

## Семинар 7: Условия Каруша–Куна–Таккера. Эквивалентные преобразования задач.

Теперь будем рассматривать задачи оптимизации не по всему пространству, а по некоторому его подмножеству  $Q \subseteq \text{Dom } f$ :

$$\min_{x \in Q} f(x). \quad (0.1)$$

Такие задачи называются *задачами с ограничениями* или *задачами условной оптимизации* и встречаются на практике, пожалуй, чаще чем безусловные задачи.

Напомним, что для безусловных гладких задач  $\min_{x \in \mathbb{R}^n} f(x)$ , у нас имеются условия существования экстремума первого порядка:

- **Необходимое условие.** Если  $x^*$  — точка локального минимума, то выполнено:  $\nabla f(x^*) = 0$ .
- **Достаточное условие.** Для *выпуклой* функции  $f(x)$ , если градиент в точке  $x^*$  равен нулю:  $\nabla f(x^*) = 0$ , то  $x^*$  — точка глобального минимума.

Нашей целью сейчас является формулирование аналогичных условий для задач оптимизации с ограничениями.

Множество  $Q$  мы будем задавать с помощью набора функциональных ограничений:

$$Q = \{x \in X \mid g_1(x) \leq 0, \dots, g_m(x) \leq 0, h_1(x) = 0, \dots, h_k(x) = 0\}, \quad (0.2)$$

здесь  $X$  — некоторое *простое открытое* множество, например, область определения всех функций:

$$X = \text{Dom } f(x) \cap \text{Dom } g_1(x) \cap \dots \cap \text{Dom } h_k(x).$$

Все функции предполагаются гладкими.

**Определение 0.1.** Точка  $x$  называется *допустимой*, если  $x \in Q$ .

**Определение 0.2.** Точка  $x^* \in Q$  называется *локальным минимумом*, если существует достаточно маленькая окрестность  $U$  точки  $x^*$  такая, что для всех  $x \in U \cap Q$  выполнено:  $f(x^*) \leq f(x)$ .

Аналогично определяется локальный максимум, *строгий* локальный минимум и *строгий* локальный максимум.

## 1 Условия Каруша–Куна–Таккера

Сформулируем утверждение, являющееся удобным инструментом для отыскания экстремумов условных задач оптимизации.

Определим для задачи (0.1), где множество  $Q$  задано функциональными ограничениями (0.2), *функцию Лагранжа*:

$$\mathcal{L}(x; \lambda, \mu) := f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^k \mu_j h_j(x).$$

**Теорема 1.1.** Пусть для этой задачи в точке  $x^* \in Q$  справедливо «условие регулярности».

Тогда если  $x^*$  есть локальный минимум, то существуют двойственные переменные  $\lambda^* \in \mathbb{R}_+^m$ ,  $\mu^* \in \mathbb{R}^k$  такие, что выполнено:

(a) *Стационарность.*

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^k \mu_j^* \nabla h_j(x^*) = 0.$$

(b) *Дополняющая нежесткость.*

$$\lambda_i^* g_i(x^*) = 0, \quad \text{для всех } i = 1, \dots, m.$$

Данная теорема является аналогом необходимого условия оптимальности первого порядка для задач оптимизации с ограничениями.

Оно утверждает, что для «хороших» задач (для которых справедливо некоторое «условие регулярности»), если точка  $x^*$  — локальный минимум, то мы можем выписать уравнения (*стационарности* и *дополняющей нежесткости*) которые обязательно разрешимы с некоторыми, неизвестными нам двойственными переменными.

**Замечание 1.1.** Обратите внимание, что двойственные переменные  $\lambda_i^*$ , отвечающие ограничениям-неравенствам обязательно неотрицательны:  $\lambda_i^* \geq 0$ .

Пречислим теперь самые популярные **Условия регулярности** в точке  $x$ . Если для нашей задачи выполнено *хотя бы одно* из этих условий в исследуемой точке  $x^*$  — задача является «хорошей» и для неё справедлива Теорема 1.1:

- Все ограничения  $g_1, \dots, g_m, h_1, \dots, h_k$  суть аффинные функции.
- В точке  $x$  градиенты всех *активных ограничений*<sup>1</sup> линейно независимы.
- Выполнено *условие Слейтера*:
  - (а) Задача является выпуклой (все  $f$  и  $g_i$  суть выпуклые функции,  $h_j$  — аффинные функции).
  - (б) Существует точка  $x_0 \in Q$  такая, что  $g_i(x_0) < 0$  для всех неаффинных  $g_i$ .

Аналогично безусловному случаю, справедливо **достаточное условие**:

**Теорема 1.2.** Пусть исходная задача является выпуклой. Тогда если для набора переменных  $(x^*, \lambda^*, \mu^*) \in Q \times \mathbb{R}_+^m \times \mathbb{R}^k$  выполнены все условия Каруша–Куна–Таккера, то  $x^*$  — точка глобального минимума.

**Пример, когда условие Слейтера не выполнено (простая):**

$$\min_x \{f(x) := x \mid g_1(x) := x^2 \leq 0\},$$

— выпуклая задача оптимизации. Единственная точка:  $x^* = 0$ . Однако, невозможно найти неотрицательный  $\lambda^* \geq 0$ , такой, что

$$\nabla f(0) + \lambda^* \nabla g_1(0) = 1 + \lambda^* = 0.$$

**Пример, когда условие Слейтера не выполнено (сложная):**

$$\min_x \{f(x) := x_2 \mid g(x) := x_2^2 \leq 0; h(x) := x_2 - x_1^2 - 1 = 0\},$$

— невыпуклая задача оптимизации.

$$L = x_1 + \lambda x_2^2 + \mu(x_2 - x_1^2 - 1)$$

**Пример: евклидова проекция на евклидов шар.**

$$\Pi_Q(x_0) := \operatorname{argmin}_{x \in Q} \|x - x_0\|_2^2$$

<sup>1</sup>Напомним, что *активными ограничениями* в точке  $x \in Q$  называются те ограничения, которые переходят в равенство в точке  $x$ , т. е.  $h_1, \dots, h_k$  и  $g_i$  для всех  $i$ , таких, что  $g_i(x^*) = 0$ .

$Q = B_2(0, 1) = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ .

Ответ:  $\frac{x_0}{\|x_0\|_2}$ .

**Пример: найти двойственную норму**

$$\|x\| := \sqrt{\langle Px, x \rangle}, \quad P \in S_{++}^n, \quad \text{Найти } \|y\|_*$$

Ответ:  $\|y\|_* = \sqrt{\langle P^{-1}y, y \rangle}$ .

**Пример: Steepest-descent методы**

$$d_k = \underset{v}{\operatorname{argmin}} \{ \langle \nabla f(x), v \rangle \mid \langle Pv, v \rangle \leq 1 \}, \quad P \in S_{++}^n.$$

Выписать формулу для направления  $d_k$ .

Ответ:

$$d_k = -\frac{P^{-1}\nabla f(x)}{\|\nabla f(x)\|_*}, \quad \|x\| := \sqrt{\langle Px, x \rangle},$$

## 2 Переход от негладкой задачи к гладкой

Одно из полезных применений условных задач: если исходная задача негладкая, то на практике довольно часто такую задачу можно переписать в эквивалентном гладком виде, с дополнительно введёнными переменными и условиями:

$$\min_{x \in \mathbb{R}^n, Ax=b} \|x\|_1 \quad \Leftrightarrow \quad \min_{y, z \in \mathbb{R}_+^n, Ay-Az=b} \sum_i y_i + \sum_i z_i.$$

**Пример 2.1.** Пусть  $f_1, \dots, f_m : \mathbb{R}^n \rightarrow \mathbb{R}$  — гладкие выпуклые функции. Рассмотрим задачу

$$\min_{x \in \mathbb{R}^n} \max_{1 \leq i \leq m} f_i(x).$$

Эта задача является выпуклой безусловной, но негладкой (из-за присутствия максимума). Тем не менее, она является эквивалентной следующей выпуклой условной гладкой:

$$\min_{x, t} t \quad \text{s. t.} \quad t \geq f_i(x), \quad 1 \leq i \leq m.$$

(Почему задача выпуклая?)

**Пример 2.2.** Рассмотрим задачу минимизации гладкой функции с  $\ell_1$ -регуляризатором:

$$\min_x \{f(x) + \|x\|_1\}.$$

Эта задача является выпуклой безусловной, но негладкой из-за присутствия  $\ell_1$ -нормы  $\|x\|_1$ . Превратим ее в выпуклую гладкую условную. Для этого введем дополнительные переменные  $x^+$  и  $x^-$ , что

$$x_i^+ = \max\{x_i, 0\}, \quad x_i^- = \max\{-x_i, 0\}.$$

Заметим, что из определения  $x^+, x^- \succeq 0$ . Тогда в терминах новых переменных

$$x_i = x_i^+ - x_i^-, \quad |x_i| = x_i^+ + x_i^-.$$

и задача переписывается следующим образом:

$$\min_{x^+, x^-} \{f(x^+ - x^-) + \langle 1_n, x^+ \rangle + \langle 1_n, x^- \rangle\} \quad \text{s. t.} \quad x^+, x^- \succeq 0.$$

(Почему эта задача эквивалентна исходной? Почему она выпуклая?)