Приложение: теоремы Куна — Таккера

Теоремы Куна — Таккера — это обобщение теоремы Лагранжа на случай задач оптимизации с ограничениями в виде неравенств. «Теоремы Куна — Таккера» — это родовое название для семейства похожих теорем, отличающихся формулировками.

Мы рассмотрим эти теоремы в дифференциальной форме (т.е. когда все функции дифференцируемы).

Пусть даны функции

$$\phi\colon X \mapsto \mathbb{R}$$

и

$$\psi_r\colon X \mapsto \mathbb{R},\ r = 1,\ldots,m,$$

где $X \subseteq \mathbb{R}^n$. Функцию $\phi(\cdot)$ будем использовать как целевую функцию, а с помощью функций $\psi_r(\cdot)$ будем задавать ограничения. При этом мы получим следующую задачу нелинейного программирования:

$$\begin{gathered} \phi(\boldsymbol{x}) \to \max \\ \psi_j(\boldsymbol{x}) \geqslant 0,\ j = 1, \ldots, m, \\ \boldsymbol{x} = (x_1, \ldots, x_n) \in X. \end{gathered} \tag{*}$$

Функция Лагранжа (лагранжиан) этой задачи имеет следующий вид:

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \lambda_0 \phi(\boldsymbol{x}) + \sum_{j=1}^{m} \lambda_j \psi_j(\boldsymbol{x}),$$

где коэффициенты λ_j называют множителями Лагранжа.

Теорема Джона.

Пусть функции $\phi(\cdot),\ \psi_1(\cdot), \ldots, \psi_n(\cdot)$ дифференцируемы, и $\bar{\boldsymbol{x}}$ — решение задачи (*), такое что $\bar{\boldsymbol{x}} \in \operatorname{int}(X)$.

Тогда существуют множители Лагранжа $\lambda_j \geqslant 0$, $j = 0, \ldots, m$, не все из которых равны нулю, такие что выполнены следующие соотношения (условия Куна — Таккера):

$$\frac{\partial L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda})}{\partial x_i} = 0,\ i = 1, \ldots, n$$

и

$$\sum_{j=1}^{m} \frac{\partial L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda})}{\partial \lambda_j} \lambda_j = 0$$

(дополняющая нежесткость).

Отметим, что условия дополняющей нежесткости можно записать в виде

$$\psi_j(\bar{\boldsymbol{x}})\lambda_j=0,\ j=1,\ldots,m.$$

Из условия дополняющей нежесткости следует, что если множитель Лагранжа положителен ($\lambda_j>0$), то соответственное ограничение выходит в оптимуме на равенство (активно), т.е. $\psi_j(\bar{\boldsymbol{x}})=0$. С другой стороны, если ограничение выполнено как строгое неравенство ($\psi_j(\bar{\boldsymbol{x}})>0$), то соответствующий множитель Лагранжа равен нулю.

Если в задаче (*) часть ограничений имеет вид ограничений на неотрицательность некоторых x_i, то для них можно не вводить множители Лагранжа. Поэтому модифицируем задачу, записав такие ограничения отдельно:

$$\begin{gathered}\phi(\boldsymbol{x})\to\max\\ \psi_j(\boldsymbol{x})\geqslant 0,\ j=1,\ldots,m,\\ \boldsymbol{x}\in X,\\ x_i\geqslant 0,\ i\in P\subseteq\{1,\ldots,n\}.\end{gathered}\qquad (**)$$

Условия первого порядка для $i\in P$ будут иметь вид

$$\frac{\partial L(\bar{\boldsymbol{x}},\boldsymbol{\lambda})}{\partial x_i}\leqslant 0.$$

Для $i\notin P$ здесь, как и в случае задачи (*), будет равенство.

Кроме того, выполнены тоже условия дополняющей нежесткости

$$\begin{gathered}\sum_{j=1}^{m}\frac{\partial L(\bar{\boldsymbol{x}},\boldsymbol{\lambda})}{\partial \lambda_j}\lambda_j=0,\\ \sum_{i=1}^{n}\frac{\partial L(\bar{\boldsymbol{x}},\boldsymbol{\lambda})}{\partial x_i}\bar{x}_i=0.\end{gathered}$$

Из второго из этих условий следует, что при $\bar{x}_i>0$

$$\frac{\partial L(\bar{\boldsymbol{x}},\boldsymbol{\lambda})}{\partial x_i}=0.$$

С другой стороны, если $\partial L(\bar{\boldsymbol{x}},\boldsymbol{\lambda})/\partial x_i<0$, то соответствующий $\bar{x}_i$ должен быть равен нулю.

Другая модификация теоремы связана с наличием в задаче ограничений в виде равенств. Обозначим множество соответствующих индексов через E. Задача принимает следующий вид:

$$\begin{gathered}\phi(\boldsymbol{x})\to\max\\ \psi_j(\boldsymbol{x})\geqslant 0,\ j\in\{1,\ldots,m\}\backslash E,\\ \psi_j(\boldsymbol{x})=0,\ j\in E,\\ \boldsymbol{x}\in X,\\ x_i\geqslant 0,\ i\in P\subseteq\{1,\ldots,n\}.\end{gathered}\qquad (***)$$

При этом в теореме Джона снимается условие, что все множители Лагранжа неотрицательны — множители Лагранжа λ_j при $j\in E$ могут иметь произвольный знак.

Теорема Джона не гарантирует, что множитель Лагранжа целевой функции, λ_0, не равен нулю. Однако если $\lambda_0=0$, то эта теорема не имеет непосредственной связи с интересующей нас задачей максимизации функции $\phi(\cdot)$, поскольку градиент самой функции $\phi(\cdot)$ «пропадает» из условий Куна — Таккера. Поэтому важно иметь условия (условия регулярности), которые гарантируют, что $\lambda_0>0$.

Одно из условий, которое это гарантирует, формулируется следующим образом: градиенты активных ограничений в точке $\bar{\boldsymbol{x}}$ линейно независимы.

Обозначим через A множество индексов тех ограничений, которые в точке оптимума $\bar{\boldsymbol{x}}$ активны, т.е.

$$\psi_j(\bar{\boldsymbol{x}})=0 \Leftrightarrow j\in A.$$

Условия регулярности состоят в том, что система векторов

$$\{\nabla\psi_j(\bar{\boldsymbol{x}})\}_{j\in A}$$

является линейно независимой.

Заметим, что если $\lambda_0>0$, то без потери общности можно принять $\lambda_0=1$, что обычно и делают. Соответствующую теорему и называют собственно (прямой) теоремой Куна — Таккера.

> **Прямая теорема Куна — Таккера** (необходимое условие оптимальности)
>
> Пусть функции $\phi(\cdot),\ \psi_1(\cdot),\dots,\psi_n(\cdot)$ дифференцируемы, и $\bar{\boldsymbol{x}}$ — решение задачи (*), такое что $\bar{\boldsymbol{x}}\in \mathrm{int}(X)$ и выполнены условия регулярности.
>
> Тогда существуют множители Лагранжа $\lambda_j\geqslant 0,\ j=1,\dots,m$, такие что при $\lambda_0=1$ выполнены следующие соотношения:
>
> $$\frac{\partial L(\bar{\boldsymbol{x}},\boldsymbol{\lambda})}{\partial x_i}=0,\ i=1,\dots,n$$
>
> и
>
> $$\sum_{j=1}^{m}\frac{\partial L(\bar{\boldsymbol{x}},\boldsymbol{\lambda})}{\partial\lambda_j}\lambda_j=0.$$

Несложно переформулировать эту теорему для задач (**) и (***). Здесь требуются такие же модификации условий Куна — Таккера, как и в теореме Джона.

Условие

$$\frac{\partial L(\bar{\boldsymbol{x}},\boldsymbol{\lambda})}{\partial x_i}=0,\ i=1,\dots,n$$

можно переписать в виде:

$$\nabla\phi(\bar{\boldsymbol{x}})=-\sum_{j=1}^{m}\lambda_j\nabla\psi_j(\bar{\boldsymbol{x}}).$$

Это условие выражает в точке оптимума градиент целевой функции в виде линейной комбинации градиентов ограничений, причем все коэффициенты этой линейной комбинации неположительны.

Обратная теорема Куна — Таккера утверждает, что при вогнутости функций $\phi(\cdot),\ \{\psi_k(\cdot)\}$, условия в допустимой точке $\bar{\boldsymbol{x}}$ нашлись множители Лагранжа удовлетворяющие требованиям прямой теоремы, то эта точка $\bar{\boldsymbol{x}}$ является оптимумом.

Обратная теорема Куна–Таккера (достаточное условие оптимальности)

Пусть функции $\phi(\cdot),\ \psi_1(\cdot), \ldots, \psi_n(\cdot)$ дифференцируемы и вогнуты, множество X выпукло и точка $\bar{\boldsymbol{x}}$ допустима в задаче (*), причем $\bar{\boldsymbol{x}} \in \operatorname{int}(X)$.

Пусть, кроме того, существуют множители Лагранжа $\lambda_j \geqslant 0$, $j = 1, \ldots, m$, такие что при $\lambda_0 = 1$ выполнены следующие соотношения:

$$\frac{\partial L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda})}{\partial x_i} = 0,\ i = 1, \ldots, n$$

и

$$\sum_{j=1}^{m} \frac{\partial L(\bar{\boldsymbol{x}}, \boldsymbol{\lambda})}{\partial \lambda_j} \lambda_j = 0.$$

Тогда $\bar{\boldsymbol{x}}$ — решение задачи (*).

Теорему можно очевидным образом переформулировать для задач (**) и (***).