Методы оптимизации, ФКН ВШЭ, зима 2017

Семинар 8: Условная оптимизации. Условия ККТ. Эквивалентные преобразования задач

28 февраля 2017 г.

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

$$\min_{x \in Q} f(x). \tag{1}$$

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:

$$Q = \{x \in X \mid g_1(x) \leqslant 0, \ldots, g_m(x) \leqslant 0, \; h_1(x) = 0, \ldots, h_k(x) = 0\}, \tag{2}$$

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

$$X = \operatorname{Dom} f(x) \cap \operatorname{Dom} g_1(x) \cap \cdots \cap \operatorname{Dom} h_k(x).$$

Все функции предполагаются гладкими.

Определение 1. Точка x называется *допустимой*, если $x \in Q$.

Определение 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.

1 Условия Кару́ша–Куна–Таккера

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

Определим для задачи (1), где множество Q задано функциональными ограничениями (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. *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:

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

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

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 1. 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 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:*
  1. Задача является выпуклой (все f и g_i суть выпуклые функции, h_j — аффинные функции).
  2. 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. *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.*

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:

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

[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 1. Рассмотрим задачу минимизации гладкой функции с ℓ_1-регуляризатором:

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

Эта задача является выпуклой безусловной, но негладкой из-за присутствия ℓ_1-нормы $\|x\|_1$. Превратим ее в выпуклую гладкую условную. Для этого введем дополнительные переменные x^+ и x^-, что

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

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

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

и задача переписывается следующим образом:

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

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