

# Условия оптимальности

**МЕТОДЫ ВЫПУКЛОЙ ОПТИМИЗАЦИИ** 

НЕДЕЛЯ 5

Даня Меркулов Пётр Остроухов



# Условия оптимальности. Ограничения равенства и неравенства. Условия ККТ.

### Семинар

Оптимизация для всех! ЦУ



# Условия оптимальности



$$f(x) \to \min_{x \in S}$$

Множество S обычно называется допустимым (бюджетным) множеством.

• Точка  $x^*$  является глобальным минимумом, если  $f(x^*) \leq f(x)$  для всех x.

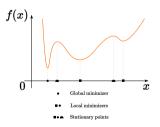


Рисунок 1. Illustration of different stationary (critical) points



$$f(x) \to \min_{x \in S}$$

Множество S обычно называется допустимым (бюджетным) множеством.

- Точка  $x^*$  является глобальным минимумом, если  $f(x^*) \leq f(x)$  для всех x.
- Точка  $x^*$  является локальным минимумом, если существует окрестность N точки  $x^*$ , такая что  $f(x^*) \leq f(x)$  для всех  $x \in N$ .

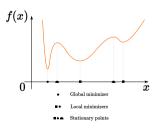


Рисунок 1. Illustration of different stationary (critical) points



$$f(x) \to \min_{x \in S}$$

Множество S обычно называется допустимым (бюджетным) множеством.

- Точка  $x^*$  является глобальным минимумом, если  $f(x^*) \leq f(x)$  для всех x.
- Точка  $x^*$  является локальным минимумом, если существует окрестность N точки  $x^*$ , такая что  $f(x^*) \leq f(x)$  для всех  $x \in N$ .
- Точка  $x^*$  является строгим локальным минимумом (также называется сильным локальным минимумом), если существует окрестность N точки  $x^*$ , такая что  $f(x^*) < f(x)$  для всех  $x \in N$  с  $x \neq x^*$ .

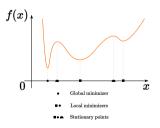


Рисунок 1. Illustration of different stationary (critical) points



$$f(x) \to \min_{x \in S}$$

Множество S обычно называется допустимым (бюджетным) множеством.

- Точка  $x^*$  является глобальным минимумом, если  $f(x^*) \leq f(x)$  для всех x.
- Точка  $x^*$  является локальным минимумом, если существует окрестность N точки  $x^*$ , такая что  $f(x^*) \leq f(x)$  для всех  $x \in N$ .
- Точка  $x^*$  является строгим локальным минимумом (также называется сильным локальным минимумом), если существует окрестность N точки  $x^*$ , такая что  $f(x^*) < f(x)$  для всех  $x \in N$  с  $x \neq x^*$ .
- Мы называем точку  $x^*$  стационарной (или критической), если  $\nabla f(x^*) = 0$ . Любой локальный минимум должен быть стационарной точкой.

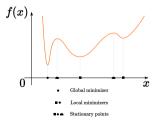


Рисунок 1. Illustration of different stationary (critical) points

### Безусловная оптимизация



Необходимое условие оптимальности первого порядка

Если  $x^*$  является локальным минимумом и f непрерывно дифференцируема в окрестности, то

$$\nabla f(x^*) = 0 \tag{1}$$

🥊 Достаточные условия оптимальности второго порядка

Предположим, что  $abla^2 f$  непрерывна в окрестности точки  $x^*$  и что

$$\nabla f(x^*) = 0 \quad \nabla^2 f(x^*) \succ 0. \tag{2}$$

Тогда  $x^{st}$  является строгим локальным минимумом функции f.



# Оптимизация с ограничениями-равенствами

# Оптимизация с ограничениями-равенствами



Рассмотрим простой, но практический случай ограничений-равенств:

$$\begin{split} f(x) &\to \min_{x \in \mathbb{R}^n} \\ \text{s.t.} \ h_i(x) &= 0, i = 1, \dots, p \end{split}$$

# Метод Лагранжа



Основная идея метода Лагранжа состоит в переходе от условной оптимизации к безусловной через увеличение размерности задачи:

$$L(x,\nu) = f(x) + \sum_{i=1}^p \nu_i h_i(x) = f(x) + \nu^T h(x) \rightarrow \min_{x \in \mathbb{R}^n, \nu \in \mathbb{R}^p}$$

## Метод Лагранжа



Основная идея метода Лагранжа состоит в переходе от условной оптимизации к безусловной через увеличение размерности задачи:

$$L(x,\nu) = f(x) + \sum_{i=1}^p \nu_i h_i(x) = f(x) + \nu^T h(x) \rightarrow \min_{x \in \mathbb{R}^n, \nu \in \mathbb{R}^p}$$

Необходимые условия:

$$\nabla_x L(x^*,\nu^*) = 0$$

$$\nabla_{\nu}L(x^*,\nu^*)=0$$

Достаточные условия:

$$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$$

$$\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^T y = 0$$



# Оптимизация с ограничениями-неравенствами

# Оптимизация с ограничениями-неравенствами



Рассмотрим простой, но практический случай ограничений-неравенств:

$$f(x) \to \min_{x \in \mathbb{R}}$$
 s.t.  $g(x) \le 0$ 

## Оптимизация с ограничениями-неравенствами



Рассмотрим простой, но практический случай ограничений-неравенств:

$$f(x) \to \min_{x \in \mathbb{R}^n}$$
 s.t.  $g(x) \le 0$ 

$$g(x) \leq 0$$
 неактивно.  $g(x^*) < 0$ : 
$$g(x^*) < 0$$
 
$$\nabla f(x^*) = 0$$
 
$$\nabla^2 f(x^*) > 0$$

$$g(x)\leq 0$$
 активно.  $g(x^*)=0$ : 
$$g(x^*)=0$$
 
$$-\nabla f(x^*)=\lambda \nabla g(x^*), \lambda>0$$
 
$$\langle y, \nabla^2_{xx} L(x^*,\lambda^*)y\rangle>0,$$
 
$$\forall y\neq 0\in \mathbb{R}^n: \nabla g(x^*)^\top y=0$$



# Условия Каруша-Куна-Таккера

# Общая формулировка



Общая задача математического программирования:

$$\begin{split} f_0(x) &\to \min_{x \in \mathbb{R}^n} \\ \text{s.t.} \ f_i(x) &\le 0, \ i=1,\dots,m \\ h_i(x) &= 0, \ i=1,\dots,p \end{split}$$

# Общая формулировка



Общая задача математического программирования:

$$\begin{split} f_0(x) &\to \min_{x \in \mathbb{R}^n} \\ \text{s.t.} \ f_i(x) &\le 0, \ i=1,\dots,m \\ h_i(x) &= 0, \ i=1,\dots,p \end{split}$$

Решение включает в себя построение функции Лагранжа:

$$L(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

# Необходимые условия ККТ



Пусть  $x^*$ ,  $(\lambda^*, \nu^*)$  является решением математической задачи программирования. Пусть также функции  $f_0, f_i, h_i$  дифференцируемы.

# Необходимые условия ККТ



Пусть  $x^*$ ,  $(\lambda^*, \nu^*)$  является решением математической задачи программирования. Пусть также функции  $f_0, f_i, h_i$  дифференцируемы.

$$\begin{split} (1)\nabla_x L(x^*,\lambda^*,\nu^*) &= 0 \\ (2)\nabla_\nu L(x^*,\lambda^*,\nu^*) &= 0 \end{split}$$

$$(3)\lambda_i^* \geq 0, i=1,\dots,m$$

$$(4)\lambda_i^*f_i(x^*)=0, i=1,\ldots,m$$

$$(5)f_i(x^*) \leq 0, i=1,\dots,m$$

## Некоторые условия регулярности



Эти условия необходимы для того, чтобы условия ККТ стали необходимыми. Некоторые из них даже превращают необходимые условия в достаточные. Например, условие Слейтера:

# Некоторые условия регулярности



Эти условия необходимы для того, чтобы условия ККТ стали необходимыми. Некоторые из них даже превращают необходимые условия в достаточные. Например, условие Слейтера:

Если для выпуклой задачи (т.е., предполагая минимизацию,  $f_0, f_i$  выпуклы и  $h_i$  аффинны), существует точка x такая что h(x)=0 и  $f_i(x)<0$  (существование строго допустимой точки), то условия ККТ становятся необходимыми и достаточными.

# Достаточные условия ККТ



Для гладких, нелинейных задач оптимизации, второе достаточное условие задается следующим образом. Решение  $x^*, \lambda^*, \nu^*$ , которое удовлетворяет условиям ККТ (выше), является локальным минимумом при ограничениях, если для функции Лагранжа

$$L(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

выполняются следующие условия:

$$\begin{split} &\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle > 0 \\ &\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top y = 0, \nabla f_0(x^*)^\top y \leq 0, \nabla f_j(x^*)^\top y = 0 \\ &i = 1, \dots, p \quad \forall j : f_j(x^*) = 0 \end{split}$$



# Задачи

## Задача 1



#### 1 Question

Функция  $f:E o\mathbb{R}$  определена как

$$f(x) = \ln \left( -Q(x) \right)$$

где 
$$E=\{x\in\mathbb{R}^n:Q(x)<0\}$$
 и

$$Q(x) = \frac{1}{2}x^{\intercal}Ax + b^{\intercal}x + c$$

$$\mathsf{c}\,A\in\mathbb{S}^n_{++},\,b\in\mathbb{R}^n,\,c\in\mathbb{R}.$$

Найдите точку максимума  $x^{st}$  функции f.

# Задача 2



i Question

Найдите явное решение следующей задачи.

$$f(x,y) = x + y \to \min$$
 s.t.  $x^2 + y^2 = 1$ 

где  $x,y\in\mathbb{R}.$ 

## Задача З



#### i Question

Найдите явное решение следующей задачи.

$$\langle c,x\rangle + \sum_{i=1}^n x_i \log x_i \to \min_{x\in\mathbb{R}^n}$$

$$\text{s.t. } \sum_{i=1}^n x_i = 1,$$

где 
$$x \in \mathbb{R}^n_{++}, c \neq 0$$
.

## Задача 4



1 Question

Пусть  $A \in \mathbb{S}^n_{++}, b>0$  покажите, что:

$$\det(X) \to \max_{X \in \mathbb{S}^n_{++}} \text{s.t.} \langle A, X \rangle \leq b$$

имеет единственное решение и найдите его.

### Задача 5



#### i Question

Даны  $y \in \{-1,1\}$ , и  $X \in \mathbb{R}^{n \times p}$ , задача об опорных векторах:

$$\begin{split} &\frac{1}{2}||w||_2^2 + C\sum_{i=1}^n \xi_i \rightarrow \min_{w,w_0,\xi_i}\\ \text{s.t. } \xi_i \geq 0, i = 1,\dots,n\\ &y_i(x_i^Tw + w_0) \geq 1 - \xi_i, i = 1,\dots,n \end{split}$$

найдите условие стационарности ККТ.



# Приложения

### Адверсариальные атаки



**Определение**: Адверсариальные атаки используются для обмана моделей DL путем добавления небольших возмущений к входным данным. Мы можем сформулировать это как задачу оптимизации с ограничениями, где целью является минимизация/максимизация функции потерь при сохранении возмущения в определенных пределах (ограничение нормы).

Метод FGSM (fast gradient-sign method) является самым простым таким методом, который генерирует adversarial examples путем применения небольшого возмущения в направлении градиента функции потерь. Формально:

$$x' = x + \varepsilon \cdot \mathrm{sgn}(\nabla_x L(x,y)), \text{s.t. } ||x - x'|| \leq \varepsilon$$

Таким образом, мы выполняем градиентный подъем на изображении (== максимизация потерь по отношению к этому изображению).



Рисунок 2. Иллюстрация

Вот код, попробуйте его сами!