точки известны приблизительно, используют приближённую интерполяцию. *Метод наименьших квадратов.*

Ищем многочлен f степени $\deg f = m < n$. Так называемая *«невязка»* $y_k - f(x_k)$. Как минимизировать невязки? Самый популярный способ — минимизировать сумму квадратов невязок. Тогда

$$\Phi(a_0, \ldots, a_m) = \sum_{k=0}^{n} (a_0 + a_1 x_k + \ldots + a_m x_k^m - y_k)^2 \to \min .$$

Максимума нет. Минимум означает, что $\frac{\partial \Phi}{\partial a_0} = 0$, …, $\frac{\partial \Phi}{\partial a_m} = 0$.

3 Одномерная оптимизация без использования производной

Часто случается, что производную сложно или невозможно найти, поскольку, например, сама имеет неизвестное уравнение.

Попробуем найти экстремумы без производной, даже не дифференцируемой.

Определение 1. Функция $f(x)\colon [a,b] \to \mathbb{R}$ называется *унимодальной*, если она непрерывна на этом отрезке и существуют такие числа $a \leqslant \alpha \leqslant \beta \leqslant b$, что

— функция $f(x)$ строго монотонно убывает на $[a, \alpha]$,

— $f(x) \equiv c$ для всех $x \in [\alpha, \beta]$,

— $f(x)$ строго монотонно возрастает на $[\beta, b]$.

Функция называется *строго унимодальной*, если $\alpha = \beta$.

Замечание 3.1. Выпуклая вниз непрерывная на $[a, b]$ функция унимодальна на этом отрезке.

Замечание 3.2. У унимодальной функции $\Omega_* = [\alpha, \beta]$ минимум $m_* = c$. Множество точек минимума.

Замечание 3.3. Есно, что если функция $f(x)$ унимодальна на $[a, b]$ и $[c, d] \subset [a, b]$, то $f(x)$ унимодальна на $[c, d]$.

1. Метод деления отрезка пополам. Пусть $f(x)$ унимодальна на $[a, b]$, где $b > a$. Выберем параметр $0 < \delta < b - a$. Выберем две точки

$$u_1 = \frac{a + b - \delta}{2}, \quad u_2 = \frac{a + b + \delta}{2}.$$

Вычислим $f(u_1)$, $f(u_2)$.

1 сл. $f(u_1) \leqslant f(u_2)$. Тогда полагаем $a_2 = a_1$,$b_2 = u_2$,

2 сл. $f(u_1) > f(u_2)$. Тогда полагаем $a_2 = u_1$, $b_2 = b_1$.

Тогда f унимодальна на $[a_2, b_2]$.

Вычисляем $f(u_{2k-1})$ и $f(u_{2k})$.

1 сл. $f(u_{2k-1}) \leqslant f(u_{2k})$. Тогда $a_{k+1} = a_k$, $b_{k+1} = u_{2k}$,

2 сл. $f(u_{2k-1}) > f(u_{2k})$. $a_{k+1} = u_{2k-1}$, $b_{k+1} = b_k$.

$l([a_{k+1}, b_{k+1}]) = \frac{b-a-\delta}{2^{k+1}} + \delta > \delta$.

4 Одномерная оптимизация с использованием дифференцирования

Пусть функция дифференцируема и мы знаем её производную. Однако по какой-то причине уравнение $f'(x) = 0$ мы решить не в состоянии.

4.1 Метод средней точки

Пусть функция непрерывно дифференцируемая на $[a, b]$ и имеет единственную стационарную точку (где $f'(x) = 0$). Пусть также $f'(a) < 0$, $f'(b) > 0$[1]. Тогда стационарная точка — точка минимума (очевидно).

Положим $a_0 = a$, $b_0 = b$. Возьмём сеередину отрезка $[a, b]$,

$$c_0 = \frac{a+b}{2}.$$

Вычислим $f'(c_0)$. Далее как обычно в методе отрезке пополам.

`Алгоритм 1 (Метод хорд).`

Замечание. Пусть $g(x)$ определена в точках x, x_1, x_2, $x_1 \neq x_2$ и $g(x_1) \cdot g(x_2) < 0$. Тогда прямая, проходящая через точки $(x_1, g(x_1))$ и $(x_2, g(x_2))$ имеет уравнение

$$y = \frac{g(x_2) - g(x_1)}{x_2 - x_1} x + \frac{g(x_1)x_2 - g(x_2)x_1}{x_2 - x_1},$$

а точка пересечения этой прямой с осью абсцисс

$$x = \frac{g(x_2)x_1 - g(x_1)x_2}{g(x_2) - g(x_1)} = x_1 - \frac{g(x_1)(x_2 - x_1)}{g(x_2) - g(x_1)} \in (x_1, x_2).$$

Пусть $f(x)$ непрерывно дифференцируема на $[a, b]$, имеет единственную стационарную точку на $[a, b]$; $f'(a) < 0$, $f'(b) > 0$.

Положим $a_0 = a$, $b_0 = b$.

$(k+1)$`-й шаг.` Рассматривается отрезок $[a_k, b_k]$, $f'(a_k) < 0$, $f'(b_k) > 0$.

Проведём прямую через точки $(a_k, f'(a_k))$, $(b_k, f'(b_k))$ и обозначим через

$$c_k = \frac{f'(b_k)a_k - f'(a_k)b_k}{f'(b_k) - f'(a_k)} \in (a_k, b_k).$$

Вычислим $f'(c_k)$.

(i). $f'(c_k) = 0$. Тогда c_k — стационарная точка.

(ii). $f'(c_k) > 0$. Тогда $a_{k+1} = a_k$, $b_{k+1} = c_k$.

(iii). $f'(c_k) < 0$. Тогда $a_{k+1} = c_k$, $b_{k+1} = b_k$.

`Условие остановки алгоритма.`

(a) Нашли стационарную точку (вряд ли),

(b) $l([a_k, b_k]) < \varepsilon > 0$, где ε задан заранееб

(c) Выполнено заданное количество итераций.

---

[1] Производная известна в том смысле, что мы можем прикинуть её знак в точках.

«Понятно, что метод сходится». — Киреева.

`Алгоритм 2 (метод Ньютона, метод касательных).` Пусть $f(x) \in C^2[a,b]$, имеет единственную стационарную точку и $f'(a) < 0$, $f'(b) > 0$.

Выберем начальную точку x_0.

(i) $f'(x_0) = 0$. Тогда x_0 стационарная.

(ii) $f'(x_0) \neq 0$. Тогда проведём касательную.

5 ???

`Алгоритм 3. Идея методов.` Пусть u_0 — начальная точка, а

$$u_{k+1} = \dots .$$

i) Если $f'(u_k) = 0$, то это стационарная точка,

ii) $f'(u_k) \neq 0$, то можно выбрать u_{k+1} так, что $f(u_{k+1}) < f(u_k)$.