МЕТОДЫ ОПТИМИЗАЦИИ (2024-2025 уч.г.)

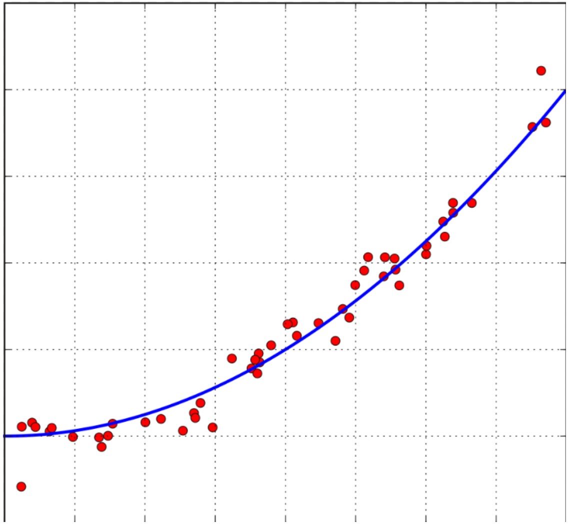
Манита Лариса Анатольевна — lmanita@yandex.ru, lmanita@hse.ru

Текущий контроль:

- Контрольная работа 1 (1 модуль) - 0,15
- Контрольная работа 2 (2 модуль) - 0,15
- Домашние задания (1-2 модули) - 0,1 (Численная оптимизация, линейное программирование, динамическое программирование)
- Коллоквиум - 0,2
- Экзамен - 0,4

1 Примеры оптимизационных задач

Регрессионный анализ



Имеем набор измерений $(x_i, y_i)_{i=1}^N$ и кривую $\varphi(x;\theta)$, где θ – набор неизвестных параметров. Необходимо подобрать θ таким образом, чтобы кривая наилучшим образом приближала экспериментальные данные.

$$\sum_{i=1}^{N} (y_i - \varphi(x_i;\theta))^2 \to \min$$

Задачи классификации

Имеем два набора векторов из $\mathbf{R}^n$, принадлежащих классам K_1 и K_2 (например, K_1 - “здоровый” человек, K_2 - с определенным заболеванием). Задача - построить функцию, которая бы “наилучшим” образом разделяла точки из классов K_1 и K_2.

Например:

Метод опорных векторов (Support Vector Machines, SVM)

$$\Downarrow$$

Конечномерная задача с ограничениями в виде неравенств

$$f\,(z) \to \min < \quad g\,(z) \geq 0$$

где z - параметры классификатора.

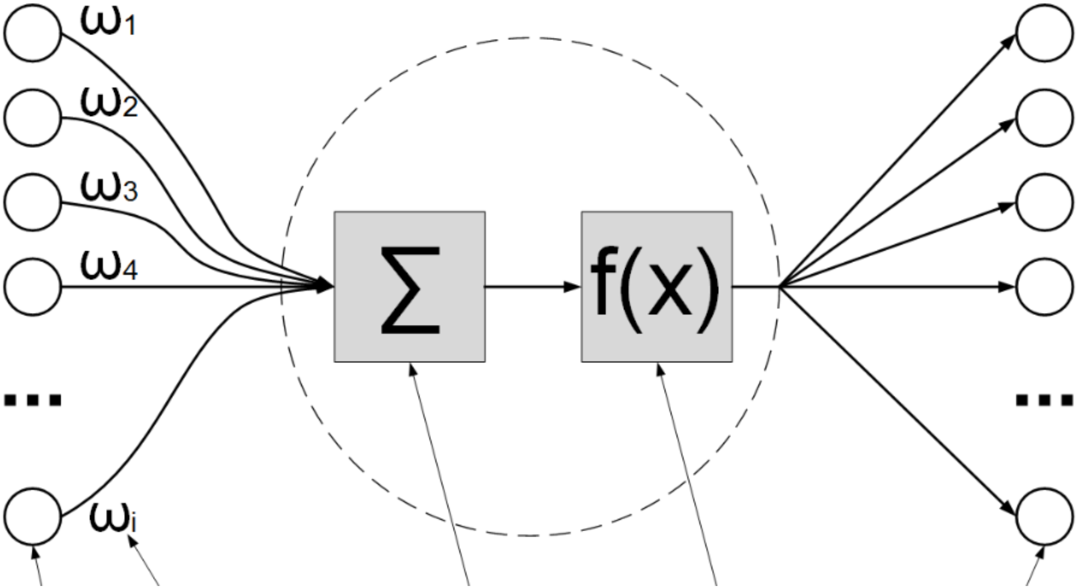
Динамическое программирование.

Принцип оптимальности Беллмана:

оптимальное поведение обладает тем свойством, что какими бы ни были первоначальное состояние системы и первоначальное решение, последующее решение должно определять оптимальное поведение относительно состояния, полученного в результате первоначального решения.

Применяется для решения задач
- распределения ресурсов;
- разработки правил управления запасами;
- составления планов текущего и капитального ремонтов оборудования и его замены;
- поиска кратчайших расстояний на транспортной сети;
- разработки алгоритмов обучения с подкреплением ("*жадный выбор по отношению к оптимальной Q-функции оптимален*")

Оптимизация в нейронных сетях



Задача о диете (задача линейного программирования)

> For a moderately active person, how much of each of a number of foods should be eaten on a daily basis so that the person's intake of nutrients will be at least equal to the recommended dietary allowances (RDAs), with the cost of the diet being minimal?

We consider the case when the only allowed foods can be found at McDonalds.

The McDonalds diet problem (Emil Gustavsson, Department of Mathematical Sciences, Chalmers University of Technology and Gothenburg University, Optimization)

Then we define the **parameters**

$$
\begin{aligned}
a_{ij} &= \text{Amount of nutrient } i \text{ in food } j,\ i \in \text{Nutrients},\ j \in \text{Foods}, \\
b_i &= \text{Recommended daily amount (RDA) of nutrient } i,\ i \in \text{Nutrients}, \\
c_j &= \text{Cost for food } j,\ j \in \text{Foods},
\end{aligned}
$$

and the **decision variables**

$$x_j = \text{Amount of food } j \text{ we should eat each day},\ j \in \text{Foods}.$$

The **model** of the diet optimization problem is then to

$$
\boxed{
\begin{aligned}
&\text{minimize} && \sum_{j \in \text{Foods}} c_j x_j, && \text{(2a)} \\
&\text{subject to} && \sum_{j \in \text{Foods}} a_{ij} x_j \ge b_i, \quad i \in \text{Nutrients}, && \text{(2b)} \\
& && x_j \ge 0, \quad j \in \text{Foods}. && \text{(2c)}
\end{aligned}
}
$$

Food	Calories	Carb	Protein	Vit A	Vit C	Calc	Iron	Cost
Big Mac	550 kcal	46g	25g	6%	2%	25%	25%	30kr
Cheeseburger	300 kcal	33g	15g	6%	2%	20%	15%	10kr
McChicken	360 kcal	40g	14g	0%	2%	10%	15%	35kr
McNuggets	280 kcal	18g	13g	0%	2%	2%	4%	40kr
Caesar Sallad	350 kcal	24g	23g	160%	35%	20%	10%	50kr
French Fries	380 kcal	48g	4g	0%	15%	2%	6%	20kr
Apple Pie	250 kcal	32g	2g	4%	25%	2%	6%	10kr
Coca-Cola	210 kcal	58g	0g	0%	0%	0%	0%	15kr
Milk	100 kcal	12g	8g	10%	4%	30%	8%	15kr
Orange Juice	150 kcal	30g	2g	0%	140%	2%	0%	15kr
RDA	2000 kcal	350g	55g	100%	100%	100%	100%	

Table 1: Given data for the diet problem

We define the **sets**

$$
\begin{aligned}
\text{Foods} &:= \{\text{Big Mac, Cheeseburger, McChicken, McNuggets, Caesar Sallad} \\
&\qquad \text{French Fried, Apple Pie, Coca-Cola, Milk, Orange Juice}\}, \\
\text{Nutrients} &:= \{\text{Calories, Carb, Protein, Vit A, Vit C, Calc, Iron.}\}
\end{aligned}
$$

(2a) We minimize the total cost, such that

(2b) we get enough of each nutrient, and such that

(2c) we don't sell anything to McDonalds.

Optimal solution

$$\boldsymbol{x} = \begin{pmatrix} x_{\text{Big Mac}} \\ x_{\text{Cheeseburger}} \\ x_{\text{McChicken}} \\ x_{\text{McNuggets}} \\ x_{\text{Caesar Sallad}} \\ x_{\text{French Fries}} \\ x_{\text{Apple Pie}} \\ x_{\text{Coca Cola}} \\ x_{\text{Milk}} \\ x_{\text{Orange Juice}} \end{pmatrix} = \begin{pmatrix} 0 \\ 7.48 \\ 0 \\ 0 \\ 0.27 \\ 0 \\ 3.03 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Total cost = 118.47 kr.

Total intake of calories = 3093.51 kcal.

integer ...

$$\boldsymbol{x} = \begin{pmatrix} x_{\text{Big Mac}} \\ x_{\text{Cheeseburger}} \\ x_{\text{McChicken}} \\ x_{\text{McNuggets}} \\ x_{\text{Caesar Sallad}} \\ x_{\text{French Fries}} \\ x_{\text{Apple Pie}} \\ x_{\text{Coca Cola}} \\ x_{\text{Milk}} \\ x_{\text{Orange Juice}} \end{pmatrix} = \begin{pmatrix} 0 \\ 7 \\ 0 \\ 0 \\ 1 \\ 0 \\ 3 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

= 150 kr.

= 3200 kcal.

Задача о найме на работу

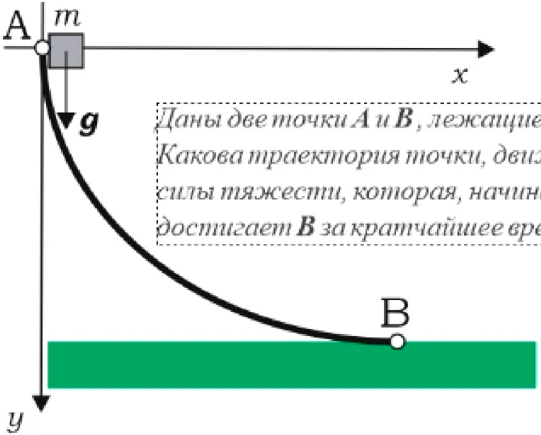
Example 1.1 (a staff planning problem) Consider a hospital ward which operates 24 hours a day. At different times of day, the staff requirement differs. Table 1.1 shows the demand for reserve wardens during six work shifts.

Table 1.1: Staff requirements at a hospital ward.

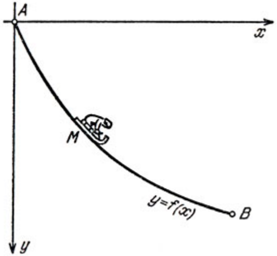
Shift	1	2	3	4	5	6
Hours	0–4	4–8	8–12	12–16	16–20	20–24
Demand	8	10	12	10	8	6

Each member of staff works in 8 hour shifts. The goal is to fulfill the demand with the least total number of reserve wardens. ■

Задача о брахистохроне (о наилучшей форме горки)



*Даны две точки **A** и **B**, лежащие в вертикальной плоскости. Какова траектория точки, движущейся только под действием силы тяжести, которая, начиная двигаться из **A**, достигает **B** за кратчайшее время?*



2 Постановка задачи оптимизации (конечномерная задача)

Задачи минимизации

$$f_0(x) \to \min, \quad x \in D, \tag{1}$$

или максимизации

$$f_0(x) \to \max, \quad x \in D \tag{2}$$

функции f_0 на множестве D.

$f_0(x)$ — скалярная функция, определенная (по крайней мере) на множестве $D \subset \mathbf{R}^n$: $f_0 : D \to \mathbf{R} \cup \{\pm\infty\}$

Если же требуется найти и точки минимума, и точки максимума, то будем писать

$$f_0(x) \to \mathrm{extr}, \quad x \in D. \tag{3}$$

Функция f_0 — целевая функция или показатель эффективности

$x = (x_1, \dots x_n)$ — вектор решений
множество D — множество допустимых решений

Определение 1 *Допустимое решение (допустимая точка) x^* является точкой локального минимума в задаче (3), если найдется ε - окрестность $U_\varepsilon(x^*)$ точки x^* $(\varepsilon > 0)$, такая, что для $\forall x \in D \bigcap U_\varepsilon(x^*)$ выполняется*

$$f_0(x) \geq f_0(x^*). \tag{4}$$

Здесь $U_\varepsilon(x^*) = \{x \in \mathbf{R}^n | \parallel x - x^* \parallel < \varepsilon\}$.

Если в определении 1 неравенство (4) выполнено строго для $x \neq x^*$, то x^* — строгий локальный минимум.

Определение 2 *Допустимая точка x^* является точкой глобального (или абсолютного) минимума в задаче (3), если для $\forall x \in D$ выполняется*

$$f_0(x) \geq f_0(x^*). \tag{5}$$

Если неравенство (5) выполнено строго для $x \neq x^*$, то x^* — строгий абсолютный минимум.

Определение 3 *Точки локального минимума и локального максимума функции f_0 называют точками локального экстремума.*

Обозначим $f_0^* = \inf_{x \in D} f_0(x)$.
Если $\exists x^* \in D \ : \quad f_0\left(x^*\right) = f_0^*$, то $f_0^* = \min_{x \in D} f_0(x)$.
Будем использовать обозначения

$$D^{\min} = argmin_{x \in D} f_0\left(x\right), \quad D^{\max} = argmax_{x \in D} f_0\left(x\right)$$

$$x^* \in \operatorname{absmin}\left(z\right)/\operatorname{absmax}\left(z\right) \Leftrightarrow x^* \in D^{\min}/x^* \in D^{\max}$$

Также будем писать

$$x^* \in locmin(z)/\mathrm{locmax}(z)$$

Из определения имеем

$$\mathrm{absmin} \Rightarrow locmin$$

$$\mathrm{absmax} \Rightarrow locmax$$

Замечание.

$$\begin{cases} f_0\,(x) \to \max \\ x \in D \end{cases} \sim \begin{cases} -f_0\,(x) \to \min \\ x \in D \end{cases}$$

А именно, допустимая точка x^* является решением задачи на максимум:

$$f_0\,(x) \to \max, \quad x \in D$$

тогда и только тогда, когда x^* — решение задачи на минимум

$$-f_0\,(x) \to \min, \quad x \in D$$

Что имеем в виду, когда говорим, что решили задачу, например, минимизации функции f_0 на множестве D?

мы нашли все точки локальных минимумов

$$\Downarrow$$

нашли значения целевой функции в каждой точке

локального минимума

$$\Downarrow$$

нашли глобальные минимумы или показали,

что глобального минимума нет.

ЛИБО

показали, что в задаче нет локальных минимумов.

3 Теоремы существования решений

Теорема 1 *(теорема Вейерштрасса)*
Пусть D — компакт в $\mathbf{R}^n$(т.е. замкнутое ограниченное множество), функция $f_0(x)$ непрерывна на D.
Тогда точка глобального минимума и точка глобального максимума функции f_0 на D существуют.

Теорема 2
Пусть D — замкнутое множество в $\mathbf{R}^n$,
$f_0(x)$ — непрерывная функция на D.
Если для некоторой точки $a \in \mathbf{R}^n$ множество

$$N(a) = \{x \in D \mid f_0(x) \leq f_0(a)\}$$

ограничено, то у функции f_0 на D существует точка глобального минимума.

Определение 4 *Функция $f_0(x)$ называется бесконечно растущей на множестве D, если на всякой последовательности $\{x_k\}_{k=1}^{\infty} \subset D$, такой, что либо*

$$\|x_k\| \longrightarrow +\infty, \quad k \to +\infty,$$

либо

$$x_k \to x^0 \in \overline{D} \backslash D,$$

выполняется

$$\lim_{k\to\infty} f_0(x_k) = +\infty. \tag{6}$$

D — открытое множество $\Rightarrow \overline{D} \backslash D$ — граничные точки множества D

D — замкнутое множество $\Rightarrow \overline{D} \backslash D = \emptyset$

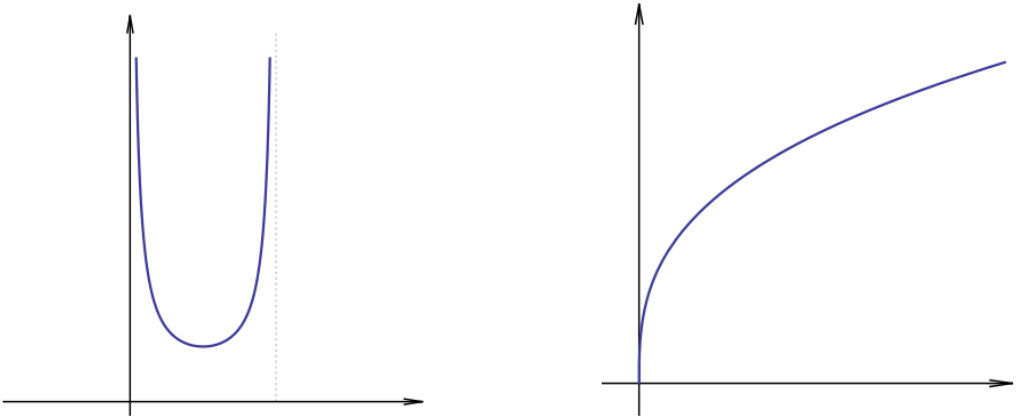


Рис. 2: примеры бесконечно растущих функций

Теорема 3 *(существование глобального минимума у бесконечно растущей функции)*
Пусть f_0 — бесконечно растущая непрерывная на D функция,
тогда на D существует глобальный минимум функции f_0.

***Пример* 1** *Функция $f_0(x_1, x_2) = x_1^2 + x_1 x_2 + x_2^2$ является бесконечно растущей в $\mathbf{R}^2$.*

Решение.
Действительно, имеем

$$x_1^2 + x_1 x_2 + x_2^2 \ \geq \ x_1^2 - \frac{x_1^2 + x_2^2}{2} + x_2^2 = \frac{1}{2}\left(x_1^2 + x^2\right) = \frac{1}{2}\left\|(x_1, x_2)\right\|^2 \ \to \ +\infty$$

при $\|(x_1, x_2)\| \to +\infty$.

4 Задачи без ограничений: $D = \mathbf{R}^n$

Одномерная задача без ограничений

$$f_0(x) \to extr, \quad x \in \mathbf{R}^1. \tag{7}$$

Теорема 4 *(Необходимые и достаточные условия локального минимума в одномерной задаче без ограничений)*
Пусть функция $f_0(x)$ s раз дифференцируема в точке x^* и

$$f_0'(x^*) = ... = f_0^{(m-1)}(x^*) = 0, \; f_0^{(m)}(x^*) \neq 0$$

для некоторого $m \in \mathbb{N}, m \leq s$. Если m нечетно, то x^* не является экстремумом; если m четно, то x^* является строгим локальным экстремумом, причем $x^* \in locmin(7)$, если $f_0^{(m)}(x^*) > 0$, и $x^* \in locmax(7)$, если $f_0^{(m)}(x^*) < 0$.

Пример 2 *Исследовать точку* $x^* = 0$ *для функций*

$$f_0(x) = \begin{cases} x^{s+\frac{1}{2}}, & x \geq 0, \\ (-x)^{s+\frac{1}{2}}, & x < 0, \end{cases}$$

и

$$f_1(x) = \begin{cases} x^{s+\frac{1}{2}}, & x \geq 0, \\ -(-x)^{s+\frac{1}{2}}, & x < 0. \end{cases}$$

Решение. Функции $f_0\,(x)$и $f_1\,(x)$ s раз непрерывно дифференцируемы в точке $x^* = 0$, и все их производные в нуле равны нулю

$$f_i'\,(0) = ... = f_i^{(s)}\,(0) = 0, \quad i = 1, 2.$$

При этом $x^* = 0$ является точкой глобального минимума функции f_0 и не является точкой экстремума функции f_1.

n – **мерная задача без ограничений** $(n \geq 2)$

$$\begin{cases} f_0(x) \to extr, \\ x \in \mathbf{R}^n, \quad n \geq 2. \end{cases} \tag{8}$$

Используем обозначения:
вектор частных производных (градиент) функции f_0 в точке x:

$$f_0'(x) = \nabla f_0 = \left(\frac{\partial f_0(x)}{\partial x_1}, ..., \frac{\partial f_0(x)}{\partial x_n}\right)^T.$$

матрица вторых производных функции f_0, вычисленная в точке x:

$$f_0''(x) = \nabla^2 f_0 = \left(\frac{\partial^2 f_0(x)}{\partial x_i \partial x_j}\right)_{i,j=1}^n.$$

Необходимое условие локального экстремума 1-го порядка в n – мерной задаче без ограничений

Теорема 5
Пусть $f_0(x)$ дифференцируема в окрестности точки x^*.
Если ***точка x^* — локальный экстремум*** *в задаче (8), то*

$$f_0'(x^*) = 0.$$

Доказательство. Пусть $x^* \in locmin$ (8).
Пусть h произвольный вектор из $\mathbf{R}^n$, α — число.
Тогда при всех достаточно малых α имеем: $f_0\left(x^*+\alpha h\right)-f_0\left(x^*\right) \geq 0$.
С другой стороны,

$$f_0\left(x^*+\alpha h\right)-f_0\left(x^*\right)=\left(f_0'\left(x^*\right), \alpha h\right)+o(\alpha)$$

Отсюда

$$\left(f_0'\left(x^*\right), \alpha h\right)+o(\alpha) \geq 0$$

Разделим обе части неравенства на $\alpha > 0$,
перейдем к пределу при $\alpha \to +0$:

$$\left(f_0'\left(x^*\right), h\right) \geq 0.$$

Пусть $f_0'\left(x^*\right) \neq 0 \Rightarrow$ положим $h = -f_0'\left(x^*\right)$. Тогда

$$\left(f_0'\left(x^*\right), -f_0'\left(x^*\right)\right) = -\left\|f_0'\left(x^*\right)\right\|^2 \geq 0.$$

Отсюда $f_0'(x^*) = 0$. Противоречие с предположением. $\square$

Точка x^*, удовлетворяющая условию $f_0'(x^*) = 0$,
называется ***стационарной*** точкой функции f_0.

Стационарная точка — решение задачи без ограничений?

Необходимы условия второго порядка

5 Знакоопределенные симметрические матрицы. Критерий Сильвестра

Пусть $A_{n\times n}$ — действительная симметрическая матрица

Определение 1
Будем говорить, что матрица A
- *неотрицательно определенная, если* $(Az, z) \geq 0 \quad \forall z \in \mathbf{R}^n$;
- *положительно определенная, если* $(Az, z) > 0 \quad \forall z \in \mathbf{R}^n$, $z \neq 0$;
- *неположительно определенная, если* $(Az, z) \leq 0 \quad \forall z \in \mathbf{R}^n$;
- *отрицательно определенная, если* $(Az, z) < 0 \quad \forall z \in \mathbf{R}^n$, $z \neq 0$;
- *знакопеременная или знаконеопределенная, если найдутся* z_1, $z_2 \in \mathbf{R}^n$, *такие, что* $(Az_1, z_1) > 0$, $(Az_2, z_2) < 0$.

$$(Az, z) = z^T Az$$

Угловые и главные миноры матрицы

Определение 2 *Угловым минором* $\triangle_{(1,2,\ldots,k)}$ *порядка* k $(1 \le k \le n)$ *матрицы* $A_{n\times n}$ *называется определитель подматрицы, которая получается из матрицы* A *вычеркиванием строк и столбцов с номерами* $k+1, \ldots, n$.

Угловой минор $\triangle_{(1,\ldots,n)}$ = определитель матрицы A.

Определение 3 *Главные миноры матрицы* A *— все возможные определители*

$$\triangle_{(k_1,\ldots,k_s)} = \det\begin{pmatrix} a_{k_1k_1} & a_{k_1k_2} & \ldots & a_{k_1k_s} \\ a_{k_2k_1} & a_{k_2k_2} & \ldots & a_{k_2k_s} \\ . & . & \ldots & . \\ a_{k_sk_1} & a_{k_sk_2} & \ldots & a_{k_sk_s} \end{pmatrix}$$

$1 \le k_1 < k_2 < \ldots < k_s \le n, \quad s = 1, \ldots, n.$

То есть, мы выписываем определитель подматрицы, которая состоит только из элементов строк и столбцов с номерами $k_1, k_2, \ldots, k_s$ (вычеркнули строки и столбцы с номерами $\{1, 2, \ldots, n\} \setminus \{k_1, k_2, \ldots, k_s\}$).

Угловые миноры являются главными

Можно использовать для главных миноров другие обозначения: $\delta_{i_1 i_2 \dots i_k}$ — удалили из A строки и столбцы с номерами $i_1, i_2, \dots, i_k$

$$(i_1, i_2, \dots, i_k \in \{1, \dots, n\}\,,\ 1 \le k \le n-1)\,.$$

Порядок $\delta_{i_1 i_2 \dots i_k}$ равен $n - k$.

Например, δ_2 — определитель подматрицы, которая получается из матрицы A вычеркиванием строки и столбца с номерами 2,
δ_{12} — вычеркиванием строк и столбцов с номерами 1 и 2,
δ_{124} — вычеркиванием строк и столбцов с номерами 1, 2 и 4, и т.д.
δ_0 — главный минор, который совпадает с определителем матрицы A.

Пример 3 *Вычислить угловые и главные миноры матрицы*

$$A = \begin{pmatrix} 4 & 2 & 3 \\ 2 & 1 & 0 \\ 3 & 0 & 2 \end{pmatrix}.$$

Решение. Угловые миноры:

$$\triangle_1 = 4, \quad \triangle_{1,2} = 0, \quad \triangle_{1,2,3} = -9.$$

Главные миноры:

$$\delta_1 = \triangle_{2,3} = 2, \quad \delta_2 = \triangle_{1,3} = -1, \quad \delta_3 = \triangle_{1,2} = 0, \quad \delta_{12} = \triangle_3 = 2,$$

$$\delta_{13} = \triangle_2 = 1, \quad \delta_{23} = \triangle_1 = 4, \quad \delta_0 = \triangle_{1,2,3} = -9.$$

Теорема 6 *Пусть $A_{n\times n}$ — действительная симметрическая матрица*
A положительно определена $\Leftrightarrow$ все ее ***угловые*** *миноры положительны.*
A отрицательно определена $\Leftrightarrow$ ее ***угловые*** *миноры имеют чередующиеся знаки, начиная с минуса $(-,+,-$ и т.д.).*
A неотрицательно определена $\Leftrightarrow$ все ее ***главные*** *миноры неотрицательны.*
A неположительно определена $\Leftrightarrow$ все ее ***главные*** *миноры* ***четного*** *порядка неотрицательны и все ее* ***главные*** *миноры* ***нечетного*** *порядка неположительны.*

Замечание 1 *Так как угловые миноры являются главными минорами, то матрица не может быть неотрицательно определенной, если среди угловых миноров есть отрицательные.*

Неотрицательность угловых миноров не является достаточным условием неотрицательной определенности матрицы.

***Пример* 4** *Проверить знакоопределенность матрицы*

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Решение. Угловые миноры неотрицательны: $0, \quad 0, \quad 1$. Следовательно, матрица A может быть неотрицательно неопределенной. Вычислим главные миноры:

$$0, -1, 0, 0, 0, -1, 1.$$

По критерию Сильвестра матрица A знакопеременная.

Знакоопределенность матрицы A связана со знаками ее собственых чисел.

Определение 4 *Число λ является собственным числом матрицы $A_{n\times n}$ с собственным вектором $v \in \mathbf{R}^n$, $v \neq 0$, если $Av = \lambda v$.*

Если $A_{n\times n}$ — действительная симметрическая матрица, то все собственные числа матрицы A действительны (спектр действительный).

A положительно определена, отрицательно определена, неотрицательно определена, неположительно определена, знаконеопределена тогда и только тогда, когда собственные значения λ_j матрицы A соответственно все положительны, все отрицательны, все неотрицательны, все неположительны, имеют различные знаки.

Необходимое условие оптимальности 2-го порядка в n – мерной задаче без ограничений

$$\begin{cases} f_0(x) \to extr, \\ x \in \mathbf{R}^n, \quad n \geq 2. \end{cases} \tag{9}$$

Теорема 7

Пусть функция f_0 дважды непрерывно дифференцируема в окрестности точки $x^* \in \mathbb{R}^n$.

Если $x^* \in locmin(9)$, то матрица $f_0''(x^*)$ неотрицательно определена.

Если $x^* \in locmax(9)$, то матрица $f_0''(x^*)$ неположительно определена.

Доказательство. Пусть $x^* - locmin \quad \Rightarrow f_0'(x^*) = 0$
и для любого $h \in \mathbf{R}^n$ и достаточно малых α имеем

$$0 \le f_0(x^* + \alpha h) - f_0(x^*) = \frac{1}{2}\left(f_0''(x^*)\,\alpha h, \alpha h\right) + o\left(\alpha^2\right).$$

Разделим обе части неравенства на α^2 и перейдем к пределу при $\alpha \to 0$, получим

$$\left(f_0''(x^*)\,h, h\right) \ge 0,$$

то есть, матрица $f_0''(x^*)$ неотрицательно определена. Для локального максимума рассуждения аналогичны. □

Замечание 2 *Если $f_0''(x^*)$ — знакопеременная матрица, то есть существуют $h^1, h^2 \in \mathbf{R}^n$, такие, что*

$$\left(f_0''(x^*)\,h^1, h^1\right) > 0,\ \left(f_0''(x^*)\,h^2, h^2\right) < 0,$$

то x^* не является локальным экстремумом.

Достаточное условие оптимальности в n – мерной задаче без ограничений

Теорема 8

Пусть функция f_0 дважды непрерывно дифференцируема в окрестности точки $x^* \in \mathbb{R}^n$.
Если $f_0'(x^*) = 0$, а матрица $f_0''(x^*)$ положительно определена, то x^* — строгий locmin (8).
Если $f_0'(x^*) = 0$, а матрица $f_0''(x^*)$ отрицательно определена, то x^* — строгий locmax (8).

Доказательство.
Для минимума!!
Допустим, что утверждение теоремы неверно, то есть, x^* не является строгим локальным минимумом.
Тогда $\exists \left\{x^k\right\}_{k=1}^{\infty}$: $x^k \neq x^*$, $x^k \to x^*$ и

$$f_0\left(x^k\right) \leq f_0\left(x^*\right). \tag{10}$$

Запишем $x^k = x^* + \alpha_k h^k$, где $\alpha_k = \left\|x^k - x^*\right\|$, $h^k = \left(x^k - x^*\right)/\alpha_k$.
$\left\|h^k\right\| = 1 \Rightarrow$ из ограниченной последовательности $\left\{h^k\right\}_{k=1}^{\infty}$ можно выбрать сходящуюся подпоследовательность $\left\{h^{k_s}\right\}_{s=1}^{\infty}$.
Считаем, что последовательность $\left\{h^k\right\}_{k=1}^{\infty}$ уже является сходящейся.
Обозначим

$$\lim_{k\to\infty} h^k = h^*.$$

Очевидно, $h^* \neq 0$. Так как $f_0'(x^*) = 0$, а по предположению верно (10), то

$$\begin{aligned} 0 &\geq f_0\left(x^k\right) - f_0\left(x^*\right) = \left(f_0'\left(x^*\right), \alpha h\right) + \frac{1}{2}\left(f_0''\left(x^*\right)\alpha_k h^k, \alpha_k h^k\right) + o\left(\alpha_k^2\right) \\ &= \frac{1}{2}\left(f_0''\left(x^*\right)\alpha_k h^k, \alpha_k h^k\right) + o\left(\alpha_k^2\right) \end{aligned}$$

Разделим на α_k^2 и перейдем к пределу при $k \to \infty$, получим

$$0 \geq \frac{1}{2}\left(f_0''\left(x^*\right)h^*, h^*\right).$$

Противоречие с условием теоремы. Следовательно, x^*— строгий локальный минимум. $\square$

Пример 5 *Найти точки экстремума функции*

$$f_0\left(x_1,x_2\right)=x_1^2+4x_1x_2+x_2^2.$$

Решение. Применим теорему о необходимых условиях первого порядка. Найдем стационарные точки функции f_0, т.е. точки, в которых производная (градиент) обращается в нуль:

$$f_0'\left(x_1,x_2\right)=\left(2x_1+4x_2,4x_1+2x_2\right);$$
$$f_0'\left(x_1,x_2\right)=0\Leftrightarrow x_1=x_2=0.$$

Итак, точка $(x_1^*,x_2^*)=(0,0)$ удовлетворяет необходимым условиям первого порядка. Проверим выполнение необходимых условий второго порядка и достаточных условий. Вычислим матрицу вторых производных:

$$f_0''\left(x_1,x_2\right)=\left(\begin{array}{cc}2&4\\4&2\end{array}\right).$$

Угловые миноры

$$\triangle_1=2>0,\ \triangle_{12}=4-16=-12<0.$$

По критерию Сильвестра матрица вторых производных является знакопеременной. Следовательно, не выполнены необходимые условия второго порядка. Таким образом, точка $(0,0)$ не является точкой экстремума.

***Пример* 6** *Найти точки экстремума функции*

$$f_0\left(x,y\right)=x^4+y^4-2x^2-2y^2.$$

Решение. Применим теорему о необходимых условиях первого порядка. Найдем стационарные точки функции $f_0\left(x,y\right)$:

$$f_0'\left(x,y\right)=\left(4x^3-4x,4y^3-4y\right);$$

$$f_0'\left(x,y\right)=0\Leftrightarrow\left\{\begin{array}{l}4x\left(x^2-1\right)=0\\4y\left(y^2-1\right)=0\end{array}\right.. \qquad (11)$$

Системе (11) удовлетворяют 9 точек:

$$\left(0,0\right),\left(0,\pm1\right),\left(1,0\right),\left(1,\pm1\right),\left(-1,\pm1\right).$$

Проверим выполнение необходимых условий второго порядка и достаточных условий. Вычислим матрицу вторых производных:

$$f_0''\left(x,y\right)=\left(\begin{array}{cc}12x^2-4&0\\0&12y^2-4\end{array}\right).$$

Проверим точку $(0,0)$:

$$f_0''(0,0) = \begin{pmatrix} -4 & 0 \\ 0 & -4 \end{pmatrix}.$$

По критерию Сильвестра матрица $f_0''(0,0)$ является отрицательно определенной, следовательно, $(0,0)$ — строгий локальный максимум.

Для точек $(0,\pm 1)$:

$$f_0''(0,\pm 1) = \begin{pmatrix} -4 & 0 \\ 0 & 8 \end{pmatrix}.$$

По критерию Сильвестра матрица $f_0''(0,\pm 1)$ является знакопеременной, следовательно, точки $(0,\pm 1)$ не являются локальными экстремумами.

Для точек $(\pm 1, 0)$:

$$f_0''(\pm 1,0) = \begin{pmatrix} 8 & 0 \\ 0 & -4 \end{pmatrix}.$$

По критерию Сильвестра матрица $f_0''(\pm 1,0)$ является знакопеременной, следовательно, $(\pm 1, 0)$ не являются локальными экстремумами.

Для точек $(\pm 1,\pm 1)$ и $(\pm 1,\mp 1)$:

$$f_0''(\pm 1,\pm 1) = f_0''(\pm 1,\mp 1) = \begin{pmatrix} 8 & 0 \\ 0 & 8 \end{pmatrix}.$$

По критерию Сильвестра матрицы $f_0''(\pm1,\pm1)$ и $f_0''(\pm1,\mp1)$ положительно определены, следовательно, $(\pm1,\pm1)$ и $(\pm1,\mp1)$ — строгие локальные минимумы.

Функция $f_0(x,y)$ является бесконечно растущей. Действительно, покажем, что $f_0(x,y)\to+\infty$, если $x^2+y^2\to+\infty$. Можем записать

$$f_0(x,y)=x^4+y^4-2x^2y^2+2x^2y^2-2x^2-2y^2=\left(x^2+y^2\right)^2-2\left(x^2+y^2\right)-2x^2y^2\geq$$

$$\geq\left(x^2+y^2\right)^2-\frac{5}{2}\left(x^2+y^2\right),$$

при оценке использовали следущий факт: $|ab|\leq\frac{1}{2}(a^2+b^2)$. Отсюда имеем $f_0(x,y)\to+\infty$ при $x^2+y^2\to+\infty$. Таким образом, у функции f_0 есть глобальный минимум. Выясним, какой из локальных минимумов является глобальным. Вычислим значения функции f_0 в точках $(\pm1,\pm1)$ и $(\pm1,\mp1)$: $f_0(\pm1,\pm1)=f_0(\pm1,\mp1)=-2$, следовательно, $(\pm1,\pm1)$ и $(\pm1,\mp1)$ — глобальные минимумы.

Локальный максимум будем проверять по определению. Рассмотрим приращение функции f_0 в нуле: $\triangle f_0=f_0(\triangle x,\triangle y)-f_0(0,0)=(\triangle x)^4+(\triangle y)^4-2(\triangle x)^2-2(\triangle y)^2$, при малых $\triangle x$, $\triangle y$ имеем: $\triangle f_0<0$, но при больших $\triangle x$, $\triangle y$ приращение $\triangle f_0>0$, следовательно, $(0,0)$ — локальный, но не глобальный максимум.

6 Выпуклые функции. Их значение для задачах оптимизации

Выпуклые множества

Определение 5 *Множество D называется* ***выпуклым****, если для всех $s, v \in D$ и любого $\alpha \in [0,1]$ выполнено $\alpha s + (1-\alpha)v \in D$.*

Геометрически выпуклость множества означает, что отрезок, соединяющий любые две точки из множества D, целиком лежит в множестве D.
Отрезок, соединяющий точки s и v, это множество точек вида

$$\{y \in \mathbf{R}^n : \; y = \alpha s + (1-\alpha)v, \, 0 \le \alpha \le 1\}$$

Пустое и одноточечное множества считаются выпуклыми по определению.

Пример 7 *Множество точек вида*

$$\Gamma = \{\, x \in \mathbf{R}^n |\, (c,x) = \gamma \} \,, \quad c \in \mathbf{R}^n, \quad c \neq 0, \quad \gamma \in \mathbf{R},$$

называемое гиперплоскостью в $\mathbf{R}^n$, *является выпуклым.*

Пример 8 *Замкнутые полупространства, порожденные гиперплоскостью* Γ,

$$\Gamma^+ = \{\, x \in \mathbf{R}^n |\, (c,x) \geq \gamma \} \,, \quad \Gamma^- = \{\, x \in \mathbf{R}^n |\, (c,x) \leq \gamma \} \,,$$

являются выпуклыми множествами в $\mathbf{R}^n$.

Определение 6 *Множества вида*

$$X = \left\{ x \in \mathbf{R}^n \middle| \left(c^i, x \right) \leq \gamma_i, \quad i = 1, \ldots, m \right\},$$

где $c^i = (c_1^i, \ldots, c_n^i), \quad \gamma_i \in \mathbf{R} \quad (i = 1, \ldots, m)$ *называются полиэдральными или полиэдрами.*

То есть полиэдр — это пересечение конечного числа полупространств. Из следующей теоремы будет следовать, что полиэдр — выпуклое множество.

Теорема 9
Пересечение (конечное или бесконечное) выпуклых множеств — выпуклое множество.

Доказательство. Пусть множества W_a выпуклы в $\mathbf{R}^n$ для всех $a \in A$. Обозначим

$$W = \bigcap_{a \in A} W_a.$$

Пусть w_1, $w_2 \in W$, тогда w_1, $w_2 \in W_a$ для всех $a \in A$. В силу выпуклости каждого W_a имеем

$$\alpha w_1 + (1 - \alpha)\, w_2 \in W_a, \quad \forall a \in A, \quad \forall \alpha \in [0, 1].$$

Отсюда следует, что $\alpha w_1 + (1 - \alpha)\, w_2 \in W$, то есть множество W выпукло. □

Объединение выпуклых множеств не обязательно выпукло.

Выпуклые функции

Пусть $f : D \to \mathbf{R}$, где D – выпуклое множество.

Определение 7 *Функция $f(x)$ называется* ***выпуклой*** *на множестве D, если $\forall\, s, v \in D$ и $\forall\, \alpha \in [0, 1]$ выполнено* ***неравенство Йенсена****:*

$$f(\alpha s + (1-\alpha)v) \le \alpha f(s) + (1-\alpha)f(v). \qquad (12)$$

Если (12) выполнено строго для всех $s, v \in D$, $s \ne v$, и для всех $0 < \alpha < 1$, то функция f называется строго выпуклой на множестве D.
Если неравенство (12) выполнено в противоположную сторону, т.е. $\ge$, то функция $f(x)$ называется вогнутой.
Если $f(x)$ — вогнутая функция, то $-f(x)$ — выпуклая функция.

***Пример* 9** *Функция $f : \mathbf{R}^n \to \mathbf{R}$, определенная формулой $f(x) = (c, x)$, $c \in \mathbf{R}^n$, является выпуклой.*

Дадим эквивалентное определение выпуклой функции через ее надграфик.

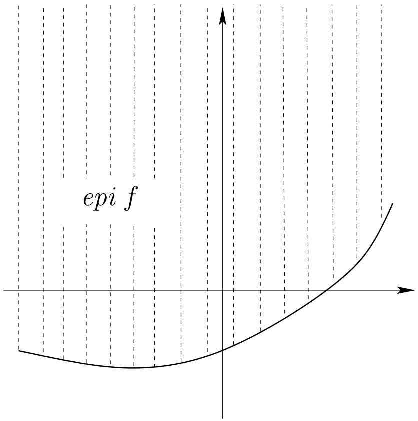
Определение 8 *Рассмотрим функцию* $f : D \to \mathbf{R}$, *где* $D \subset \mathbf{R}^n$.

Надграфиком функции f *назовем множество*

$$epi\, f = \left\{(x, c) \in \mathbf{R}^{n+1} \ : x \in D,\, c \in \mathbf{R},\, c \geq f\,(x)\right\}.$$

Теорема 10 *Пусть* D *— выпуклое множество,* $f \, : D \to \mathbf{R}$.

Функция f *выпукла тогда и только тогда,*
когда $epi\, f$ *— выпуклое множество в* $\mathbf{R}^{n+1}$.



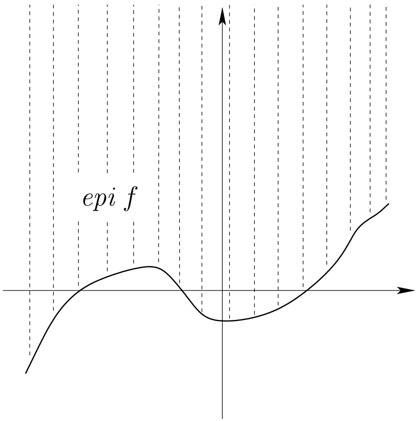


Рис. 1: слева надграфик выпуклой функции, справа надграфик функции, которая не является выпуклой

Критерий выпуклости дифференцируемой функции

Теорема 11
Пусть $f\ : D \to \mathbf{R}$ *— дифференцируемая функция,*
$D \subset \mathbf{R}^n$ *— выпуклое множество.*
Функция $f\,(x)$ *является выпуклой тогда и только тогда, когда*

$$f\,(x) - f\,(y) \geq \left(f'\,(y)\,, x - y\right)\ \forall\, x,\, y \in D.$$

Критерий выпуклости

дважды дифференцируемой функции

Теорема 12
Пусть $D \subset \mathbf{R}^n$ *— выпуклое множество,*
$f\ : D \to \mathbf{R}$ *дважды дифференцируемая функция.*
Функция f *выпукла* $\Longleftrightarrow$ $f''\,(x)$ *неотрицательно определена* $\forall x \in D$.

Доказательство.

$\Rightarrow$ Пусть f — выпуклая функция.

$y, x \in D$, обозначим $h = x - y$, тогда

$$y + \alpha h = y + \alpha(x - y) = \alpha x + (1 - \alpha) y \in D \quad (0 \le \alpha \le 1)$$

В силу критерия выпуклости для дифференцируемой функции (Т. 11)

$$\forall\, y,\, x \in D \quad f(y + \alpha h) - f(y) - (f'(y), \alpha h) \ge 0.$$

С другой стороны, применяя формулу Тейлора, имеем при $\alpha \to +0$:

$$f(y + \alpha h) - f(y) - (f'(y), \alpha h) = \frac{1}{2}(f''(y)\alpha h, \alpha h) + o\left(\alpha^2\right).$$

Отсюда

$$\frac{1}{2}(f''(y)\alpha h, \alpha h) + o\left(\alpha^2\right) \ge 0.$$

Разделим обе части неравенства на α^2 и рассмотрим предел при $\alpha \to +0$, получим $(f''(y)h, h) \ge 0$.

$\Leftarrow$ Пусть $f''(s)$ — неотрицательно определенная матрица $\forall\, s \in D$.

$y,\, x \in D,\ h = x - y$. Запишем для функции f формулу Тейлора с остаточным членом в форме Лагранжа

$$f(x) - f(y) = (f'(y), h) + \frac{1}{2}(f''(y + \alpha h)\, h, h),$$

где α – некоторое число, $0 < \alpha < 1$.

Точка $y + \alpha h \in D$ в силу выпуклости множества D.

Так как $f''(y + \alpha h)$ неотрицательно определенная матрица, то

$$f(x) - f(y) - (f'(y), x - y) \geq 0,$$

Применяем критерий выпуклости дифференцируемой функции

$\Rightarrow f(x)$ — выпуклая функция. $\square$

***Пример* 10** *Исследовать на выпуклость функцию*

$$f\left(x,y,z\right)=x^2+xy+y^2+2z.$$

Решение.

$$f''\left(x,y,z\right)=\begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Критерий Сильвестра $\Rightarrow$ угловые миноры:

$$\triangle_1=2>0,\quad \triangle_{12}=2\cdot 2-1\cdot 1=3>0,\quad \triangle_{123}=0.$$

Матрица $f''\left(x,y,z\right)$ может быть неотрицательно определенной, поэтому вычислим главные миноры:

$$\delta_1=0,\quad \delta_2=0,\quad \delta_3=3,\quad \delta_{12}=0,\quad \delta_{13}=2,\quad \delta_{23}=2,\quad \delta_0=0.$$

Так как все главные миноры неотрицательны, то $f''\left(x,y,z\right)$ неотрицательно определена $\Rightarrow$ функция $f\left(x,y,z\right)$ выпуклая.

Выпуклые функции и глобальные минимумы

Задача (z_c)

$$f_0(x) \to \min, \quad x \in D \subset \mathbf{R}^n \quad (\mathbf{z_c})$$

называется выпуклой задачей, если D – выпуклое множество, f_0 – выпуклая на D функция.

Теорема 13 *Пусть x^* — локальный минимум в выпуклой задаче (z_c). Тогда x^* — абсолютный минимум в (z_c).*

Доказательство.
$x^* \in locmin \Rightarrow$ найдется $U_\varepsilon(x^*)$: $f_0(x) \geq f_0(x^*)$ для всех $x \in D \bigcap U_\varepsilon(x^*)$. Возьмем любую точку $\tilde{x} \in D$. Соединим точки x^* и $\tilde{x}$ отрезком:
$\alpha\tilde{x} + (1-\alpha)x^* \in D$ для всех $0 \leq \alpha \leq 1$. Пусть $\alpha > 0$ и $\alpha \to 0$.
Тогда точки $x_\alpha = \alpha\tilde{x} + (1-\alpha)x^*$ попадут в $U_\varepsilon(x^*)$.
Для этих точек $f_0(x^*) \leq f_0(x_\alpha)$. Используем выпуклость $f_0 \quad \Rightarrow$

$$f_0(x^*) \quad \leq \quad f_0(\alpha\tilde{x} + (1-\alpha)x^*) \leq \alpha f_0(\tilde{x}) + (1-\alpha) f_0(x^*)$$

$$\Rightarrow \quad \alpha f_0(x^*) \leq \alpha f_0(\tilde{x}) \quad \Rightarrow \quad f_0(x^*) \leq f_0(\tilde{x}) \quad \Rightarrow \quad x^* \in \mathrm{absmin}$$

□

!!! Задача1. *Пусть* $f_0(x)$ — *выпуклая на* $\mathbf{R}^n$ *функция, дифференцируемая в точке* x^*. *Доказать, что если* $f_0'(x^*) = 0$, *то* x^* *является точкой глобального минимума* $f_0(x)$ *на* $\mathbf{R}^n$, *а следовательно, и на любом множестве* $D \subset \mathbf{R}^n$, *содержащем точку* x^*.

Свойства выпуклых функций

Сумма выпуклых функций

Теорема 14
Пусть $D \subset \mathbf{R}^n$ *— выпуклое множество, функции* $f_i : D \to \mathbf{R}$ *выпуклы для любого* $i = 1, \ldots, m$.
Пусть $c_i \geq 0 \ \ (i = 1, \ldots, m)$, *тогда функция*

$$f(x) = \sum_{i=1}^{m} c_i\, f_i(x)$$

является выпуклой на D.

Доказательство.
Пусть $s, v \in D$ и $\alpha \in [0,1]$. Тогда

$$f(\alpha s + (1-\alpha)v) = \sum_{i=1}^{m} c_i\, f_i\,(\alpha s + (1-\alpha)\,v) \leq_{(1)}$$

$$\sum_{i=1}^{m} c_i\, \left(\alpha f_i\,(s) + (1-\alpha)\, f_i\,(v)\right) = \alpha \sum_{i=1}^{m} c_i f_i\,(s) + (1-\alpha) \sum_{i=1}^{m} c_i f_i\,(v) =$$

$$= \alpha f\,(s) + (1-\alpha)\, f\,(v)\,.$$

Таким образом, для функции $f\,(x)$ выполнено неравенство Йенсена.

Неравенство (1) — выпуклость функций f_i и неотрицательность c_i. $\square$

Максимум из выпуклых функций

Теорема 15
Пусть на выпуклом множестве D определены выпуклые функции $f_a(x)$, $a \in A$ (A — некоторое множество индексов).

Тогда функция

$$f(x) = \sup_{a \in A} f_a(x)$$

является выпуклой на D.

Доказательство.

Первый способ.

$$epi \sup_{a \in A} f_a = \bigcap_{a \in A} epi\, f_a.$$

Так как $epi\, f_a$ — выпуклое множество для любого $a \in A$,
а пересечение выпуклых множеств — выпуклое множество,
то $epi \sup_{a \in A} f_a = epi\, f$ — выпуклое множество $\Longrightarrow f(x)$ — выпуклая функция.

Второй способ.

Пусть $s, v \in D$ и $\alpha \in [0,1]$. Тогда

$$f_a\left(\alpha s+(1-\alpha) v\right) \leq_{(1)} \alpha f_a\left(s\right)+(1-\alpha) f_a\left(v\right) \leq_{(2)} \alpha f\left(s\right)+(1-\alpha) f\left(v\right) \tag{13}$$

Неравенство (1) выполнено в силу выпуклости функций f_a,

неравенство (2) следует из определения функции $f\left(x\right)$.

Возьмем sup по a от левой и правой частей (13):

$$\sup_{a \in A} f_a\left(\alpha s+(1-\alpha) v\right) \leq \alpha f\left(s\right)+(1-\alpha) f\left(v\right),$$

$\Longrightarrow f\left(\alpha s+(1-\alpha) v\right) \leq \alpha f\left(s\right)+(1-\alpha) f\left(v\right).$ $\square$

Суперпозиция выпуклых функций

Теорема 16

Пусть на выпуклом множестве D_1 определена выпуклая функция $f_1(x)$, на выпуклом множестве D_2 $(f_1(D_1) \subset D_2)$ определена неубывающая выпуклая функция $f_2(y)$. Тогда функция $f(x) = f_2(f_1(x))$ является выпуклой на D_1.

Доказательство. Проверим выполнение неравенства Йенсена:

$$f\ (\alpha s + (1-\alpha)v) = f_2(f_1(\alpha s + (1-\alpha)v)) \leq_{(1)} f_2(\alpha f_1(s) + (1-\alpha) f_1(v)) \leq$$
$$\leq_{(2)} \alpha f_2(f_1(s)) + (1-\alpha) f_2(f_1(v)) = \alpha f(s) + (1-\alpha) f(v)\,.$$

Неравенство (1) — выпуклость $f_1(x)$

+ неубывание $f_2(y)$ $(y \leq \bar{y} \Rightarrow f_2(y) \leq f_2(\bar{y}))$,

неравенство (2) — выпуклость $f_2(x)$. □

***Пример* 11** *Заданы функции* $f(x) = x^2 - 1$ *и* $g(y) = |y|$.

Рассмотрим их композицию:

$$\varphi(x) = g(f(x)) = \left|x^2 - 1\right|.$$

Функции $f(x)$ *и* $g(y)$*выпуклые, однако* $g(f(x))$ *не выпуклая функция.*

***Пример* 12** *Заданы функции* $f(x_1, x_2) = x_1^2 + x_2^2$ *и* $g(y) = e^y$.

Рассмотрим их композицию:

$$\varphi(x) = g(f(x)) = e^{x_1^2 + x_2^2}.$$

Функции $f(x)$ *и* $g(y)$ *выпуклые,*

$g(y)$ *— возрастающая функция* $\Longrightarrow$

выполнены условия теоремы 16

$\Longrightarrow$ *композиция* $\varphi(x) = e^{x_1^2 + x_2^2}$ *— выпуклая функция.*

!!! Задача 2. Пусть функция $f : D \to \mathbf{R}$, где $D \subset \mathbf{R}^n$ — выпуклое множество, выпукла. Доказать обобщенное неравенство Йенсена: для любых

$$s_i \in D, \quad \alpha_i \geq 0 \quad (i = 1, \ldots, m), \quad \sum_{i=1}^{m} \alpha_i = 1 \quad (m = 1, 2, \ldots)$$

верно неравенство

$$f\left(\sum_{i=1}^{m} \alpha_i s_i\right) \leq \sum_{i=1}^{m} \alpha_i f\left(s_i\right). \tag{14}$$

Ю. Е. Нестеров Методы выпуклой оптимизации

“Задачи оптимизации совершенно естественно возникают в различных прикладных областях. Во многих жизненных ситуациях у нас появляется желание или необходимость организовать свою деятельность наилучшим из возможных способов. Это намерение, облеченное в математическую форму, приобретает вид той или иной оптимизационной задачи. В зависимости от конкретной области приложения это может быть задача оптимального управления или задача оптимального размещения, составление оптимальной диеты или задача оптимального раскроя. Однако уже следующий шаг — нахождение решения поставленной модельной задачи — совсем нетривиален. На первый взгляд, все выглядит просто: на рынке имеется огромное количество легкодоступных коммерческих программных оптимизационных пакетов, и любой пользователь может получить «решение» задачи простым нажатием на иконку на экране своего персонального компьютера. ***Вопрос заключается в том, что именно он получит в качестве решения и насколько можно доверять результату.*** Одна из целей данной книги — показать, что, несмотря на всю свою привлекательность, «решения» общих оптимизационных задач, получаемые таким образом, очень часто не соответствуют ожиданиям доверчивого пользователя.

Во многих практических приложениях процесс формализации и приведения реальной проблемы к какому-либо стандартному виду требует большого времени и усилий. Поэтому исследователь должен иметь ясное представление о свойствах модели, которую он строит. На этапе моделирования обычно применяются различные средства для аппроксимации реального явления, и при этом совершенно необходимо осознавать, к каким вычислительным последствиям приведет каждое из принимаемых решений. ***Очень часто приходится выбирать между «хорошей» модельной задачей, которую не удается решить, и «плохой» задачей, решение которой заведомо возможно. Какая из них лучше?***

Я попытаюсь убедить читателя в том, что для успешного применения оптимизационных формулировок задач необходимо иметь определенные сведения из теории оптимизации, которая помогает понять, чего можно и чего нельзя достигнуть при решении задачи оптимизации."

7 Введение в методы численной оптимизации.

1. **An Introduction to Continuous Optimization: Foundations and Fundamental Algorithms**, Niclas Andreasson, Anton Evgrafov, Michael Patriksson, http://www.math.chalmers.se/Math/Grundutb/CTH/tma947/0405/kompendium_sub.pdf
2. **Численные методы оптимизациию** или **Курс методов оптимизации.** Сухарев А.Г., Тимохов А.В., Федоров В.В.
3. **Численные методы оптимизации**, Измаилов А.Ф., Солодов М.В.
4. **Numerical Optimization,** Jorge Nocedal, Stephen J. Wright.

Методы одномерной оптимизации

$$f_0(x) \to \min, \quad x \in \mathbb{R}$$

- Метод перебора на равномерной сетке, Метод ломаных
- Метод деления пополам, метод парабол (для унимодальных функций)

Алгоритмы численной оптимизации (n- мерная задача без ограничений)

$$f_0(x) \to \min, \quad x \in \mathbb{R}^n$$

Определяем процедуру построения последовательности $x^0, x^1, \dots, x^k$:

$$x^{k+1} = x^k + \alpha_k h^k, \quad \alpha_k \in \mathbb{R}, \quad h^k \in \mathbb{R}^n \tag{15}$$

- Алгоритм определяется заданием x^0,
- правилами выбора векторов h^k и чисел α_k.
- Условия остановки:

$\left\|x^{k+1} - x^k\right\| \leq \epsilon$ или

$$\left\|f_0\left(x^{k+1}\right) - f_0\left(x^k\right)\right\| \leq \epsilon$$

или $\left\|f_0'\left(x^{k+1}\right)\right\| \leq \epsilon$ и др.

Методы спуска

Если последовательность в (15) $x^0, x^1, \ldots, x^k$ выбирается таким образом, что $f_0\left(x^0\right) > f_0\left(x^1\right) > \ldots > f_0\left(x^k\right)$, то алгоритм оптимизации (15) называется методом спуска.

Методы спуска: покоординатный, случайный поиск, градиентный, сопряженных градиентов, ...

Алгоритмы нулевого порядка
Алгоритмы первого порядка
Алгоритмы второго порядка

Определение 9 *Вектор $h \in \mathbf{R}^n$ определяет* ***направление убывания*** *функции f_0 в точке x ($f_0\left(x\right) < \infty$), если $\exists\, \delta > 0$: $f_0\left(x + \alpha h\right) < f_0\left(x\right)$ для всех $\alpha \in (0, \delta)$.*

Выбор длины шага α_k

$x^{k+1} = x^k + \alpha_k h^k$, $\alpha_k \in \mathbb{R}$, $h^k \in \mathbb{R}^n$

1. **Минимизация функции вдоль заданного направления**

$$f_0\left(x^k + \alpha_k h^k\right) = \min_{\alpha \geq 0} f_0\left(x^k + \alpha h^k\right)$$

На каждом шаге численно решается одномерная задача минимизации

2. **Априорный выбор длины шага**

- постоянный шаг;
- в ряде методов можно выбрать коэффициенты α с такими свойствами:

$$\alpha_k > 0,\ k = 0, 1, \ldots; \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty$$

например,

$$\alpha_k = \frac{c}{k+1}$$

3. **Дробление шага**

Пусть h^k - направление убывания в точке x^k. Выбираем $\beta > 0$ и $0 < \lambda < 1$. Положим $\alpha = \beta$, проверим

$$f_0\left(x^k + \alpha h^k\right) < f_0\left(x^k\right)?$$

Если выполнено, то положим $\alpha_k = \alpha$. Если нет, то дробим шаг: $\alpha = \lambda \cdot \beta$. Опять проверяем

$$f_0\left(x^k + \alpha h^k\right) < f_0\left(x^k\right)?$$

Процесс дробления продолжаем до тех пор, пока не получим

$$f_0\left(x^k + \alpha h^k\right) < f_0\left(x^k\right)$$

Почему за конечное число шагов мы получим это неравенство???

7.1 Методы нулевого порядка

Метод циклического покоординатного спуска

- Выбираем начальную точку x^0. Множество направлений - координатные направления.
- Цикл состоит из n шагов
- На 1-ом шаге проводят спуск по координате x_1. Значения всех переменных (кроме x_1) фиксированы:
$$x_2 = x_2^0,\ x_3 = x_3^0, \dots,\ x_n = x_n^0$$
Выбираем x_1^1 из условия минимизации функции f по переменной x_1:
$$f\left(x_1^1, x_2^0, \dots, x_n^0\right) = \min_{x_1 \in \mathbb{R}} f\left(x_1, x_2^0, \dots, x_n^0\right)$$
- На 2-ом шаге проводят спуск по координате x_2. Значения всех переменных (кроме x_2) фиксированы:
$$x_1 = x_1^1,\ x_3 = x_3^0, \dots,\ x_n = x_n^0$$

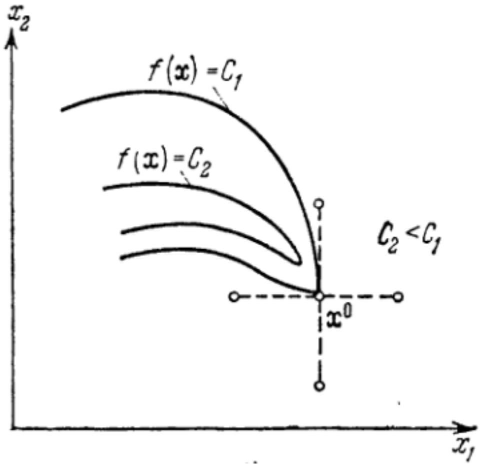
Выбираем x_2^1 из условия минимизации функции f по переменной x_2:

$$f\left(x_1^1, x_2^1, x_3^0 \ldots, x_n^0\right) = \min_{x_2 \in \mathbb{R}} f\left(x_1^1, x_2, x_3^0 \ldots, x_n^0\right)$$

- На n-ом шаге проводят спуск по координате x_2. Значения всех переменных (кроме x_n) фиксированы:
$$x_1 = x_1^1,\ x_2 = x_2^1, \ldots,\ x_{n-1} = x_{n-1}^1$$
Выбираем x_n^1 из условия минимизации функции f по переменной x_n:
$$f\left(x_1^1, x_2^1, x_3^1 \ldots, x_n^1\right) = \min_{x_n \in \mathbb{R}} f\left(x_1^1, x_2^1, \ldots, x_{n-1}^1 x_n\right)$$

- В результате получаем новую точку $x^1 = (x_1^1, x_2^1, \ldots, x_n^1)$.

- Повторяем цикл


x_2
$f(x) = C_1$
$f(x) = C_2$
$C_2 < C_1$
x^0
x_1

Случайный поиск

- Задается начальная точка x_0, число M - максимальное число неудачных испытаний на данной итерации, τ - минимальная величина шага, N - максимальное число итераций, параметр $\beta \in (0, 1)$.

- **На каждой итерации** k генерится случайный вектор $\xi = (\xi_1, \xi_2, \ldots, \xi_n)$ единичной длины (т.е. случайная точка (равномерно распределенная с.в.) на единичной сфере). Анализируется точка
$$y^k = x^k + \alpha_k \xi$$
Если в полученной точке значение не меньше, чем в x^k, то испытание считается неудачным и генерится новый случайный вектор. Если число неудачных испытаний достигает M, то дальнейший поиск продолжается с уменьшением шага ($\alpha_k \to \beta\alpha_k$), пока шаг не станет меньше τ.
Если в полученной точке y^k значение функции меньше, чем в x^k, то испытание считается удачным. Полагаем $x^{k+1} = y^k$ и переходим к следующей итерации.

Метод наилучшей пробы

- Задается начальная точка x_0, число M - число генерируемых точек на каждой итерации, τ - минимальная величина шага, N - максимальное число итераций, параметр $\beta \in (0, 1)$.

- **На каждой итерации** k генерится M случайных векторов $\xi^1 = (\xi_1^1, \xi_2^1, \ldots, \xi_n^1)$, ξ^2, ..., ξ^M единичной длины .
  Анализируются точки
  $$y^j = x^k + \alpha_k \xi^j$$
  Среди точек выбираем ту, в которой значение функции наименьшее: $y^s: \ f(y^s) \leq f(y^j)$ для всех y^s.

- Если в полученной точке y^s значение функции меньше, чем в x^k, то шаг считается удачным. Полагаем $x^{k+1} = y^k$ и переходим к следующей итерации. Если в полученной точке y^s значение не меньше, чем в x^k, то уменьшаем длину шага ($\alpha_k \to \beta\alpha_k$) (если $\alpha_k \geq \tau$) и генерим опять M случайных векторов. Если длина шага менее τ, про итерации завершаются. .

Метод статистического градиента

- генерится M случайных векторов $\xi^1 = (\xi_1^1, \xi_2^1, \ldots, \xi_n^1)$, ξ^2, ..., ξ^M единичной длины
- вычисляем разности
$$\triangle_j f = f\left(x^k + \delta \xi^j\right) - f\left(x^k\right)$$
- в качестве направления поиска используем вектор статистического антиградиента
$$d^k = -\frac{\sum_{j=1}^{M} \xi^j \triangle_j f}{\delta}$$

Метод сопряженных направлений

Пусть $A_{n\times n}$ - симметричная положительно определенная матрица.

Определение 10 *Векторы* h' *и* h'' *называются* сопряженными относительно матрицы A, *если они отличны от* 0 *и* $(Ah', h'') = 0$. *Векторы* $h^0, h^1, \ldots, h^k$ *называются* взаимно сопряженными относительно матрицы A, *если все они отличны от* 0 *и* $(Ah^i, h^j) = 0$, $i \neq j$, $\quad 0 \leq i, j \leq k$.

Лемма. Пусть векторы $h^0, h^1, \ldots, h^k$ взаимно сопряжены относительно матрицы A. Тогда они линейно независимы.

Доказательство.

Пусть лемма неверна. Тогда найдется ненулевой набор постоянных $c_0, c_1, \ldots, c_k$, такой что

$$c_0h^0 + c_1h^1 + \ldots + c_kh^k = 0$$

Выберем ненулевую постоянную c_j и рассмотрим

$$0 = \left(Ah^j, c_0h^0 + c_1h^1 + \ldots + c_kh^k\right) = \sum_{i=0}^{k} c_i \left(Ah^j, h^i\right) = c_j \left(Ah^j, h^j\right) \neq 0$$

Противоречие $\Rightarrow h^0, h^1, \ldots, h^k$ - линейно независимы.

Рассмотрим задачу

$$f_0(x) = \frac{1}{2}(Ax, x) + (b, x) \to \min, \quad x \in \mathbb{R}^n \tag{16}$$

$A_{n\times n}$ - симметричная положительно определенная матрица.

$$x^{k+1} = x^k + \alpha_k h^k, \quad \alpha_k \in \mathbb{R},\ h^k \in \mathbb{R}^n$$

$$f_0\left(x^k + \alpha_k h^k\right) = \min_{\alpha} f_0\left(x^k + \alpha h^k\right) \tag{17}$$

Теорема 17 *Если векторы $h^0, h^1, \ldots, h^{n-1}$ взаимно сопряжены относительно матрицы A, то для любой начальной точки x^0 алгоритм (17) находит точку минимума квадратичной функции f_0 не более чем за n шагов:*

$$f_0(x^n) = \min_{x\in\mathbb{R}^n} f_0(x)$$

Как строить сопряженные направления? Если есть набор $v^0, v^1, \dots, v^k$ линейно независимых векторов, то можно из них построить набор векторов $h^0, h^1, \dots, h^k$, взаимно сопряженных относительно матрицы A , причем подпространства, натянутые на них, совпадают.

Положим $h^0 = v^0$. Предположим, что мы уже имеем $v^0, v^1, \dots, v^j$ и $h^0, h^1, \dots, h^j$.

Построим

$$h^{j+1} = v^{j+1} + \sum_{i=0}^{j} c_i^{j+1} h^i$$

где c_i^{j+1} выбираем так, чтобы h^{j+1} был сопряжен с $h^0, \dots, h^j$, то есть, для всех $m = 0, 1, , \dots, j$

$$\left(h^{j+1}, Ah^m\right) = \left(v^{j+1} + \sum_{i=0}^{j} c_i^{j+1} h^i, Ah^m\right) = 0$$

$$\Rightarrow \left(v^{j+1}, Ah^i\right) + c_m^{j+1} \left(h^m, Ah^m\right) = 0$$

$$\Rightarrow c_m^{j+1} = -\frac{(v^{j+1}, Ah^i)}{(h^m, Ah^m)}$$