Методы оптимизации, ВМК, осень 2018

Семинар 7: Условия Каруша–Куна–Таккера. Двойственность Лагранжа

1 Введение

Будем рассматривать задачи оптимизации не по всему пространству, а по некоторому его собственному подмножеству Q:

$$\min_{x \in Q} f(x). \tag{1.1}$$

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 имееются 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:

$$Q = \{x \in X \mid g_1(x) \leqslant 0, \dots, g_m(x) \leqslant 0, \ h_1(x) = 0, \dots, h_k(x) = 0\}, \tag{1.2}$$

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

$$X = \operatorname{Dom} f \cap \operatorname{Dom} g_1 \cap \dots \cap \operatorname{Dom} h_k.$$

Все функции предполагаются гладкими.

Определение 1.1. Точка x называется *допустимой*, если $x \in Q$.

Определение 1.2. 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: $f(x^*) \leqslant f(x)$.

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

2 Условия Кару́ша–Куна–Таккера

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

Определим для задачи (1.1), где множество Q задано функциональными ограничениями (1.2), *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 2.1. *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 нежёсткость.*

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

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 нежёсткости*) которые обязательно разрешимы с некоторыми, неизвестными нам двойственными переменными.

Замечание 2.2. Обратите внимание, что двойственные переменные λ_i^*, отвечающие ограничениям-неравенствам обязательно неотрицательны: $\lambda_i^* \geqslant 0$.

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 2.1:

- Все ограничения $g_1, \ldots, g_m, \; h_1, \ldots, h_k$ суть аффинные функции.
- В точке x градиенты всех *активных ограничений*[1] линейно независимы.
- Выполнено *условие Слейтера:*
  (a) Задача является выпуклой (все f и g_i суть выпуклые функции, h_j — аффинные функции).
  (b) 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 2.3. *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:

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

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

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

Пример: евклидова проекция на евклидов шар.

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

$Q = B_2(0,1) = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$.
Ответ: $\frac{x_0}{\|x_0\|_2}$.

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

Пример: найти двойственную норму

$$\|x\| := \sqrt{\langle Px, x\rangle},\ P \in S^n_{++}, \qquad \text{Найти } \ \|y\|_*$$

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

Пример: Steepest-descent методы

$$d_k \ = \ \operatorname*{argmin}_v \{\langle \nabla f(x), v\rangle \mid \langle Pv, v\rangle \leqslant 1\}, \qquad P \in \mathbb{S}^n_{++}.$$

Выписать формулу для направления d_k.
Ответ:

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

3 Двойственность Лагранжа

Ещё одно важное понятие, возникающее в условной оптимизации — двойственная задача.

Мы предложим универсальный способ: как для любой условной задачи оптимизации $\min_{x\in Q} f(x)$ записать новую *двойственную* задачу $\max_{y\in\Omega} g(y)$, при этом, двойственная задача будет оценивать прямую задачу снизу:

$$g(y) \ \leqslant \ f(x) \qquad \text{для всех} \quad y \in \Omega,\ x \in Q.$$

Поскольку неравенство справедливо для всех допустимых x и y, значит:

$$\max_{y\in\Omega} g(y) \ \leqslant \ \min_{x\in Q} f(x).$$

Прежде чем перейти к конкретному (одному из возможных) способов построения двойственной задачи, обсудим, зачем это может быть полезно.

(a) *Построение оценки снизу на решение прямой задачи.* Решить исходную задачу может быть очень сложно. Но если у нас есть двойственная к ней, то мы можем взять произвольный $y \in \Omega$ и подставить его в $g(y)$ — получим некоторую оценку снизу:

$$g(y) \ \leqslant \ \min_{x\in Q} f(x).$$

(b) *Проверка на допустимость задачи и ограниченность решения.* Из неравенства $\max_{y\in\Omega} g(y) \ \leqslant \min_{x\in Q} f(x)$ следует: если $\min_{x\in Q} f(x) = -\infty$, значит $\Omega = \emptyset$ и наоборот.

(c) *Двойственную задачу бывает решить легче, чем прямую.* При этом, если выполнена *сильная двойственность*: $g(y^*) = f(x^*)$ то мы ничего не теряем.

(d) *Получение оценки сверху на невязку по функции:* $f(x) - f^* \ \leqslant \ f(x) - g(y)$ для произвольного $y \in \Omega$.

Опишем возможный способ построения двойственной задачи для случая, когда Q задаётся функциональными ограничениями.

Заметим, что в этом случае, исходная условная задача $\min_{x\in Q} f(x)$ эквивалента следующей безусловной задаче:

$$\min_{x\in\mathbb{R}^n} \ \max_{\lambda\in\mathbb{R}^m_+,\, \mu\in\mathbb{R}^k} \mathcal{L}(x; \lambda, \mu).$$

Поменяем минимум и максимум местами — получим оценку снизу (это универсальное правило, которое всегда можно использовать):

$$\max_{\lambda\in\mathbb{R}^m_+,\,\mu\in\mathbb{R}^k}\ \min_{x\in\mathbb{R}^n}\mathcal{L}(x;\lambda,\mu)\ \leqslant\ \min_{x\in R^n}\ \max_{\lambda\in\mathbb{R}^m_+,\,\mu\in\mathbb{R}^k}\mathcal{L}(x;\lambda,\mu).$$

Обозначим внутреннюю задачу оптимизации в оценке снизу за $g(\lambda,\mu)$:

$$g(\lambda,\mu)\ :=\ \min_{x\in\mathbb{R}^n}\mathcal{L}(x;\lambda,\mu).$$

Получили *двойственную задачу*:

$$\max_{\lambda\in\mathbb{R}^m_+,\,\mu\in\mathbb{R}^k} g(\lambda,\mu).$$

Замечание 3.1. Заметим, что двойственная функция $g(\lambda,\mu)$ всегда будет вогнутой, как минимум аффинных функций.

По построению всегда выполнена *слабая двойственность*: $g^*\leqslant f^*$.

Если исходная задача <u>выпуклая</u>, и для нее выполнено <u>условие Слейтера</u>, то имеет место *сильная двойственность*: $g^*=f^*$.

Замечание 3.2. Если выполнена сильная двойственность ($g^*=f^*$), то решения прямой задачи x^* и двойственной (λ^*,μ^*) тоже связаны:

$$x^*\in\operatorname*{Argmin}_{x\in\mathbb{R}^n}\mathcal{L}(x,\lambda^*,\mu^*),$$

— решения прямой задачи принадлежат множеству минимумов лагранжиана по прямым переменным, при фиксированных оптимальных двойственных.