

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

## Семинар

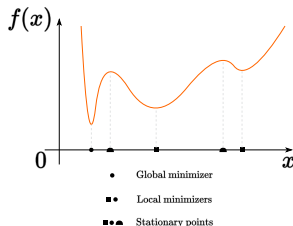
Оптимизация для всех! ЦУ

## Важные понятия

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

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.



# Безусловная оптимизация

## 💡 Необходимое условие оптимальности первого порядка

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

$$\nabla f(x^*) = 0 \quad (1)$$

## 💡 Достаточные условия оптимальности второго порядка

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

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

Тогда  $x^*$  является строгим локальным минимумом функции  $f$ .

# Оптимизация с ограничениями-равенствами

Рассмотрим простой, но практический случай ограничений-равенств:

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

# 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:

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

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

# Оптимизация с ограничениями-неравенствами

Рассмотрим простой, но практический случай ограничений-неравенств:

$$\begin{aligned} f(x) &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) &\leq 0 \end{aligned}$$

# Оптимизация с ограничениями-неравенствами

Рассмотрим простой, но практический случай ограничений-неравенств:

$$\begin{aligned} f(x) &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) &\leq 0 \end{aligned}$$

$g(x) \leq 0$  **неактивно**.  $g(x^*) < 0$ :

$$\begin{aligned} g(x^*) &< 0 \\ \nabla f(x^*) &= 0 \\ \nabla^2 f(x^*) &> 0 \end{aligned}$$

$g(x) \leq 0$  **активно**.  $g(x^*) = 0$ :

$$\begin{aligned} g(x^*) &= 0 \\ -\nabla f(x^*) &= \lambda \nabla g(x^*), \lambda > 0 \\ \langle y, \nabla_{xx}^2 L(x^*, \lambda^*) y \rangle &> 0, \\ \forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y &= 0 \end{aligned}$$



# Общая формулировка

Общая задача математического программирования:

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

# Общая формулировка

Общая задача математического программирования:

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

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 с нулевым двойственным разрывом (оптимальное значение для приоритетной задачи  $p^*$  равно оптимальному значению для двойственной задачи  $d^*$ ). Пусть также функции  $f_0, f_i, h_i$  дифференцируемы.

# Необходимые условия ККТ

Пусть  $x^*$ ,  $(\lambda^*, \nu^*)$  является решением математической задачи программирования с нулевым двойственным разрывом (оптимальное значение для приоритетной задачи  $p^*$  равно оптимальному значению для двойственной задачи  $d^*$ ). Пусть также функции  $f_0, f_i, h_i$  дифференцируемы.

$$(1) \nabla_x L(x^*, \lambda^*, \nu^*) = 0$$

$$(2) \nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$$

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

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

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

## 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:

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

$$\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top y = 0, \nabla f_0(x^*)^\top y \leq 0, \nabla f_j(x^*)^\top y = 0$$

$$i = 1, \dots, p \quad \forall j : f_j(x^*) = 0$$

# Задача 1

## i Question

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

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

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

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

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

Найдите точку максимума  $x^*$  функции  $f$ .



## Задача 2

### Question

Найдите явное решение следующей задачи.

$$\begin{aligned} f(x, y) = x + y &\rightarrow \min \\ \text{s.t. } x^2 + y^2 &= 1 \end{aligned}$$

где  $x, y \in \mathbb{R}$ .

## Задача 3

### i Question

Найдите явное решение следующей задачи.

$$\begin{aligned} \langle c, x \rangle + \sum_{i=1}^n x_i \log x_i &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } \sum_{i=1}^n x_i &= 1, \end{aligned}$$

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

## Задача 4

### Question

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

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

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:

$$\frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \xi_i \rightarrow \min_{w, w_0, \xi_i}$$

$$\text{s.t. } \xi_i \geq 0, i = 1, \dots, n$$

$$y_i(x_i^T w + w_0) \geq 1 - \xi_i, i = 1, \dots, n$$

найдите условие стационарности ККТ.

## Задача 6

### i Question

Покажите, что следующая задача оптимизации с ограничениями имеет единственное решение и найдите его.

$$\langle C^{-1}, X \rangle - \log \det(X) \rightarrow \min_{X \in \mathbb{S}_{++}^n} \text{ s.t. } a^T X a \leq 1$$

$$C \in \mathbb{S}_{++}^n, a \neq 0$$

Вы должны избежать явного обратного матрицы  $C$  в ответе.

## Задача 7 (БОНУС)

Для некоторых  $\Sigma, \Sigma_0 \in \mathbb{S}_{++}^n$  определите расхождение Кульбака-Лейблера между двумя гауссовыми распределениями как:

$$D(\Sigma, \Sigma_0) = \frac{1}{2}(\langle \Sigma_0^{-1}, \Sigma \rangle - \log \det(\Sigma_0^{-1} \Sigma) - n)$$

Теперь пусть  $H \in \mathbb{S}_{++}^n$  и  $y, x \in \mathbb{R}^n : \langle y, s \rangle > 0$

Мы хотим решить следующую задачу минимизации с ограничениями.

$$\min_{X \in \mathbb{S}_{++}^n} \{D(X^{-1}, H^{-1}) | Xy = s\}$$

Докажите, что она имеет единственное решение и оно равно:

$$(I_n - \frac{sy^T}{y^T s})H(I_n - \frac{ys^T}{y^T s}) + \frac{ss^T}{y^T s}$$

## Задача 8 (БОНУС)

### Question

Пусть  $e_1, \dots, e_n$  будет стандартным базисом в  $\mathbb{R}^n$ . Покажите, что:

$$\max_{X \in \mathbb{S}_{++}^n} \det(X) : \|Xe_i\| \leq 1 \forall i \in 1, \dots, n$$

имеет единственное решение  $I_n$ , и выведите неравенство Гильберта:

$$\det(X) \leq \prod_{i=1}^n \|Xe_i\| \forall X \in \mathbb{S}_{++}^n$$

## 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 (быстрого знака градиента) является самым простым таким методом, который генерирует adversarial examples путем применения небольшого возмущения в направлении градиента функции потерь. Формально:

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

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

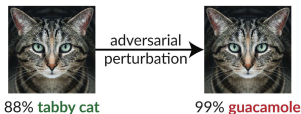


Рис. 2: Иллюстрация