Условия оптимальности. Теория двойственности

Александр Катруца

Московский физико-технический институт,
Факультет Управления и Прикладной Математики

17 октября 2017 г.

Напоминание

- Выпуклые множества
- Выпуклые функции
- Критерии выпуклости
- Операции, сохраняющие выпуклость

Мотивация

Вопрос 0

Когда существует решение оптимизационной задачи?

Вопрос 1

Как проверить, что точка является решением оптимизационной задачи?

Вопрос 2

Из каких условий можно найти решение оптимизационной задачи?

Существование решения

Теорема Вейерштрасса

Пусть $X \subset R^n$ компактное множество и пусть $f(x)$ непрерывная функция на X. Тогда точка глобального минимума функции $f(x)$ на X существует.

Эта теорема гарантирует, что решение подавляющего большинства разумных задач существует.

Условия оптимальности

Определение

Условием оптимальности будем называть некоторое выражение, выполнимость которого даёт необходимое и (или) достаточное условие экстремума.

Классы задач:

- Общая задача минимизации
- Задача безусловной минимизации
- Задача минимизации с ограничениями типа равенств
- Задача минимизации с ограничениями типа равенств и неравенств

Общая задача минимизации

Задача

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

Критерий оптимальности

Пусть $f(x)$ определена на множестве $X \subset \mathbb{R}^n$. Тогда

1. если x^* точка минимума $f(x)$ на X, то $\partial_X f(x^*) \neq \emptyset$ и $0 \in \partial_X f(x^*)$
2. если для некоторой точки $x^* \in X$ существует субдифференциал $\partial_X f(x^*)$ и $0 \in \partial_X f(x^*)$, то x^* — точка минимума $f(x)$ на X.

Какие недостатки у приведённого критерия?

Задача безусловной минимизации

Задача: $f(x) \to \min\limits_{x \in \mathbb{R}^n}$.

Критерий оптимальности для выпуклых функций

Пусть $f(x)$ выпуклая функция на $\mathbb{R}^n$. Тогда точка x^* решение задачи безусловной минимизации $\Leftrightarrow 0 \in \partial f(x^*)$.

Следствие

Если $f(x)$ выпукла и дифференцируема на $\mathbb{R}^n$. Тогда точка x^* решение задачи безусловной минимизации $\Leftrightarrow$ $\nabla f(x^*) = 0$.

Достаточное условие для невыпуклых функций

Пусть f дважды дифференцируема на $\mathbb{R}^n$ и x^* такая что $\nabla f(x^*) = 0$. Тогда если $\nabla^2 f(x^*) \succ 0$, то x^* точка строгого локального минимума $f(x)$ на $\mathbb{R}^n$.

Задача

$$f(x) \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } g_i(x) = 0, \; i = 1, \dots, m$$

Лагранжиан

$$L(x, \boldsymbol{\lambda}) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x)$$

Критерий оптимальности

Пусть $f(x)$ и $g_i(x)$ дважды дифференцируемы в точке x^* и непрерывно дифференцируемы в некоторой окрестности x^*. Пусть также $\nabla_x L(x^*, \boldsymbol{\lambda}) = 0$. Тогда если $\mathbf{h}^\top \nabla^2 L(x^*, \boldsymbol{\lambda}) \mathbf{h} > 0$, где $\mathbf{h} \in T(\mathbf{x}^* | G)$ — касательный конус, то x^* — точка локального минимума.

Возможные варианты

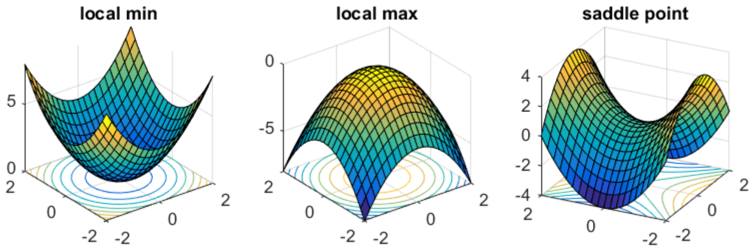


Рис.: Рисунок взят из блога
http://www.offconvex.org/2016/03/22/saddlepoints/

Задача минимизации с ограничениями типа равенств и неравенств

Задача

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } g_i(x) = 0, \ i = 1, \ldots, m$$
$$h_j(x) \leq 0, \ j = 1, \ldots, p$$

Лагранжиан

$$L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Условия оптимальности

Необходимое условие (Каруша-Куна-Такера)

Пусть x^* решение задачи математического программирования, и функции f, h_j, g_i дифференцирумы. Тогда найдутся такие $\boldsymbol{\mu}^*$ и $\boldsymbol{\lambda}^*$, что выполнены следующие условия:

- $g_i(x^*) = 0$
- $h_j(x^*) \leq 0$
- $\mu_j^* \geq 0$
- $\mu_j^* h_j(x^*) = 0$
- $\nabla_x L(x^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0$

Если задача выпуклая, то это же условие является достаточным.

Условия оптимальности (cont'd)

Если задача невыпуклая, то

Достаточное условие первого порядка

Если для стационарной точки $(x^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ число активных неравенств $|J|$ такое что $n = m + |J|$ и $\mu_j > 0$, $j \in J$, то эта точка является точкой минимума.

Достаточное условие второго порядка

Если в задаче математического программирования число активных ограничений меньше размерности задачи, то точка x^* яляется решением задачи, если выполнены условия

$$\mathbf{z}^\mathsf{T} \nabla^2_{xx} L(x^*) \mathbf{z} > 0$$

для

- $\mathbf{z} \neq 0$ и $\nabla g_i^\mathsf{T}(x^*)\mathbf{z} = 0$
- при $j \in J$ и $\mu_j > 0$, $\nabla h_j^\mathsf{T}(x^*)\mathbf{z} = 0$
- при $j \in J$ и $\mu_j = 0$, $\nabla h_j^\mathsf{T}(x^*)\mathbf{z} \leq 0$

Двойственность: обозначения

Задача

$$\min_{x \in \mathcal{D}} f(x) = p^*$$

$$\text{s.t. } g_i(x) = 0, \; i = 1, \ldots, m$$

$$h_j(x) \leq 0, \; j = 1, \ldots, p$$

Лагранжиан

$$L(x, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{p} \mu_j h_j(x)$$

Двойственные переменные

Вектора $\boldsymbol{\mu}$ и $\boldsymbol{\lambda}$ называются двойственными переменными.

Двойственная функция

Функция $g(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \inf_{x \in \mathcal{D}} L(x, \boldsymbol{\lambda}, \boldsymbol{\mu})$ называется двойственной функцией Лагранжа.

Свойства двойственной функции

Вогнутость

Двойственная функция является вогнутой как инфимум аффинных функций по $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ вне зависимости от того, является ли исходная задача выпуклой.

Нижняя граница

Для любого $\boldsymbol{\lambda}$ и для $\boldsymbol{\mu} \geq 0$ выполнено $g(\boldsymbol{\mu}, \boldsymbol{\lambda}) \leq p^*$.

Двойственная задача

$$\max g(\boldsymbol{\mu}, \boldsymbol{\lambda}) = d^*$$
$$\text{s.t. } \boldsymbol{\mu} \geq 0$$

Зачем?

- Двойственная задача выпукла независимо от того, выпукла ли прямая
- Нижняя оценка **может** достигаться

Примеры

Найти двойственную функцию:

- Решение СЛУ минимальной нормы

$$
\begin{aligned}
\min \|\mathbf{x}\|_2^2 \\
\text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}
\end{aligned}
$$

- Линейное программирование

$$
\begin{aligned}
\min \mathbf{c}^\top \mathbf{x} \\
\text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b} \\
\mathbf{x} \geq 0
\end{aligned}
$$

- Задача разбиения

$$
\begin{aligned}
\min \mathbf{x}^\top \mathbf{W} \mathbf{x} \\
\text{s.t. } x_i^2 = 1, \; i = 1, \ldots, n
\end{aligned}
$$

Слабая и сильная двойственность

Определение

Оптимальные значения целевой функции в прямой и двойственной задаче связаны соотношением

$$d^* \leq p^*.$$

Если $d^* < p^*$, то свойство называют слабой двойственностью.
Если $d^* = p^*$, то — сильной двойственностью.

Замечание

Слабая двойственность есть всегда по построению двойственной задачи.

Вопросы

- При каких условиях выполняется сильная двойственность?
- Как использовать двойственность для проверки оптимальности?

Критерий субоптимальности

По построению $p^* \geq g(\boldsymbol{\lambda}, \boldsymbol{\mu})$, поэтому
$f_0(x) - p^* \leq f_0(x) - g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \varepsilon$.

Определение

Разность $f_0(x) - g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ называется *двойственным зазором* и является оценкой сверху для разности текущего и оптимального значения функции.

Способы использования:

- критерий остановки в итерационном процессе
- теоретическая оценка сходимости алгоритма
- проверка оптимальности данной точки

Условия Слейтера

Теорема

Если задача выпуклая и существует x, лежащий внутри допустимой области, т.е. ограничения типа неравенств выполнены как строгие неравенства, то выполнено свойство сильной двойственности.

- Решение СЛАУ наименьшей нормы
- Линейное программирование
- Квадратичное программирование с квадратичными огранчиениями
- Невыпуклая задача с сильной двойственностью

Условия дополняющей нежёсткости

Пусть $\mathbf{x}^*$ и $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ решения прямой и двойственной задачи. То есть

$$f(\mathbf{x}^*) = g(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq$$

$$f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* g_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j^* h_j(\mathbf{x}^*) \leq$$

$$f(\mathbf{x}^*), \qquad \boldsymbol{\mu} \geq 0$$

Условия дополняющей нежёсткости

$$\mu_j^* h_j(\mathbf{x}^*) = 0, \qquad j = 1, \dots, p$$

Для каждого неравенства

- либо множитель Лагранжа равен нулю
- либо оно активно.

Условия Каруша-Куна-Таккера

Необходимые условия ККТ:

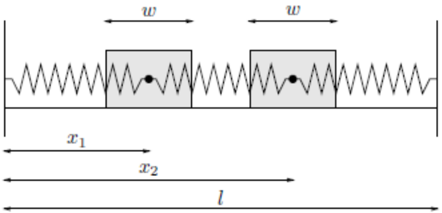
1. $g_i(x^*) = 0$ — допустимость в прямой задаче
2. $h_j(x^*) \leq 0$ — допустимость в прямой задаче
3. $\mu_j^* \geq 0$ — допустимость в двойственной задаче
4. $\mu_j^* h_j(x^*) = 0$ — условие дополняющей нежёсткости
5. $\nabla_x L(x^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0$ — стационарность лагранжиана по прямым переменным

Пример ($\mathbf{P} \in \mathcal{S}^n_+$)

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\mathbf{x}^\top \mathbf{P} \mathbf{x} + \mathbf{q}^\top \mathbf{x} + r$$

$$\text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}$$

Механическая интерпретация



Поиск устойчивого положения системы:

$$
\begin{aligned}
&\min_{\mathbf{x}\in\mathbb{R}^3} \frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2 \\
\text{s.t. } &\frac{w}{2} - x_1 \leq 0 \\
&w + x_1 - x_2 \leq 0 \\
&\frac{w}{2} - l + x_2 \leq 0
\end{aligned}
$$

Примеры

- Орицательная энтропия при линейных ограничениях
$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} & \sum_{i=1}^n x_i \log x_i \\ \text{s.t. } & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & \mathbf{1}^\top \mathbf{x} = 1 \end{aligned}$$
- Сформулировать двойственную задачу и по её решению найти решение прямой задачи:
$$\begin{aligned} & \min \frac{1}{2}x^2 + 2y^2 + \frac{1}{2}z^2 + x + y + 2z \\ \text{s.t. } & x + 2y + z = 4 \end{aligned}$$
- Релаксация Лагранжа для задачи бинарного линейного программирования:
$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} & \mathbf{c}^\top \mathbf{x} \\ \text{s.t. } & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\ & x_i \in \{0, 1\}, \quad i = 1, \ldots, n \end{aligned}$$

Резюме

- Существование решения оптимизационной задачи
- Условия оптимальности для
  - общей задачи оптимизации
  - задачи безусловной оптимизации
  - задачи оптимизации с ограничениями типа равенств
  - задачи оптимизации с ограничениями типа равенств и неравенств
- Двойственая задача: что это такое и зачем оно надо?
- Сильная и слабая двойственность
- Условия Слейтера