Условия оптимальности и введение в теорию двойственности

Александр Катруца

18 марта 2018 г.

1. Условия оптимальности для выпуклых задач

Условия оптимальности дают необходимые и/или достаточные условия минимума для выпуклых задач. То есть формализуют поиск решения задачи минимизации и/или дают условия, при которых точка не является решением. Отметим важный факт о решении задачи выпуклой оптимизации

Теорема 1 *Если f — выпуклая функция и $\mathbf{x}^*$ точка локального минимума, то $\mathbf{x}^*$ точка глобального минимума*

Рассмотрим два класса задач выпуклой оптимизации и сформулируем для них условия оптимальности.

1.1 Безусловная оптимизация

Теорема 2 *Для безусловной задачи выпуклой оптимизации вида*

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}),$$

где f дифференцируема, точка $\mathbf{x}^*$ является решением тогда и только тогда, когда выполнено $f'(\mathbf{x}^*) = 0$.

1.2 Условная оптимизация

Теорема 3 *Для выпуклой задачи условной оптимизации вида*

$$\begin{aligned} &\min f_0(\mathbf{x}) \\ \text{s.t. } &\mathbf{A}\mathbf{x} = \mathbf{b} \\ &f_i(\mathbf{x}) \leq 0,\ i = 1, \ldots, p \end{aligned}$$

где $f_i,\ i = 0, \ldots, p$ — выпуклые функции точка $\mathbf{x}^*$ является решением тогда и только тогда, когда существуют вектора $\boldsymbol{\mu}^*$ и $\boldsymbol{\lambda}^*$, такие что выполнены следующие условия

1. $f_i(\mathbf{x}^*) \leq 0,\ i = 1, \ldots, p$

2. $\mathbf{A}\mathbf{x}^* = \mathbf{b}$

3. $\boldsymbol{\lambda}^* \geq 0$

4. $\lambda_i^* f_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, p$

5. $L'_x(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0,$

где L *— лагранжиан, который имеет вид*

$$L = f_0(\mathbf{x}) + \boldsymbol{\mu}^\intercal(\mathbf{A}\mathbf{x} - \mathbf{b}) + \sum_{i=1}^{p} \lambda_i f_i(\mathbf{x})$$

Эти условия называются условиями Каруша-Куна-Таккера, сокращённо ККТ.

Первые два условия означают, что $\mathbf{x}^*$ лежит в допустимом множестве. Последнее условие означает стационарность лагранжиана в точке $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. Условия 3 и 4 называются условиями допустимости в двойственной задачи и условием дополняющей нежёсткости, соответственно. Откуда они берутся станет ясно после рассмотрения того, как строится двойственная задача и при каких условиях выполняется сильная двойственность.

Обратите внимание на следующие три обстоятельства:

1. Если рассмотреть не выпуклую задачу, а произвольную, то можно записать аналогичные 5 условий. Однако в этом случае они будут только необходимыми, но не достаточными условиями минимума точки $\mathbf{x}^*$

2. Слагаемые лагранжиана, в которые входят вектора $\boldsymbol{\mu}$ и $\boldsymbol{\lambda}$, можно записать как скалярные произведения, поэтому аналогичные условия можно записать для выпуклых функции, заданных на некотором выпуклом множестве матриц. В этом случае скалярное произведение будет записано для матриц в виде $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{Tr}(\mathbf{X}^\intercal \mathbf{Y})$, а условие 3 превратится в условие неотрицательной определённости матрицы $\boldsymbol{\Lambda}$.

   Упражнение 1 *Запишите условия ККТ для задачи нахождения матрицы ковариаций многомерного нормального распределения по принципу максимума правдоподобия при известном среднем.*

3. Условие 4 можно переформулировать как альтернативу: либо $\lambda_i = 0$, либо $f_i(\mathbf{x}^*) = 0$. Одновременное равенство нулю возможно, но не несёт в себе никакой новой информации, поскольку равенство 0 λ_i означает, что градиент лагранжиана вырождается в градиент целевой функции, равенство нулю которого эквивалентно тому, что точка является решением. При этом $f_i(\mathbf{x}^*)$ может быть равно нулю, но проверять это отдельно ненужно. Подробнее написано тут.

2. Введение в теорию двойственности

Теория двойственности позволяет переформулировать *любую* задачу выпуклой оптимизации в виде *выпуклой* задачи оптимизации (она называется двойственной), целевая функция которой оценивает снизу целевую функцию исходной задачи. Можно провести аналогию между двойственной задачей и описанием выпуклого множества как пересечения полупространств, образованных касательными гиперплоскостями в граничных точках множества.

Как строить выпуклую задачу в общем виде описано в презентации. Ниже расммотрены несколько примеров, для которых получим двойственную задачу и в некоторых случаях её аналитическое (!) решение.

2.1 Примеры

1. Задача поиска решения недоопределённой системы линейных уравнений минимальной нормы

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

   Запишем лагранжиан

$$L = \|\mathbf{x}\|_2^2 + \boldsymbol{\lambda}^\intercal(\mathbf{A}\mathbf{x} - \mathbf{b})$$

   Двойственная функция

$$g(\boldsymbol{\lambda}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \min_{\mathbf{x}} \|\mathbf{x}\|_2^2 + \boldsymbol{\lambda}^\intercal(\mathbf{A}\mathbf{x} - \mathbf{b})$$

   В данном случае лагранжиан является выпуклой функцией по $\mathbf{x}$, поэтому для нахождения минимума достаточно найти стационарные точки, то есть корни уравнения $L'_x = 0$:

$$L'_x = 2\mathbf{x}^* + \mathbf{A}^\intercal\boldsymbol{\lambda} = 0$$

   Тогда $\mathbf{x}^* = -\frac{1}{2}\mathbf{A}^\intercal\boldsymbol{\lambda}$ и двойственная функция

$$g(\lambda) = \frac{1}{4}\boldsymbol{\lambda}^\intercal\mathbf{A}\mathbf{A}^\intercal\boldsymbol{\lambda} - \frac{1}{2}\boldsymbol{\lambda}^\intercal\mathbf{A}\mathbf{A}^\intercal\boldsymbol{\lambda} - \boldsymbol{\lambda}^\intercal\mathbf{b} = -\frac{1}{4}\boldsymbol{\lambda}^\intercal\mathbf{A}\mathbf{A}^\intercal\boldsymbol{\lambda} - \boldsymbol{\lambda}^\intercal\mathbf{b}$$

   Заметим, что как и ожидалось из теории, она является вогнутой. Двойственная задача примет вид

$$\max_{\boldsymbol{\lambda}} -\frac{1}{4}\boldsymbol{\lambda}^\intercal\mathbf{A}\mathbf{A}^\intercal\boldsymbol{\lambda} - \boldsymbol{\lambda}^\intercal\mathbf{b}$$

   Таким образом, выпуклую задачу с ограничениями типа равенства свели к выпуклой задачи без ограничений меньшей размерности. В силу выпуклости исходной задачи выполняется сильная двойственность и решение прямой задачи можно легко восстановить из решения двойственной задачи по формуле $\mathbf{x}^* = -\frac{1}{2}\mathbf{A}^\intercal\boldsymbol{\lambda}^*$.

   Так как в исходной задаче нет ограничений типа неравенств, то отсутствуют условия дополняющей нежёсткости и ограничение на допустимость в двойственной задаче.

2. Задача линейного программирования

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

   Лагранжиан

$$L = \mathbf{c}^\intercal\mathbf{x} + \boldsymbol{\mu}^\intercal(\mathbf{A}\mathbf{x} - \mathbf{b}) - \sum_{i=1}^n x_i\lambda_i = (\mathbf{c} + \mathbf{A}^\intercal\boldsymbol{\mu} - \boldsymbol{\lambda})^\intercal\mathbf{x} - \boldsymbol{\mu}^\intercal\mathbf{b}.$$

   Лагранжиан снова выпуклая функция по $\mathbf{x}$, точнее линейная, поэтому в общем случае она является неограниченной снизу. Для того, чтобы минимуму лагранжиана был конечен необходимо наложить дополнительное ограничение $\mathbf{c} + \mathbf{A}^\intercal \boldsymbol{\mu} - \boldsymbol{\lambda} = 0$. Таким образом, двойственная функция примет вид

$$g(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \begin{cases} -\mathbf{b}^\intercal \boldsymbol{\mu}, & \mathbf{c} + \mathbf{A}^\intercal \boldsymbol{\mu} - \boldsymbol{\lambda} = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Двойственная задача

$$\begin{aligned} &\min \mathbf{b}^\intercal \boldsymbol{\mu} \\ \text{s.t.} &\mathbf{c} + \mathbf{A}^\intercal \boldsymbol{\mu} - \boldsymbol{\lambda} = 0 \\ &\boldsymbol{\lambda} \geq 0 \end{aligned}$$

или

$$\begin{aligned} &\min \mathbf{b}^\intercal \boldsymbol{\mu} \\ \text{s.t.} &\mathbf{c} + \mathbf{A}^\intercal \boldsymbol{\mu} \geq 0 \end{aligned}$$

Двойственная задача снова является задачей линейного программирования, но меньшей размерности, так как в матрице $\mathbf{A}$ число строк меньше числа столбцов, и с меньшим числом ограничений.

Так как исходная задача была выпуклой, то выполняется сильная двойственность.

3. Задача о разбиении

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

где $\mathbf{W} \in \mathbf{S}^n$.

Обратите внимание что это задача дискретной оптимизации, то есть для её точного решения неизвестны быстрые алгоритмы, однако и для таких задач можно строить двойственные задачи, которые будут выпуклыми и будут давать оценки снизу на оптимальные значения целевых функций в прямых задачах.

Итак, лагранжиан для нашей задачи имеет вид

$$L = \mathbf{x}^\intercal \mathbf{W} \mathbf{x} + \sum_{i=1}^{n} \lambda_i (x_i^2 - 1) = \mathbf{x}^\intercal (\mathbf{W} + \operatorname{diag}(\boldsymbol{\lambda})) \mathbf{x} - \mathbf{1}^\intercal \boldsymbol{\lambda}$$

Лагранжиан является выпуклой функцией по $\mathbf{x}$, если матрица $\mathbf{W} + \operatorname{diag}(\boldsymbol{\lambda})$ неотрицательно определена, иначе лагранжиан является неограниченной снизу функцией и его минимум равен $-\infty$. Таким образом,

$$g(\boldsymbol{\lambda}) = \begin{cases} -\mathbf{1}^\intercal \boldsymbol{\lambda}, & \mathbf{W} + \operatorname{diag}(\boldsymbol{\lambda}) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Двойственная задача имеет вид

$$\begin{aligned} &\min_{\boldsymbol{\lambda}} \mathbf{1}^\intercal \boldsymbol{\lambda} \\ \text{s.t. } &\mathbf{W} + \operatorname{diag}(\boldsymbol{\lambda}) \succeq 0 \end{aligned}$$

Задача дискретной оптимизации была сведена к выпуклой задаче минимизации, решение которой даёт оценку снизу на оптимальное значение целевой функции в исходной задаче.

Также отметим, что для любого $\boldsymbol{\lambda}$ из допустимого множества двойственная функция даёт оценку снизу на оптимальное значение исходной целевой функции. Самым простым примером $\boldsymbol{\lambda}$ из допустимого множества является вектор из минимальных собственных значений матрицы $\mathbf{W}$ (проверьте почему?). Соответственно значение $n\lambda_{\min}(\mathbf{W})$ оценивает снизу минимальное значение функции $\mathbf{x}^\mathsf{T}\mathbf{W}\mathbf{x}$ для $x_i = \pm 1$.

Минимальное собственное значение матрицы можно найти за $\mathcal{O}(n^3)$ операций, что существенно быстрее полного перебора.