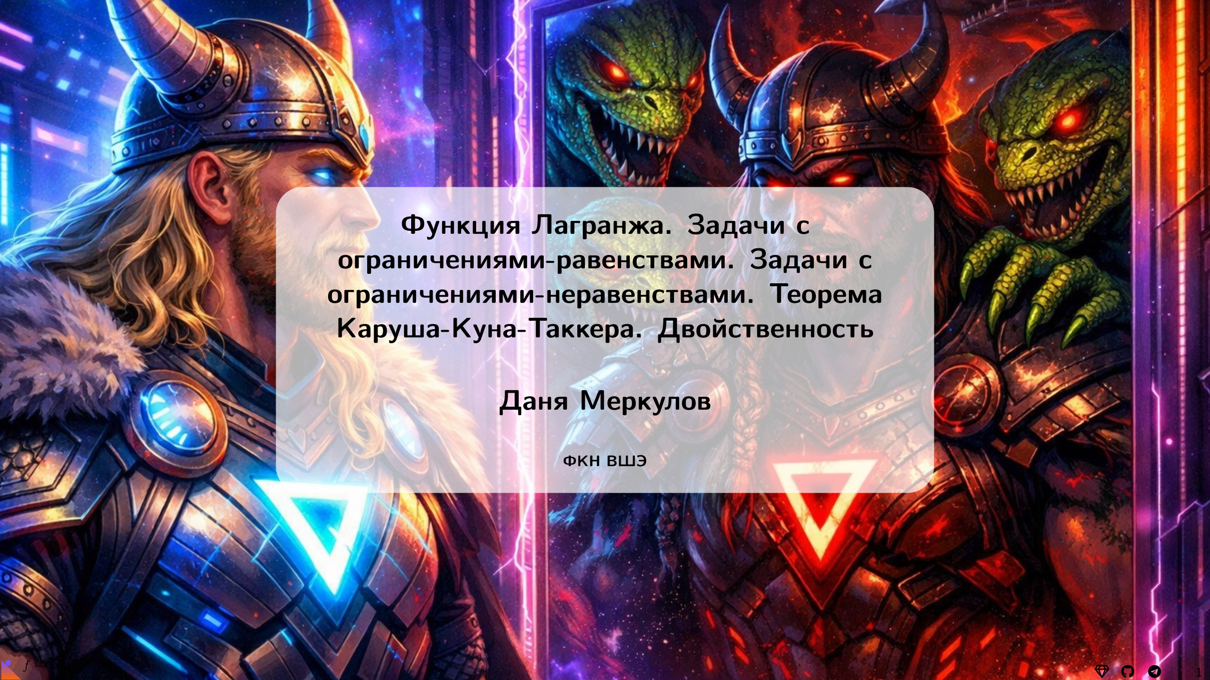


Функция Лагранжа. Задачи с ограничениями-равенствами. Задачи с ограничениями-неравенствами. Теорема Каруша-Куна-Таккера. Двойственность

Даня Меркулов

ФКН ВШЭ

Условная оптимизация

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

1. Тогда для любого допустимого направления $d \in \mathbb{R}^n$ в x^* выполняется $\nabla f(x^*)^T d \geq 0$.

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.
Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

1. Тогда для любого допустимого направления $d \in \mathbb{R}^n$ в x^* выполняется $\nabla f(x^*)^T d \geq 0$.
2. Если, кроме того, S выпукло, то

$$\nabla f(x^*)^T (x - x^*) \geq 0, \forall x \in S.$$

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

1. Тогда для любого допустимого направления $d \in \mathbb{R}^n$ в x^* выполняется $\nabla f(x^*)^T d \geq 0$.
2. Если, кроме того, S выпукло, то

$$
\nabla f(x^*)^T (x - x^*) \geq 0, \forall x \in S.
$$

Общее условие локальной оптимальности первого порядка

Вектор $d \in \mathbb{R}^n$ является допустимым направлением в точке $x^* \in S \subseteq \mathbb{R}^n$, если малые шаги вдоль d не выводят нас за пределы S.

Рассмотрим множество $S \subseteq \mathbb{R}^n$ и функцию $f : \mathbb{R}^n \to \mathbb{R}$. Предположим, что $x^* \in S$ является точкой локального минимума f на S, и что f непрерывно дифференцируема в окрестности x^*.

1. Тогда для любого допустимого направления $d \in \mathbb{R}^n$ в x^* выполняется $\nabla f(x^*)^T d \geq 0$.
2. Если, кроме того, S выпукло, то

$$\boxed{\nabla f(x^*)^T (x - x^*) \geq 0, \forall x \in S.}$$

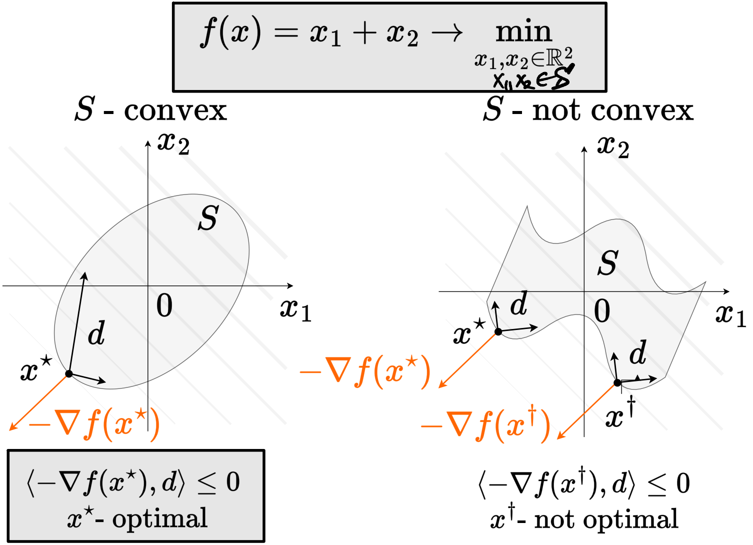


Figure 1: Общее условие локальной оптимальности первого порядка

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Ещё один важный результат для выпуклого случая звучит следующим образом: если $f(x) : S \to \mathbb{R}$ - выпуклая функция, определённая на выпуклом множестве S, то:

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Ещё один важный результат для выпуклого случая звучит следующим образом: если $f(x) : S \to \mathbb{R}$ - выпуклая функция, определённая на выпуклом множестве S, то:

- Любой локальный минимум является глобальным.

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Ещё один важный результат для выпуклого случая звучит следующим образом: если $f(x) : S \to \mathbb{R}$ - выпуклая функция, определённая на выпуклом множестве S, то:

- Любой локальный минимум является глобальным.
- Множество локальных (= глобальных) минимумов S^* выпукло.

Выпуклый случай

Следует отметить, что в **выпуклом** случае (то есть при выпуклых f и S) необходимое условие становится достаточным.

Ещё один важный результат для выпуклого случая звучит следующим образом: если $f(x) : S \to \mathbb{R}$ - выпуклая функция, определённая на выпуклом множестве S, то:

- Любой локальный минимум является глобальным.
- Множество локальных (= глобальных) минимумов S^* выпукло.
- Если $f(x)$ - строго или сильно выпуклая функция, то S^* содержит только одну точку: $S^* = \{x^*\}$.

Задачи с ограничениями-равенствами

В задачах без ограничений всё довольно интуитивно. В этом разделе мы добавим одно ограничение-равенство, то есть:

Задачи с ограничениями-равенствами

В задачах без ограничений всё довольно интуитивно. В этом разделе мы добавим одно ограничение-равенство, то есть:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & h(x) = 0
\end{split}
$$

Задачи с ограничениями-равенствами

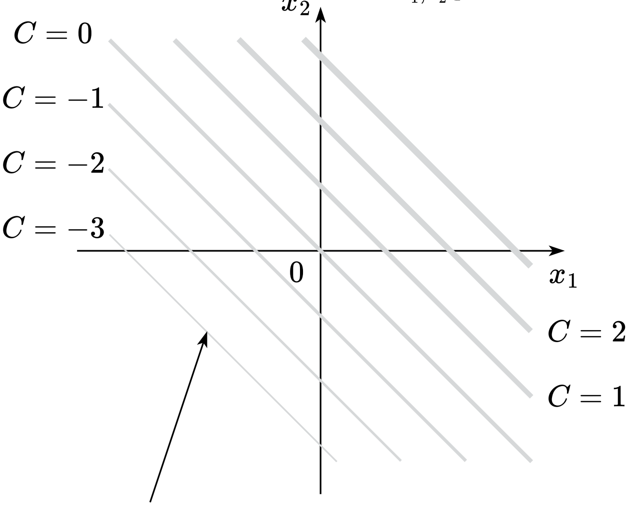
В задачах без ограничений всё довольно интуитивно. В этом разделе мы добавим одно ограничение-равенство, то есть:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & h(x) = 0
\end{split}
$$

Мы попробуем проиллюстрировать подход к решению этой задачи через простой пример с $f(x) = x_1 + x_2$ и $h(x) = x_1^2 + x_2^2 - 2$.

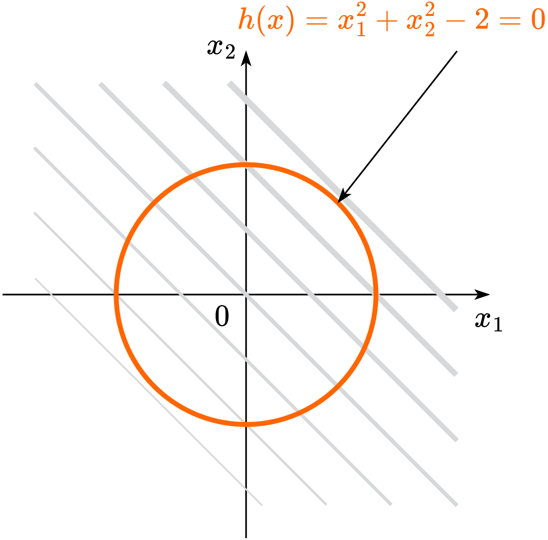
Задачи с ограничениями-равенствами

$$f(x) = x_1 + x_2 \to \min_{x_1, x_2 \in \mathbb{R}^2}$$

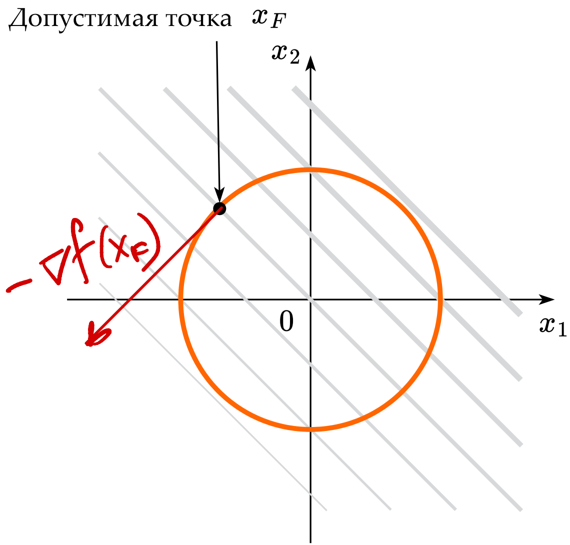


$$f(x) = x_1 + x_2 = C$$

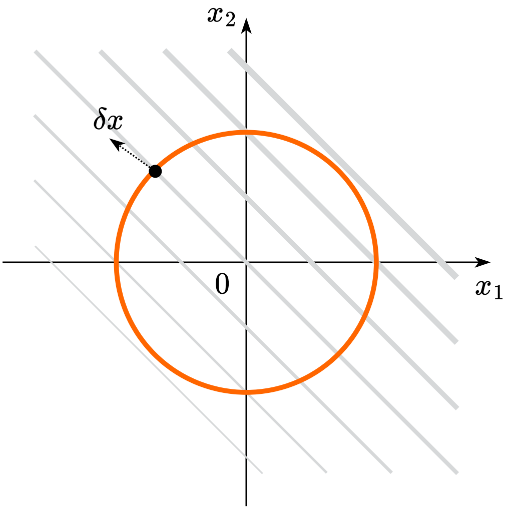
Задачи с ограничениями-равенствами



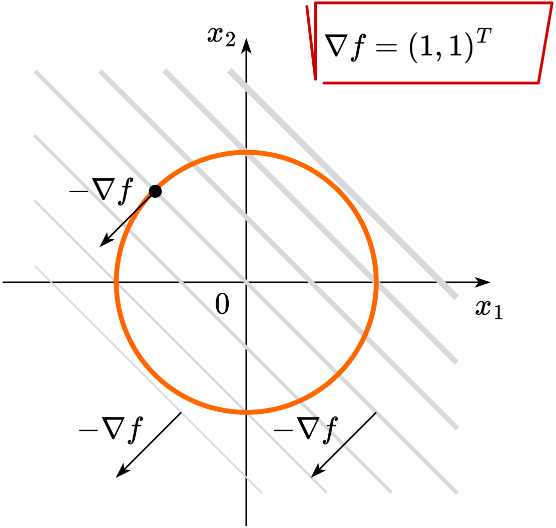
Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

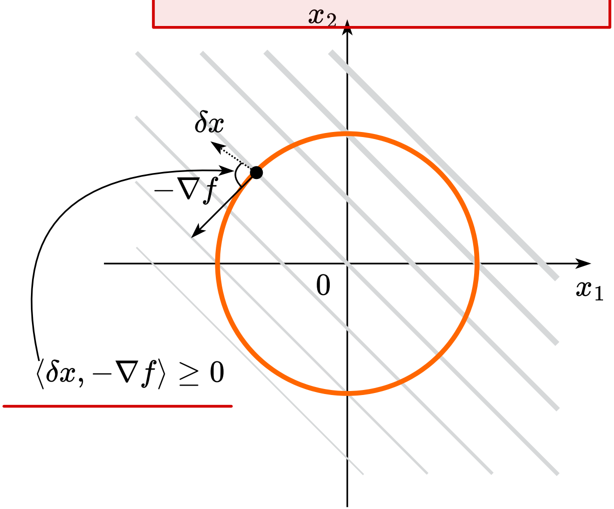


Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

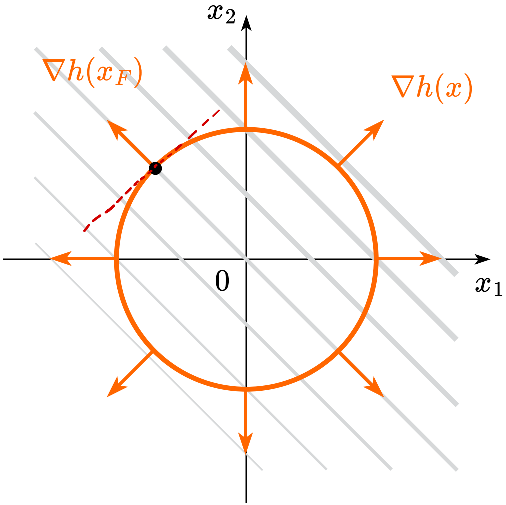
Мы хотим: $f(x_F + \delta x) \leq f(x_F)$



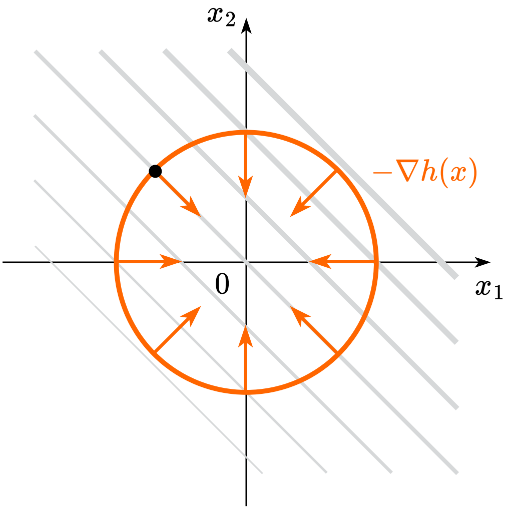
Задачи с ограничениями-равенствами

$h(x) = x_1^2 + x_2^2 - 2$

$$\nabla h = (2x_1, 2x_2)^T$$

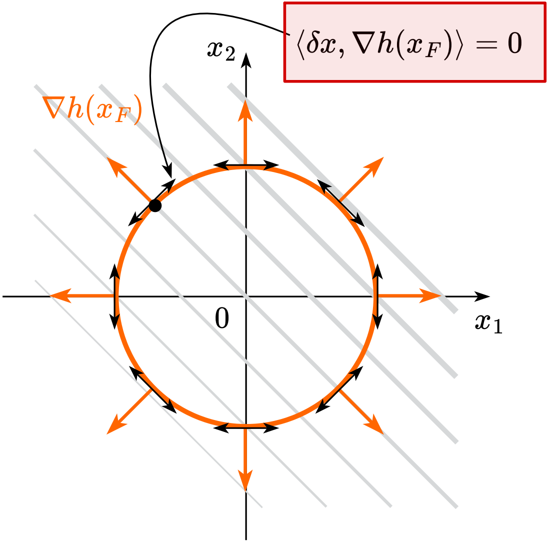


Задачи с ограничениями-равенствами



Задачи с ограничениями-равенствами

условие
нак.
в бюдж.
мн-ве



Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Предположим, что в процессе такого движения мы пришли в точку, где

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Предположим, что в процессе такого движения мы пришли в точку, где

$$-\nabla f(x) = \nu \nabla h(x)$$

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Предположим, что в процессе такого движения мы пришли в точку, где

$$-\nabla f(x) = \nu \nabla h(x)$$

$$\langle \delta x, -\nabla f(x) \rangle = \langle \delta x, \nu \nabla h(x) \rangle = 0$$

Задачи с ограничениями-равенствами

В общем случае, чтобы двигаться от x_F вдоль допустимого множества и уменьшать значение функции, необходимо обеспечить два условия:

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

Предположим, что в процессе такого движения мы пришли в точку, где

$$-\nabla f(x) = \nu \nabla h(x)$$

$$\langle \delta x, -\nabla f(x) \rangle = \langle \delta x, \nu \nabla h(x) \rangle = 0$$

Тогда мы достигли такой точки допустимого множества, из которой нельзя уменьшить значение функции при допустимых малых сдвигах. Это и есть условие локального минимума в задаче с ограничением.

Задачи с ограничениями-равенствами

$f: \mathbb{R}^2 \to \mathbb{R}$

$L: \mathbb{R}^3 \to \mathbb{R}$ функция Лагранжа

$L(x, \nu)$

ν „НЮ"

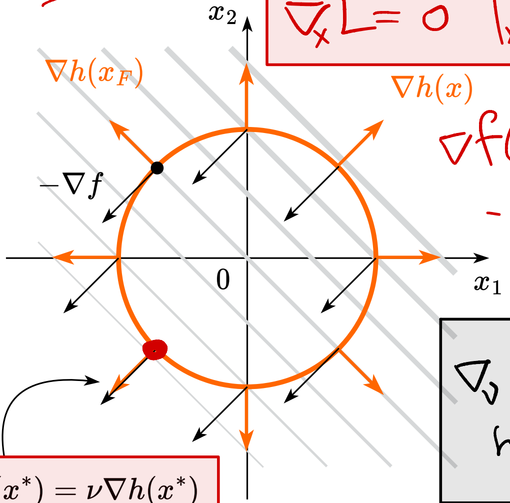
μ „МЮ"

$$L = f(x) + \nu h(x)$$

$$\boxed{\nabla_x L = 0 \Big|_{x = x^*}}$$

$$\nabla f(x^*) + \nu \cdot \nabla h(x^*) = 0$$

$$-\nabla f(x^*) = \nu \nabla h(x^*)$$



$$\boxed{-\nabla f(x^*) = \nu \nabla h(x^*)}$$

$$\boxed{\begin{gathered}\nabla_\nu L = 0 \\ h(x) = 0\end{gathered}}$$

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$\boxed{L(x, \nu) = f(x) + \nu h(x)}$$

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача *регулярная* (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача *регулярная* (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

$\min f(x)$
$x \in \mathbb{R}^n$
$h(x) = 0$

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача *регулярная* (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

Давайте определим лагранжиан (для удобства):

$h(x) = 0$

$x \in S$

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача *регулярная* (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

$\nabla_\nu L(x^*, \nu^*) = 0$ бюджетное ограничение

$\nabla L = 0$

всегда $h(x) = 0$

Важно отметить, что $\boxed{L(x^*, \nu^*) = f(x^*).}$

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача *регулярная* (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

$\nabla_\nu L(x^*, \nu^*) = 0$ бюджетное ограничение

Достаточные условия

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача *регулярная* (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

$\nabla_\nu L(x^*, \nu^*) = 0$ бюджетное ограничение

Достаточные условия

$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Лагранжиан

Давайте определим лагранжиан (для удобства):

$$L(x, \nu) = f(x) + \nu h(x)$$

Если задача *регулярная* (мы определим это понятие позже) и точка x^* является локальным минимумом для описанной выше задачи, то существует ν^*:

Необходимые условия

$\nabla_x L(x^*, \nu^*) = 0$ это мы уже написали выше

$\nabla_\nu L(x^*, \nu^*) = 0$ бюджетное ограничение

Достаточные условия

$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$

$\forall y \neq 0 \in \mathbb{R}^n : \nabla h(x^*)^T y = 0$

Важно отметить, что $L(x^*, \nu^*) = f(x^*)$.

Задачи с ограничениями-равенствами

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & h_i(x) = 0, \; i = 1, \ldots, p
\end{split}
\tag{ECP}
$$

$$
L(x, \nu) = f(x) + \sum_{i=1}^{p} \nu_i h_i(x) = f(x) + \nu^T h(x)
$$

Пусть $f(x)$ и $h_i(x)$ дважды дифференцируемы в точке x^* и непрерывно дифференцируемы в некоторой окрестности x^*. Условия локального минимума для $x \in \mathbb{R}^n, \nu \in \mathbb{R}^p$ записываются как

Необходимые условия

$$\nabla_x L(x^*, \nu^*) = 0$$

$$\nabla_\nu L(x^*, \nu^*) = 0$$

ECP: Достаточные условия

$$\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0,$$

$$\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^T y = 0$$

Задача наименьших квадратов

Example

Поставим задачу оптимизации и решим её для линейной системы $Ax = b$, $A \in \mathbb{R}^{m \times n}$, в трёх случаях (предполагая, что матрица имеет полный ранг):

- $m < n$

Задача наименьших квадратов

Example

Поставим задачу оптимизации и решим её для линейной системы $Ax = b$, $A \in \mathbb{R}^{m \times n}$, в трёх случаях (предполагая, что матрица имеет полный ранг):

- $m < n$
- $m = n$

Задача наименьших квадратов

Example

Поставим задачу оптимизации и решим её для линейной системы $Ax = b$, $A \in \mathbb{R}^{m \times n}$, в трёх случаях (предполагая, что матрица имеет полный ранг):

- $m < n$
- $m = n$
- $m > n$

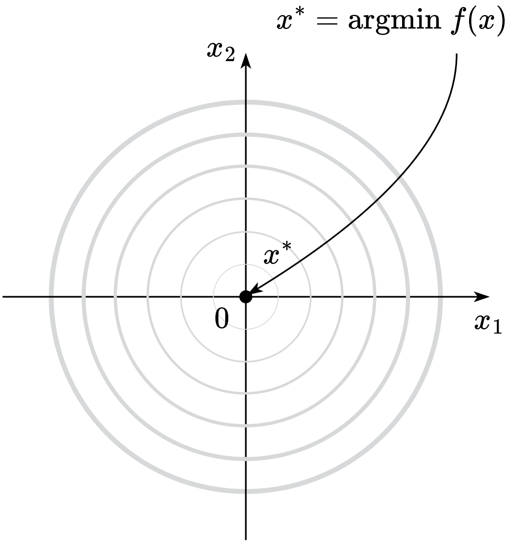
Задачи с ограничениями-неравенствами

Пример задачи с ограничениями-неравенствами

$$
f(x) = x_1^2 + x_2^2 \quad g(x) = x_1^2 + x_2^2 - 1
$$

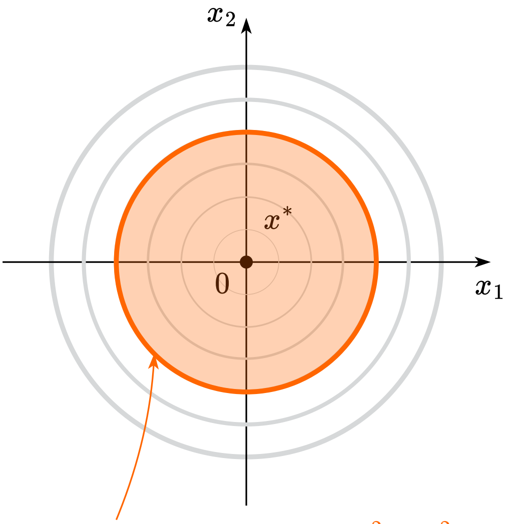
$$
\boxed{\begin{aligned} f(x) &\to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) &\leq 0 \end{aligned}}
$$

Задачи с ограничениями-неравенствами



Линии уровня $f(x) = x_1^2 + x_2^2 = C$

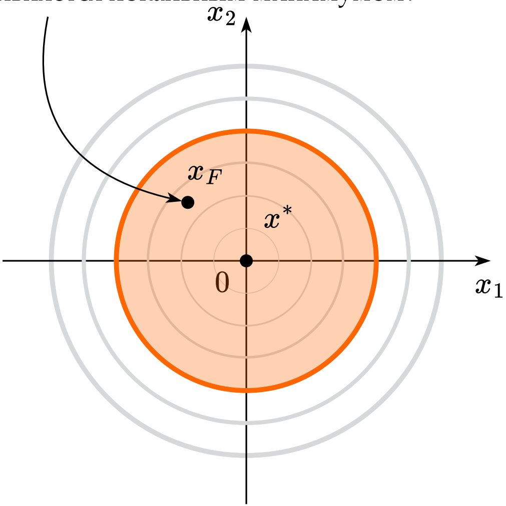
Задачи с ограничениями-неравенствами



Бюджетное множество $g(x) = x_1^2 + x_2^2 - 1 \leq 0$

Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая точка является локальным минимумом?



Задачи с ограничениями-неравенствами

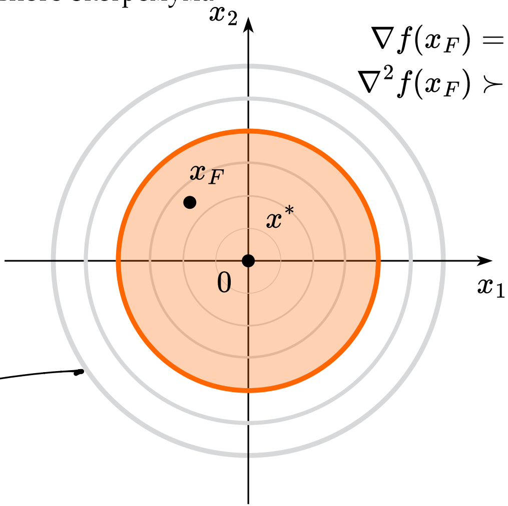
Просто! Проверим достаточные условия локального экстремума

$$\nabla f(x_F) = 0$$
$$\nabla^2 f(x_F) \succ 0$$

ограничение

$g(x) \leq 0$

НЕ АКТИВНО



Задачи с ограничениями-неравенствами

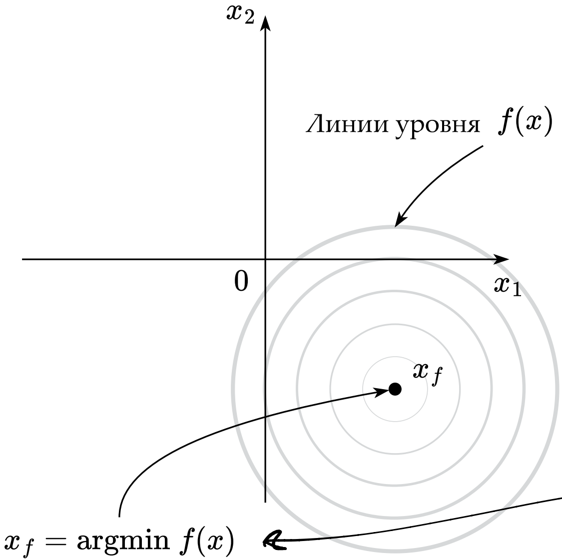
Таким образом, если ограничения типа неравенства неактивны в условной задаче, то мы можем решать задачу без ограничений. Однако так бывает не всегда. Рассмотрим второй простой пример.

$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

$$\begin{aligned} f(x) &\to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) &\leq 0 \end{aligned}$$

Задачи с ограничениями-неравенствами

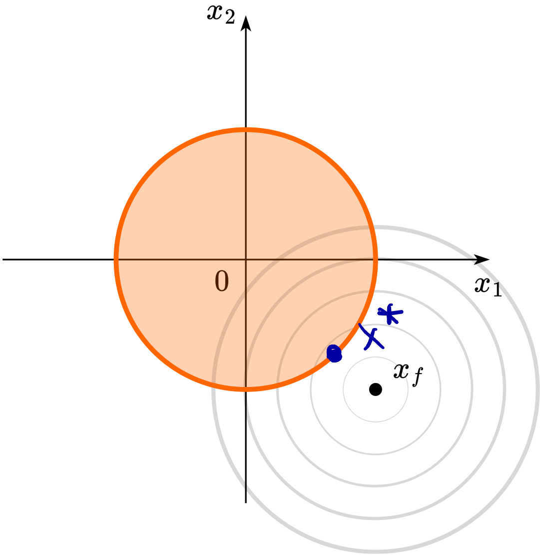
$$f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 = C$$



оптимум
безусловн
зад.

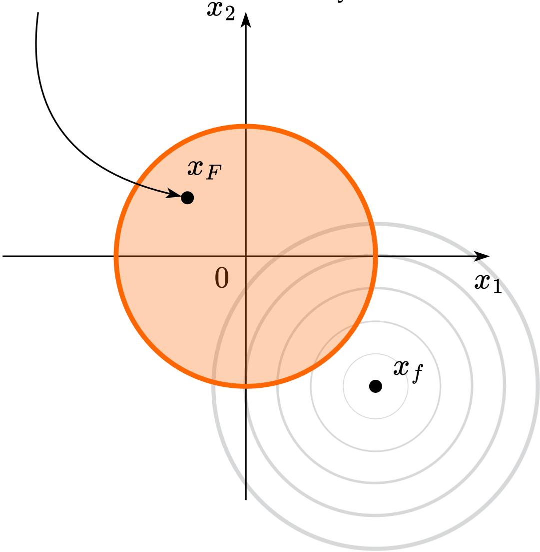
Задачи с ограничениями-неравенствами

Бюджетное множество $g(x) = x_1^2 + x_2^2 - 1 \leq 0$



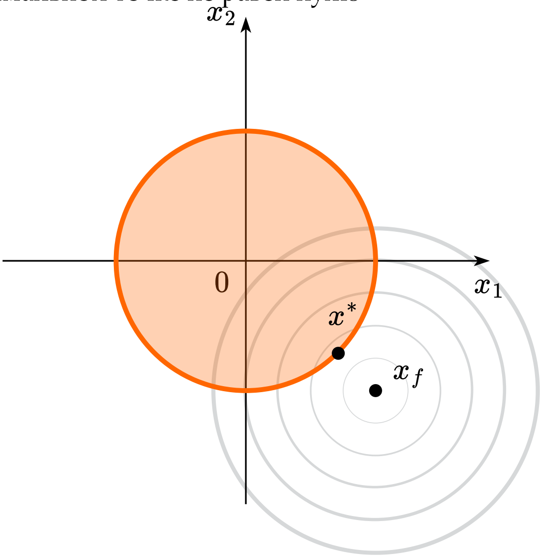
Задачи с ограничениями-неравенствами

Как понять, что некоторая допустимая точка является локальным минимумом?



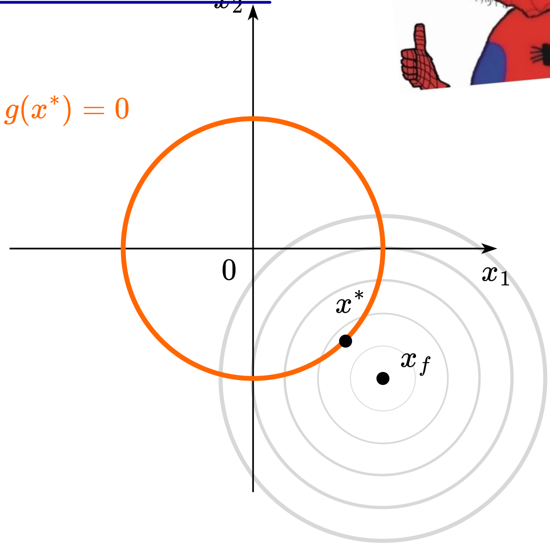
Задачи с ограничениями-неравенствами

Не так просто! Даже градиент
в оптимальной точке не равен нулю 😭

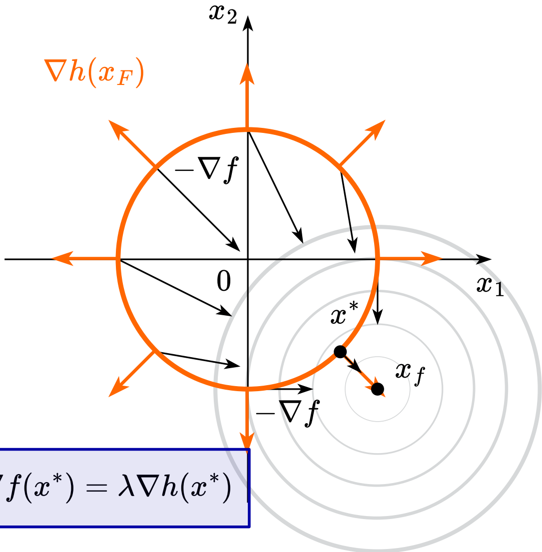


Задачи с ограничениями-неравенствами

Фактически имеем задачу
с ограничением-равенством 💡



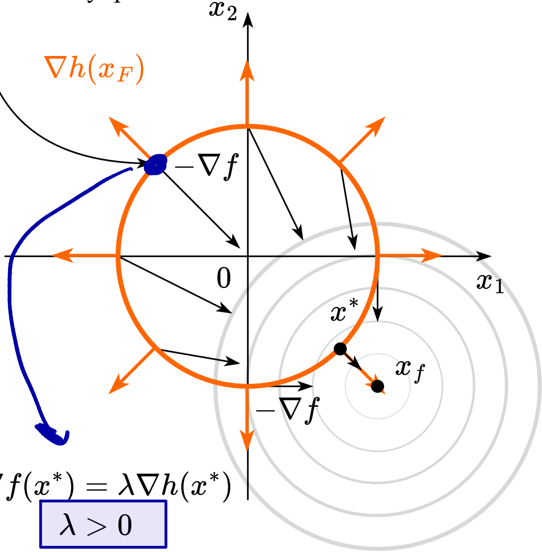
Задачи с ограничениями-неравенствами



$$\boxed{-\nabla f(x^*) = \lambda \nabla h(x^*)}$$

Задачи с ограничениями-неравенствами

Не является локальным минимумом, т.к. $-\nabla f(x)$ направлен внутрь бюджетного множества



$$-\nabla f(x^*) = \lambda \nabla h(x^*)$$

$$\boxed{\lambda > 0}$$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$\text{s.t. } g(x) \leq 0$$

Два возможных случая:

$\boxed{g(x) \leq 0 \text{ неактивно:}}$ $\underline{g(x^*) < 0}$

- $g(x^*) < 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$\text{s.t. } g(x) \leq 0$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

$g(x) \leq 0$ активно: $g(x^*) = 0$

- $g(x^*) = 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

$g(x) \leq 0$ активно: $g(x^*) = 0$

- $g(x^*) = 0$
- Необходимые условия: $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$

Задачи с ограничениями-неравенствами

Итак, у нас есть задача:

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

$$
\text{s.t. } g(x) \leq 0
$$

Два возможных случая:

$g(x) \leq 0$ неактивно: $g(x^*) < 0$

- $g(x^*) < 0$
- $\nabla f(x^*) = 0$
- $\nabla^2 f(x^*) \succ 0$

$g(x) \leq 0$ активно: $g(x^*) = 0$

- $g(x^*) = 0$
- Необходимые условия: $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$
- ~~Достаточные условия:~~ $\cancel{\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0, \forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^T y = 0}$

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

$$f(x) \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } g(x) \leq 0$$

Определим функцию Лагранжа:

$$L(x, \lambda) = f(x) + \lambda g(x)$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

$$\boxed{\begin{aligned} & f(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } & g(x) \leq 0 \end{aligned}}$$

λ – множители

λ – Лагранжа

Определим функцию Лагранжа:

$$\boxed{L(x, \lambda) = f(x) + \lambda g(x)}$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Если x^* является локальным минимумом для описанной выше задачи, то существует единственный множитель Лагранжа λ^* такой, что:

$$(1)\ \nabla_x L(x^*, \lambda^*) = 0$$

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

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

$$\text{s.t. } g(x) \leq 0$$

Определим функцию Лагранжа:

$$L(x, \lambda) = f(x) + \lambda g(x)$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Если x^* является локальным минимумом для описанной выше задачи, то существует единственный множитель Лагранжа λ^* такой, что:

$$(1)\ \nabla_x L(x^*, \lambda^*) = 0$$

$$(2)\ \lambda^* \geq 0$$

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Определим функцию Лагранжа:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Если x^* является локальным минимумом для описанной выше задачи, то существует единственный множитель Лагранжа λ^* такой, что:

$$
\begin{split}
& (1) \; \nabla_x L(x^*, \lambda^*) = 0 \\
& (2) \; \lambda^* \geq 0 \\
& (3) \; \lambda^* g(x^*) = 0
\end{split}
$$

Лагранжиан для задач с ограничениями-неравенствами

Объединяя два возможных случая, мы можем записать общие условия для задачи:

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

$$
\text{s.t. } g(x) \leq 0
$$

Определим функцию Лагранжа:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

Классические условия оптимальности первого и второго порядка Каруша-Куна-Таккера для локального минимума x^*, сформулированные при некоторых условиях регулярности, можно записать следующим образом.

Если x^* является локальным минимумом для описанной выше задачи, то существует единственный множитель Лагранжа λ^* такой, что:

$$
\begin{cases}
(1)\ \nabla_x L(x^*, \lambda^*) = 0 \\
(2)\ \lambda^* \geq 0 \\
(3)\ \lambda^* g(x^*) = 0 \\
(4)\ g(x^*) \leq 0
\end{cases}
$$

$\lambda^* = 0$

ограничение неактивно

$g(x^*) \leq 0$

$\lambda^* > 0$

ограничение активно

$g(x^*) = 0$

Общая формулировка

0

$$
\begin{aligned}
f_0(x) &\to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } f_i(x) &\le 0, \; i = 1, \ldots, m \\
h_i(x) &= 0, \; i = 1, \ldots, p
\end{aligned}
$$

Данная формулировка представляет собой общую задачу математического программирования.

Решение включает в себя построение лагранжиана:

$$
L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) =
$$

$$
= f_0(x) + \lambda^T f(x) + \nu^T h(x)
$$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$ $h_i(x) = 0$

$\cancel{\nabla_\lambda L = 0}$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda_i^* \geq 0, i = 1, \ldots, m$

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda_i^* \geq 0, i = 1, \ldots, m$
- $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$ ← условия дополняющей нежесткости

Необходимые условия

Пусть x^*, (λ^*, ν^*) является решением задачи математического программирования с нулевым двойственным зазором (оптимальное значение прямой задачи p^* равно оптимальному значению двойственной задачи d^*). Пусть также функции f_0, f_i, h_i дифференцируемы.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$ — огр. рав-ва
- $\lambda_i^* \geq 0, i = 1, \ldots, m$
- $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$ — огр. нер-ва
- $f_i(x^*) \leq 0, i = 1, \ldots, m$

Задача

$$Ax = b$$

$A \in \mathbb{R}^{m \times n}$

$m < n$

решений ∞ много

$$\frac{1}{2}\|x\|^2 \to \min$$

$Ax = b$ | m равенств

$h(x) = 0$

$$L = \frac{1}{2}x^T x + \lambda^T(Ax - b)$$

$$\frac{\partial L}{\partial x} = 0 \Rightarrow \begin{cases} x + A^T\lambda = 0 \\ Ax - b = 0 \end{cases}$$

$$\frac{\partial L}{\partial \lambda} = 0 \Rightarrow$$

$$x = -A^T\lambda$$

$$A(-A^T\lambda) - b = 0$$

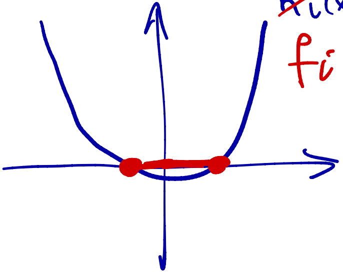
$$-AA^T\lambda = b \to \lambda = -(AA^T)^{-1} b$$

$$x = -A^T(-(AA^T)^{-1} b) = A^T(AA^T)^{-1} b$$

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, при наличии регулярности можно записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределённым* гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (при минимизации, с выпуклыми f_0, f_i и аффинными h_i) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то двойственный зазор равен нулю и условия Каруша-Куна-Таккера становятся необходимыми и достаточными.



~~$h_i(x)$~~ f_i

$\boxed{h_i(x) = 0}$

аффинное

$f_i(x) \leq 0$

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, при наличии регулярности можно записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределённым* гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (при минимизации, с выпуклыми f_0, f_i и аффинными h_i) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то двойственный зазор равен нулю и условия Каруша-Куна-Таккера становятся необходимыми и достаточными.
- **Условие линейной квалификации ограничений.** Если f_i и h_i являются аффинными функциями, то никаких других условий не требуется.

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, при наличии регулярности можно записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределённым* гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (при минимизации, с выпуклыми f_0, f_i и аффинными h_i) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то двойственный зазор равен нулю и условия Каруша-Куна-Таккера становятся необходимыми и достаточными.
- **Условие линейной квалификации ограничений.** Если f_i и h_i являются аффинными функциями, то никаких других условий не требуется.
- **Условие линейной независимости ограничений.** Градиенты активных ограничений неравенства и градиенты ограничений равенства линейно независимы в точке x^*.

Некоторые условия регулярности

Эти условия нужны для того, чтобы условия Каруша-Куна-Таккера стали необходимыми условиями. Некоторые из них даже превращают необходимые условия в достаточные (например, условие Слейтера). Кроме того, при наличии регулярности можно записать необходимые условия второго порядка $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ с *полуопределённым* гессианом лагранжиана.

- **Условие Слейтера.** Если для выпуклой задачи (при минимизации, с выпуклыми f_0, f_i и аффинными h_i) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то двойственный зазор равен нулю и условия Каруша-Куна-Таккера становятся необходимыми и достаточными.
- **Условие линейной квалификации ограничений.** Если f_i и h_i являются аффинными функциями, то никаких других условий не требуется.
- **Условие линейной независимости ограничений.** Градиенты активных ограничений неравенства и градиенты ограничений равенства линейно независимы в точке x^*.
- Для других примеров см. wiki.

Пример. Проекция на гиперплоскость

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Пример. Проекция на гиперплоскость

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.
$$

Решение

Лагранжиан:

Пример. Проекция на гиперплоскость

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Решение

Лагранжиан:

$$L(\mathbf{x}, \nu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 + \nu(\mathbf{a}^T\mathbf{x} - b)$$

Пример. Проекция на гиперплоскость

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Решение

Лагранжиан:

$$L(\mathbf{x}, \nu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 + \nu(\mathbf{a}^T\mathbf{x} - b)$$

Производная L по $\mathbf{x}$:

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu\mathbf{a}$$

Пример. Проекция на гиперплоскость

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.
$$

Решение

Лагранжиан:

$$
L(\mathbf{x}, \nu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 + \nu(\mathbf{a}^T\mathbf{x} - b)
$$

Производная L по $\mathbf{x}$:

$$
\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu\mathbf{a}
$$

$$
\mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{y} - \nu\mathbf{a}^T\mathbf{a} \qquad \nu = \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}
$$

Пример. Проекция на гиперплоскость

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{a}^T\mathbf{x} = b.$$

Решение

Лагранжиан:

$$L(\mathbf{x}, \nu) = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 + \nu(\mathbf{a}^T\mathbf{x} - b)$$

Производная L по $\mathbf{x}$:

$$\frac{\partial L}{\partial \mathbf{x}} = \mathbf{x} - \mathbf{y} + \nu\mathbf{a} = 0, \qquad \mathbf{x} = \mathbf{y} - \nu\mathbf{a}$$

$$\mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{y} - \nu\mathbf{a}^T\mathbf{a} \qquad \nu = \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}$$

$$\mathbf{x} = \mathbf{y} - \frac{\mathbf{a}^T\mathbf{y} - b}{\|\mathbf{a}\|^2}\mathbf{a}$$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1}=1, \quad \mathbf{x}\geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^T\mathbf{1}-1)
$$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^T\mathbf{1} - 1)
$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^T\mathbf{1} - 1)
$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$

Пример. Проекция на единичный симплекс

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0.$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^T\mathbf{1} - 1)$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$

Пример. Проекция на единичный симплекс

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0.$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^T\mathbf{1} - 1)$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $\mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0$

Пример. Проекция на единичный симплекс

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0.$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^T\mathbf{1} - 1)$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $\mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0$

Пример. Проекция на единичный симплекс

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1}=1, \quad \mathbf{x}\geq 0.$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^T\mathbf{1}-1)$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $\mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0$

Question

Решите систему выше за $O(n \log n)$.

Пример. Проекция на единичный симплекс

$$
\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2, \quad \text{s.t.} \quad \mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0.
$$

Условия ККТ

Лагранжиан задаётся следующим образом:

$$
L = \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|^2 - \sum_i \lambda_i x_i + \nu(\mathbf{x}^T\mathbf{1} - 1)
$$

Взяв производную L по x_i и записав ККТ, мы получаем:

- $\frac{\partial L}{\partial x_i} = x_i - y_i - \lambda_i + \nu = 0$
- $\lambda_i x_i = 0$
- $\lambda_i \geq 0$
- $\mathbf{x}^T\mathbf{1} = 1, \quad \mathbf{x} \geq 0$

Question

Решите систему выше за $O(n\log n)$.

Question

Решите систему выше за $O(n)$.

Ссылки (условная оптимизация)

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" @ KTH.

Ссылки (условная оптимизация)

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе “Элементы статистического обучения” @ KTH.
- Однострочное доказательство ККТ

Ссылки (условная оптимизация)

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе “Элементы статистического обучения” @ KTH.
- Однострочное доказательство ККТ
- О втором порядке оптимальности для задач оптимизации с ограничениями неравенства

Ссылки (условная оптимизация)

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" @ KTH.
- Однострочное доказательство ККТ
- О втором порядке оптимальности для задач оптимизации с ограничениями неравенства
- О втором порядке оптимальности в нелинейной оптимизации

Ссылки (условная оптимизация)

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе "Элементы статистического обучения" @ KTH.
- Однострочное доказательство ККТ
- О втором порядке оптимальности для задач оптимизации с ограничениями неравенства
- О втором порядке оптимальности в нелинейной оптимизации
- Численная оптимизация Jorge Nocedal, Stephen J. Wright.

Ссылки (условная оптимизация)

- Лекция по условиям ККТ (очень интуитивное объяснение) в курсе “Элементы статистического обучения” @ KTH.
- Однострочное доказательство ККТ
- О втором порядке оптимальности для задач оптимизации с ограничениями неравенства
- О втором порядке оптимальности в нелинейной оптимизации
- Численная оптимизация Jorge Nocedal, Stephen J. Wright.
- Лекция Duality Uses and Correspondences, Ryan Tibshirani, курс.

Двойственность

Мотивация

Двойственность позволяет связать любую задачу условной оптимизации с задачей **вогнутой максимизации**, решения которой дают **нижнюю границу** для оптимального значения исходной задачи.
Интересно то, что существуют случаи, когда можно решить **прямую задачу**, сначала решив **двойственную**.
Рассмотрим общую задачу оптимизации с ограничениями:

Мотивация

Двойственность позволяет связать любую задачу условной оптимизации с задачей **вогнутой максимизации**, решения которой дают **нижнюю границу** для оптимального значения исходной задачи.
Интересно то, что существуют случаи, когда можно решить **прямую задачу**, сначала решив **двойственную**.
Рассмотрим общую задачу оптимизации с ограничениями:

$$\boxed{\text{Primal: } f(x) \to \min_{x \in S}} \qquad \boxed{\text{Dual: } g(y) \to \max_{y \in \Omega}}$$

Мотивация

Двойственность позволяет связать любую задачу условной оптимизации с задачей **вогнутой максимизации**, решения которой дают **нижнюю границу** для оптимального значения исходной задачи.
Интересно то, что существуют случаи, когда можно решить **прямую задачу**, сначала решив **двойственную**.
Рассмотрим общую задачу оптимизации с ограничениями:

$$
\text{Primal: } f(x) \to \min_{x \in S} \qquad \text{Dual: } g(y) \to \max_{y \in \Omega}
$$

Строим функцию $g(y)$, удовлетворяющую **универсальному (равномерному) неравенству**:

$$
\boxed{g(y) \leq f(x) \qquad \forall x \in S, \forall y \in \Omega}
$$

Мотивация

Двойственность позволяет связать любую задачу условной оптимизации с задачей **вогнутой максимизации**, решения которой дают **нижнюю границу** для оптимального значения исходной задачи.
Интересно то, что существуют случаи, когда можно решить **прямую задачу**, сначала решив **двойственную**.
Рассмотрим общую задачу оптимизации с ограничениями:

$$\text{Primal: } f(x) \to \min_{x \in S} \qquad \text{Dual: } g(y) \to \max_{y \in \Omega}$$

Строим функцию $g(y)$, удовлетворяющую **универсальному (равномерному) неравенству**:

$$g(y) \leq f(x) \qquad \forall x \in S, \forall y \in \Omega$$

В результате получаем:

$$\boxed{\max_{y \in \Omega} g(y) \leq \min_{x \in S} f(x)}$$

Лагранжева двойственность

Рассмотрим один из возможных способов построения функции $g(y)$ в случае, когда у нас есть общая задача математического программирования с функциональными ограничениями:

Лагранжева двойственность

Рассмотрим один из возможных способов построения функции $g(y)$ в случае, когда у нас есть общая задача математического программирования с функциональными ограничениями:

$$
\begin{aligned}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m \\
& h_i(x) = 0, \; i = 1, \ldots, p
\end{aligned}
$$

Лагранжева двойственность

Рассмотрим один из возможных способов построения функции $g(y)$ в случае, когда у нас есть общая задача математического программирования с функциональными ограничениями:

$$
\begin{aligned}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \; i = 1, \ldots, m \\
& h_i(x) = 0, \; i = 1, \ldots, p
\end{aligned}
$$

$$p^* = f_0(x^*)$$

И лагранжиан, соответствующий этой задаче:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) = f_0(x) + \lambda^T f(x) + \nu^T h(x)$$

Двойственная функция

Пусть $\mathcal{D} = \bigcap\limits_{i=0}^{m} \mathbf{dom}\ f_i \cap \bigcap\limits_{i=1}^{p} \mathbf{dom}\ h_i$ — непустое множество.

Определим **лагранжеву двойственную функцию** (или просто **двойственную функцию**) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ как инфимум лагранжиана по x, для $\lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$

Двойственная функция

инфимум

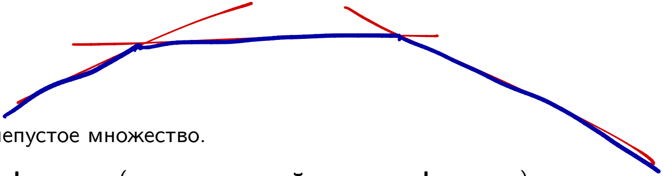
Пусть $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\ f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\ h_i$ — непустое множество.

Определим **лагранжеву двойственную функцию** (или просто **двойственную функцию**) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ как инфимум лагранжиана по x, для $\lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$

опр.

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Двойственная функция



Пусть $\mathcal{D} = \bigcap_{i=0}^{m} \textbf{dom } f_i \cap \bigcap_{i=1}^{p} \textbf{dom } h_i$ — непустое множество.

Определим **лагранжеву двойственную функцию** (или просто **двойственную функцию**) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ как инфимум лагранжиана по x, для $\lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Если лагранжиан не ограничен снизу по x, то двойственная функция принимает значение $-\infty$.
Так как двойственная функция является **поточечным инфимумом семейства аффинных функций** от (λ, ν), она является **вогнутой**, даже если исходная задача не является выпуклой.

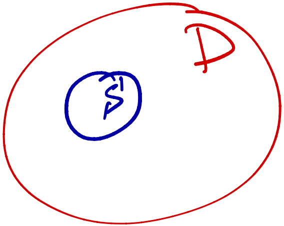
Двойственная функция как нижняя граница

Покажем, что двойственная функция даёт **нижнюю границу** для оптимального значения p^* исходной задачи при любых $\lambda \succeq 0, \nu$.

$p^* \geq g(\lambda, \nu)$

Пусть некоторая точка $\hat{x}$ является допустимой для исходной задачи, то есть $f_i(\hat{x}) \leq 0$ и $h_i(\hat{x}) = 0$. Тогда при $\lambda \succeq 0$ имеем:

$\hat{x} \in S \in D$



Двойственная функция как нижняя граница

Покажем, что двойственная функция даёт **нижнюю границу** для оптимального значения p^* исходной задачи при любых $\lambda \succeq 0, \nu$.

Пусть некоторая точка $\hat{x}$ является допустимой для исходной задачи, то есть $f_i(\hat{x}) \leq 0$ и $h_i(\hat{x}) = 0$. Тогда при $\lambda \succeq 0$ имеем:

≥ 0 ≤ 0

$$
\underline{L(\hat{x}, \lambda, \nu)} = f_0(\hat{x}) + \underbrace{\lambda^T f(\hat{x})}_{\leq 0} + \underbrace{\nu^T h(\hat{x})}_{=0} \leq f_0(\hat{x})
$$

Двойственная функция как нижняя граница

Покажем, что двойственная функция даёт **нижнюю границу** для оптимального значения p^* исходной задачи при любых $\lambda \succeq 0, \nu$.

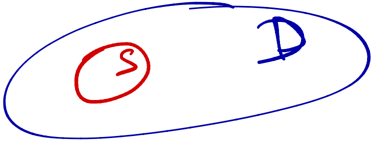
Пусть некоторая точка $\hat{x}$ является допустимой для исходной задачи, то есть $f_i(\hat{x}) \leq 0$ и $h_i(\hat{x}) = 0$. Тогда при $\lambda \succeq 0$ имеем:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^T f(\hat{x})}_{\leq 0} + \underbrace{\nu^T h(\hat{x})}_{=0} \leq f_0(\hat{x})$$

Следовательно,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})$$

Двойственная функция как нижняя граница



Покажем, что двойственная функция даёт **нижнюю границу** для оптимального значения p^* исходной задачи при любых $\lambda \succeq 0, \nu$.
Пусть некоторая точка $\hat{x}$ является допустимой для исходной задачи, то есть $f_i(\hat{x}) \leq 0$ и $h_i(\hat{x}) = 0$. Тогда при $\lambda \succeq 0$ имеем:

$\forall \hat{x} \in S$

$$
L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^T f(\hat{x})}_{\leq 0} + \underbrace{\nu^T h(\hat{x})}_{= 0} \leq f_0(\hat{x})
$$

в том числе и для $\hat{x} = x^*$

Следовательно,

$$
g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})
$$

$\rightarrow g(\lambda, \nu) \leq f_0(\hat{x})$

$$
\boxed{g(\lambda, \nu) \leq p^*}
$$

Двойственная функция как нижняя граница

Покажем, что двойственная функция даёт **нижнюю границу** для оптимального значения p^* исходной задачи при любых $\lambda \succeq 0, \nu$.
Пусть некоторая точка $\hat{x}$ является допустимой для исходной задачи, то есть $f_i(\hat{x}) \leq 0$ и $h_i(\hat{x}) = 0$. Тогда при $\lambda \succeq 0$ имеем:

$$
L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^T f(\hat{x})}_{\leq 0} + \underbrace{\nu^T h(\hat{x})}_{=0} \leq f_0(\hat{x})
$$

Следовательно,

$$
g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})
$$

$$
g(\lambda, \nu) \leq p^*
$$

Возникает естественный вопрос: **какую наилучшую нижнюю границу** можно получить из лагранжевой двойственной функции?
Это приводит нас к следующей задаче оптимизации:

Двойственная функция как нижняя граница

Покажем, что двойственная функция даёт **нижнюю границу** для оптимального значения p^* исходной задачи при любых $\lambda \succeq 0, \nu$.

Пусть некоторая точка $\hat{x}$ является допустимой для исходной задачи, то есть $f_i(\hat{x}) \leq 0$ и $h_i(\hat{x}) = 0$. Тогда при $\lambda \succeq 0$ имеем:

$$L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^T f(\hat{x})}_{\leq 0} + \underbrace{\nu^T h(\hat{x})}_{=0} \leq f_0(\hat{x})$$

Следовательно,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})$$

$$g(\lambda, \nu) \leq p^*$$

Возникает естественный вопрос: **какую наилучшую нижнюю границу** можно получить из лагранжевой двойственной функции?

Это приводит нас к следующей задаче оптимизации:

$$g(\lambda, \nu) \to \max_{\lambda \in \mathbb{R}^m, \, \nu \in \mathbb{R}^p}$$

$$\text{s.t. } \lambda \succeq 0$$

$d^* = g(\lambda^*, \nu^*)$

Двойственная функция как нижняя граница

Покажем, что двойственная функция даёт **нижнюю границу** для оптимального значения p^* исходной задачи при любых $\lambda \succeq 0, \nu$.

Пусть некоторая точка $\hat{x}$ является допустимой для исходной задачи, то есть $f_i(\hat{x}) \leq 0$ и $h_i(\hat{x}) = 0$. Тогда при $\lambda \succeq 0$ имеем:

$$
L(\hat{x}, \lambda, \nu) = f_0(\hat{x}) + \underbrace{\lambda^T f(\hat{x})}_{\leq 0} + \underbrace{\nu^T h(\hat{x})}_{=0} \leq f_0(\hat{x})
$$

Следовательно,

$$
g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\hat{x}, \lambda, \nu) \leq f_0(\hat{x})
$$

$$
g(\lambda, \nu) \leq p^*
$$

Возникает естественный вопрос: **какую наилучшую нижнюю границу** можно получить из лагранжевой двойственной функции?

Это приводит нас к следующей задаче оптимизации:

$$
\begin{aligned}
g(\lambda, \nu) &\to \max_{\lambda \in \mathbb{R}^m, \, \nu \in \mathbb{R}^p} \\
\text{s.t. } \lambda &\succeq 0
\end{aligned}
$$

Теперь становится понятным термин **«допустимая для двойственной задачи» (dual feasible)**, обозначающий пару (λ, ν), для которой $\lambda \succeq 0$ и $g(\lambda, \nu) > -\infty$.

Это означает, как следует из названия, что (λ, ν) является допустимым решением для двойственной задачи.

Если пара (λ^*, ν^*) оптимальна для этой задачи, то её называют **оптимальными лагранжевыми множителями** или **оптимальными двойственными переменными**.

Резюме

	Прямая задача	Двойственная задача
Функция	$f_0(x)$	$g(\lambda, \nu) = \inf\limits_{x \in \mathcal{D}} L(x, \lambda, \nu)$
Переменные	$x \in S \subseteq \mathbb{R}^n$	$\lambda \in \mathbb{R}^m_+, \nu \in \mathbb{R}^p$
Ограничения	$f_i(x) \leq 0,\ i = 1, \ldots, m;\ h_i(x) = 0,$ $i = 1, \ldots, p$	$\lambda_i \geq 0, i = 1, \ldots, m$
Задача	$\begin{aligned} & f_0(x) \to \min_{x \in \mathbb{R}^n} \\ \text{s.t. } & f_i(x) \leq 0,\ i = 1, \ldots, m \\ & h_i(x) = 0,\ i = 1, \ldots, p \end{aligned}$	$\begin{aligned} & g(\lambda, \nu) \to \max_{\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p} \\ \text{s.t. } & \lambda \succeq 0 \end{aligned}$
Оптимум	x^* если допустимо, $p^* = f_0(x^*)$	λ^*, ν^* если максимум достигается, $d^* = g(\lambda^*, \nu^*)$

Пример. Линейный МНК

Мы рассматриваем задачу в пределах непустого множества допустимых решений, определяемого следующим образом:

Пример. Линейный МНК

Мы рассматриваем задачу в пределах непустого множества допустимых решений, определяемого следующим образом:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

где матрица $A \in \mathbb{R}^{m \times n}$.

Пример. Линейный МНК

Мы рассматриваем задачу в пределах непустого множества допустимых решений, определяемого следующим образом:

$$
\begin{aligned}
&\text{min} \quad x^T x \\
&\text{s.t.} \quad Ax = b,
\end{aligned}
$$

где матрица $A \in \mathbb{R}^{m \times n}$.

В данной задаче отсутствуют неравенства — присутствуют только m **линейных равенств**.
Лагранжиан имеет вид $L(x, \nu) = x^T x + \nu^T (Ax - b)$ и определён на области $\mathbb{R}^n \times \mathbb{R}^m$.
Двойственная функция обозначается как $g(\nu) = \inf_x L(x, \nu)$.
Так как $L(x, \nu)$ является **выпуклой квадратичной функцией** по x, минимизирующее значение x можно найти из **условия оптимальности**

Пример. Линейный МНК

Мы рассматриваем задачу в пределах непустого множества допустимых решений, определяемого следующим образом:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

где матрица $A \in \mathbb{R}^{m \times n}$.

В данной задаче отсутствуют неравенства — присутствуют только m **линейных равенств**.
Лагранжиан имеет вид $L(x, \nu) = x^T x + \nu^T (Ax - b)$ и определён на области $\mathbb{R}^n \times \mathbb{R}^m$.
Двойственная функция обозначается как $g(\nu) = \inf_x L(x, \nu)$.
Так как $L(x, \nu)$ является **выпуклой квадратичной функцией** по x, минимизирующее значение x можно найти из **условия оптимальности**

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

Пример. Линейный МНК

Мы рассматриваем задачу в пределах непустого множества допустимых решений, определяемого следующим образом:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

где матрица $A \in \mathbb{R}^{m \times n}$.

В данной задаче отсутствуют неравенства — присутствуют только m **линейных равенств**.
Лагранжиан имеет вид $L(x, \nu) = x^T x + \nu^T (Ax - b)$ и определён на области $\mathbb{R}^n \times \mathbb{R}^m$.
Двойственная функция обозначается как $g(\nu) = \inf_x L(x, \nu)$.
Так как $L(x, \nu)$ является **выпуклой квадратичной функцией** по x, минимизирующее значение x можно найти из **условия оптимальности**

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

откуда следует $x = -(1/2)A^T \nu$.
В результате двойственная функция имеет вид:

Пример. Линейный МНК

Мы рассматриваем задачу в пределах непустого множества допустимых решений, определяемого следующим образом:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

где матрица $A \in \mathbb{R}^{m \times n}$.

В данной задаче отсутствуют неравенства — присутствуют только m **линейных равенств**.
Лагранжиан имеет вид $L(x, \nu) = x^T x + \nu^T (Ax - b)$ и определён на области $\mathbb{R}^n \times \mathbb{R}^m$.
Двойственная функция обозначается как $g(\nu) = \inf_x L(x, \nu)$.
Так как $L(x, \nu)$ является **выпуклой квадратичной функцией** по x, минимизирующее значение x можно найти из **условия оптимальности**

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

откуда следует $x = -(1/2)A^T \nu$.
В результате двойственная функция имеет вид:

$$
g(\nu) = L(-(1/2)A^T \nu, \nu) = -(1/4)\nu^T A A^T \nu - b^T \nu,
$$

Пример. Линейный МНК

Мы рассматриваем задачу в пределах непустого множества допустимых решений, определяемого следующим образом:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

где матрица $A \in \mathbb{R}^{m \times n}$.

В данной задаче отсутствуют неравенства — присутствуют только m **линейных равенств**.
Лагранжиан имеет вид $L(x, \nu) = x^T x + \nu^T (Ax - b)$ и определён на области $\mathbb{R}^n \times \mathbb{R}^m$.
Двойственная функция обозначается как $g(\nu) = \inf_x L(x, \nu)$.
Так как $L(x, \nu)$ является **выпуклой квадратичной функцией** по x, минимизирующее значение x можно найти из **условия оптимальности**

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

откуда следует $x = -(1/2) A^T \nu$.
В результате двойственная функция имеет вид:

$$
g(\nu) = L(-(1/2) A^T \nu, \nu) = -(1/4) \nu^T A A^T \nu - b^T \nu,
$$

что представляет собой **вогнутую квадратичную функцию** на области $\mathbb{R}^m$. Согласно свойству нижней границы, для любого $\nu \in \mathbb{R}^m$ выполняется неравенство:

Пример. Линейный МНК

Мы рассматриваем задачу в пределах непустого множества допустимых решений, определяемого следующим образом:

$$
\begin{aligned}
\text{min} \quad & x^T x \\
\text{s.t.} \quad & Ax = b,
\end{aligned}
$$

где матрица $A \in \mathbb{R}^{m \times n}$.

В данной задаче отсутствуют неравенства — присутствуют только m **линейных равенств**.
Лагранжиан имеет вид $L(x, \nu) = x^T x + \nu^T (Ax - b)$ и определён на области $\mathbb{R}^n \times \mathbb{R}^m$.
Двойственная функция обозначается как $g(\nu) = \inf_x L(x, \nu)$.
Так как $L(x, \nu)$ является **выпуклой квадратичной функцией** по x, минимизирующее значение x можно найти из **условия оптимальности**

$$
\nabla_x L(x, \nu) = 2x + A^T \nu = 0,
$$

откуда следует $x = -(1/2)A^T \nu$.
В результате двойственная функция имеет вид:

$$
g(\nu) = L(-(1/2)A^T \nu, \nu) = -(1/4)\nu^T A A^T \nu - b^T \nu,
$$

что представляет собой **вогнутую квадратичную функцию** на области $\mathbb{R}^m$. Согласно свойству нижней границы, для любого $\nu \in \mathbb{R}^m$ выполняется неравенство:

$$
-(1/4)\nu^T A A^T \nu - b^T \nu \leq \inf\{x^T x \mid Ax = b\}.
$$

Это простая, но нетривиальная **нижняя граница** без необходимости решать саму задачу.

Пример. Задача двустороннего разбиения

Мы рассматриваем (невыпуклую) задачу:

$$
\begin{aligned}
\text{min} \quad & x^T W x \\
\text{s.t.} \quad & x_i^2 = 1, \quad i = 1, \ldots, n,
\end{aligned}
$$

Пример. Задача двустороннего разбиения

Мы рассматриваем (невыпуклую) задачу:

$$
\begin{aligned}
\min \quad & x^T W x \\
\text{s.t.} \quad & x_i^2 = 1, \quad i = 1, \ldots, n,
\end{aligned}
$$

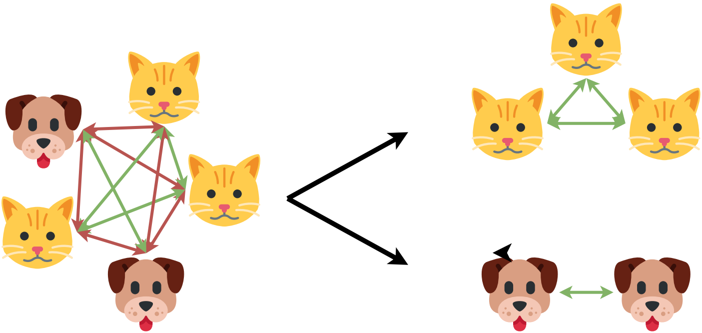


Figure 23: Иллюстрация задачи двустороннего разбиения

Пример. Задача двустороннего разбиения

Мы рассматриваем (невыпуклую) задачу:

$$\begin{aligned} \min \quad & x^T W x \\ \text{s.t.} \quad & x_i^2 = 1, \quad i = 1, \ldots, n, \end{aligned}$$

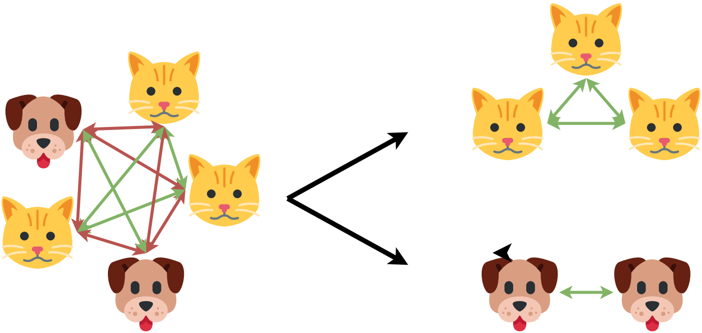


Figure 23: Иллюстрация задачи двустороннего разбиения

Эту задачу можно рассматривать как задачу двустороннего разбиения множества из n элементов, обозначаемого как $\{1, \ldots, n\}$: допустимый вектор x соответствует разбиению

$$\{1, \ldots, n\} = \{i \mid x_i = -1\} \cup \{i \mid x_i = 1\}.$$

Пример. Задача двустороннего разбиения

Мы рассматриваем (невыпуклую) задачу:

$$
\begin{aligned}
\min \quad & x^T W x \\
\text{s.t.} \quad & x_i^2 = 1, \quad i = 1, \ldots, n,
\end{aligned}
$$

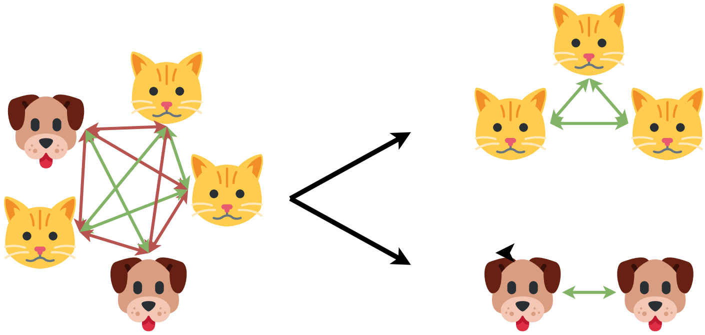


Figure 23: Иллюстрация задачи двустороннего разбиения

Эту задачу можно рассматривать как задачу двустороннего разбиения множества из n элементов, обозначаемого как $\{1, \ldots, n\}$: допустимый вектор x соответствует разбиению

$$
\{1, \ldots, n\} = \{i \mid x_i = -1\} \cup \{i \mid x_i = 1\}.
$$

Коэффициент W_{ij} в матрице представляет собой стоимость помещения элементов i и j в одну и ту же группу, тогда как $-W_{ij}$ означает цену их разделения. Целевая функция отражает суммарную стоимость по всем парам элементов, и поставленная задача состоит в том, чтобы найти разбиение, минимизирующее общие затраты.

Пример. Задача двустороннего разбиения

Теперь выведем двойственную функцию для этой задачи. Лагранжиан имеет вид

$$L(x, \nu) = x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) = x^T (W + \mathrm{diag}(\nu)) x - \mathbf{1}^T \nu.$$

Пример. Задача двустороннего разбиения

Теперь выведем двойственную функцию для этой задачи. Лагранжиан имеет вид

$$L(x, \nu) = x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu.$$

Минимизируя по x, получаем двойственную функцию Лагранжа:

$$g(\nu) = \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{если } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{в противном случае,} \end{cases}$$

Пример. Задача двустороннего разбиения

Теперь выведем двойственную функцию для этой задачи. Лагранжиан имеет вид

$$L(x,\nu) = x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu.$$

Минимизируя по x, получаем двойственную функцию Лагранжа:

$$g(\nu) = \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{если } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{в противном случае,} \end{cases}$$

Используется тот факт, что инфимум квадратичной формы равен нулю, если форма положительно полуопределённая, и равен $-\infty$, если нет.

Пример. Задача двустороннего разбиения

Теперь выведем двойственную функцию для этой задачи. Лагранжиан имеет вид

$$
L(x, \nu) = x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu.
$$

Минимизируя по x, получаем двойственную функцию Лагранжа:

$$
g(\nu) = \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{если } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{в противном случае,} \end{cases}
$$

Используется тот факт, что инфимум квадратичной формы равен нулю, если форма положительно полуопределённая, и равен $-\infty$, если нет.

Эта двойственная функция задаёт нижние границы на оптимальное значение задачи. Например, можно выбрать конкретное значение двойственной переменной:

$$
\nu = -\lambda_{\text{min}}(W) \mathbf{1}
$$

Пример. Задача двустороннего разбиения

Теперь выведем двойственную функцию для этой задачи. Лагранжиан имеет вид

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu.$$

Минимизируя по x, получаем двойственную функцию Лагранжа:

$$g(\nu) = \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{если } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{в противном случае,} \end{cases}$$

Используется тот факт, что инфимум квадратичной формы равен нулю, если форма положительно полуопределённая, и равен $-\infty$, если нет.

Эта двойственная функция задаёт нижние границы на оптимальное значение задачи. Например, можно выбрать конкретное значение двойственной переменной:

$$\nu = -\lambda_{\min}(W) \mathbf{1}$$

Это значение допустимо в двойственной задаче, так как $W + \text{diag}(\nu) = W - \lambda_{\min}(W) I \succeq 0$.

Пример. Задача двустороннего разбиения

Теперь выведем двойственную функцию для этой задачи. Лагранжиан имеет вид

$$L(x,\nu) = x^T W x + \sum_{i=1}^n \nu_i(x_i^2 - 1) = x^T(W + \text{diag}(\nu))x - \mathbf{1}^T\nu.$$

Минимизируя по x, получаем двойственную функцию Лагранжа:

$$g(\nu) = \inf_x x^T(W + \text{diag}(\nu))x - \mathbf{1}^T\nu = \begin{cases} -\mathbf{1}^T\nu & \text{если } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{в противном случае,} \end{cases}$$

Используется тот факт, что инфимум квадратичной формы равен нулю, если форма положительно полуопределённая, и равен $-\infty$, если нет.

Эта двойственная функция задаёт нижние границы на оптимальное значение задачи. Например, можно выбрать конкретное значение двойственной переменной:

$$\nu = -\lambda_{\text{min}}(W)\mathbf{1}$$

Это значение допустимо в двойственной задаче, так как $W + \text{diag}(\nu) = W - \lambda_{\text{min}}(W)I \succeq 0$.

Отсюда получается простая нижняя граница для оптимального значения p^*: $p^* \geq -\mathbf{1}^T\nu = n\lambda_{\text{min}}(W)$.

Пример. Задача двустороннего разбиения

Теперь выведем двойственную функцию для этой задачи. Лагранжиан имеет вид

$$L(x,\nu) = x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) = x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu.$$

Минимизируя по x, получаем двойственную функцию Лагранжа:

$$g(\nu) = \inf_x x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu = \begin{cases} -\mathbf{1}^T \nu & \text{если } W + \text{diag}(\nu) \succeq 0 \\ -\infty & \text{в противном случае,} \end{cases}$$

Используется тот факт, что инфимум квадратичной формы равен нулю, если форма положительно полуопределённая, и равен $-\infty$, если нет.

Эта двойственная функция задаёт нижние границы на оптимальное значение задачи. Например, можно выбрать конкретное значение двойственной переменной:

$$\nu = -\lambda_{\text{min}}(W)\mathbf{1}$$

Это значение допустимо в двойственной задаче, так как $W + \text{diag}(\nu) = W - \lambda_{\text{min}}(W) I \succeq 0$.

Отсюда получается простая нижняя граница для оптимального значения p^*: $p^* \geq -\mathbf{1}^T \nu = n\lambda_{\text{min}}(W)$.

Код для этой задачи доступен здесь: Open in Colab

Сильная двойственность

Сильная двойственность

Неравенство между оптимальными значениями прямой и двойственной задач обычно называется **слабой двойственностью**. Для задачи выполняется:

$$
p^* \geq d^*
$$

Сильная двойственность

Неравенство между оптимальными значениями прямой и двойственной задач обычно называется **слабой двойственностью**. Для задачи выполняется:

$$
p^* \geq d^*
$$

Разность между ними часто называют **зазором двойственности** (*duality gap*):

$$
p^* - d^* \geq 0
$$

Сильная двойственность

Неравенство между оптимальными значениями прямой и двойственной задач обычно называется **слабой двойственностью**. Для задачи выполняется:

$$p^* \geq d^*$$

Разность между ними часто называют **зазором двойственности** (*duality gap*):

$$p^* - d^* \geq 0$$

Заметьте, что слабая двойственность выполняется всегда, если прямая и двойственная задачи корректно сформулированы. Это означает, что если нам удалось решить двойственную задачу (двойственная функция которой всегда вогнута, независимо от того, была ли исходная задача выпуклой), то мы получаем некоторую нижнюю границу. Удивительно, но существуют случаи, когда эти решения совпадают.

Сильная двойственность

Неравенство между оптимальными значениями прямой и двойственной задач обычно называется **слабой двойственностью**. Для задачи выполняется:

$$
p^* \geq d^*
$$

Разность между ними часто называют **зазором двойственности** (*duality gap*):

$$
p^* - d^* \geq 0
$$

Заметьте, что слабая двойственность выполняется всегда, если прямая и двойственная задачи корректно сформулированы. Это означает, что если нам удалось решить двойственную задачу (двойственная функция которой всегда вогнута, независимо от того, была ли исходная задача выпуклой), то мы получаем некоторую нижнюю границу. Удивительно, но существуют случаи, когда эти решения совпадают.

Сильная двойственность имеет место, если зазор двойственности равен нулю: $p^* = d^*$.

Сильная двойственность

Неравенство между оптимальными значениями прямой и двойственной задач обычно называется **слабой двойственностью**. Для задачи выполняется:

$$p^* \geq d^*$$

Разность между ними часто называют **зазором двойственности** (*duality gap*):

$$p^* - d^* \geq 0$$

Заметьте, что слабая двойственность выполняется всегда, если прямая и двойственная задачи корректно сформулированы. Это означает, что если нам удалось решить двойственную задачу (двойственная функция которой всегда вогнута, независимо от того, была ли исходная задача выпуклой), то мы получаем некоторую нижнюю границу. Удивительно, но существуют случаи, когда эти решения совпадают.

Сильная двойственность имеет место, если зазор двойственности равен нулю: $p^* = d^*$.

Замечание: как p^*, так и d^* могут быть равны ∞.

- Известно несколько достаточных условий!

Сильная двойственность

Неравенство между оптимальными значениями прямой и двойственной задач обычно называется **слабой двойственностью**. Для задачи выполняется:

$$
p^* \geq d^*
$$

Разность между ними часто называют **зазором двойственности** (*duality gap*):

$$
p^* - d^* \geq 0
$$

Заметьте, что слабая двойственность выполняется всегда, если прямая и двойственная задачи корректно сформулированы. Это означает, что если нам удалось решить двойственную задачу (двойственная функция которой всегда вогнута, независимо от того, была ли исходная задача выпуклой), то мы получаем некоторую нижнюю границу. Удивительно, но существуют случаи, когда эти решения совпадают.

Сильная двойственность имеет место, если зазор двойственности равен нулю: $p^* = d^*$.

Замечание: как p^*, так и d^* могут быть равны ∞.

- Известно несколько достаточных условий!
- "Простые" необходимые и достаточные условия — неизвестны.

Сильная двойственность в задаче линейных наименьших квадратов

Exercise

В примере с решением системы линейных уравнений методом наименьших квадратов выше вычислите оптимум прямой задачи p^* и оптимум двойственной задачи d^* и проверьте, выполняется ли для этой задачи сильная двойственность.

Полезные свойства двойственности

- **Построение нижней границы для решения прямой задачи.**

  Иногда исходную задачу решить крайне сложно. Но если у нас есть двойственная задача, можно взять произвольное $y \in \Omega$ и подставить его в $g(y)$ — мы сразу получим некоторую нижнюю границу.

Полезные свойства двойственности

- **Построение нижней границы для решения прямой задачи.**

  Иногда исходную задачу решить крайне сложно. Но если у нас есть двойственная задача, можно взять произвольное $y \in \Omega$ и подставить его в $g(y)$ — мы сразу получим некоторую нижнюю границу.

- **Проверка разрешимости задачи и достижимости решения.**

  Из неравенства $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ следует: если $\min\limits_{x \in S} f_0(x) = -\infty$, то $\Omega = \emptyset$, и наоборот.

Полезные свойства двойственности

- **Построение нижней границы для решения прямой задачи.**

  Иногда исходную задачу решить крайне сложно. Но если у нас есть двойственная задача, можно взять произвольное $y \in \Omega$ и подставить его в $g(y)$ — мы сразу получим некоторую нижнюю границу.

- **Проверка разрешимости задачи и достижимости решения.**

  Из неравенства $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ следует: если $\min\limits_{x \in S} f_0(x) = -\infty$, то $\Omega = \emptyset$, и наоборот.

- **Иногда проще решить двойственную задачу, чем прямую.**

  В этом случае, если выполняется сильная двойственность $g(y^*) = f_0(x^*)$, мы ничего не теряем.

Полезные свойства двойственности

- **Построение нижней границы для решения прямой задачи.**

  Иногда исходную задачу решить крайне сложно. Но если у нас есть двойственная задача, можно взять произвольное $y \in \Omega$ и подставить его в $g(y)$ — мы сразу получим некоторую нижнюю границу.

- **Проверка разрешимости задачи и достижимости решения.**

  Из неравенства $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ следует: если $\min\limits_{x \in S} f_0(x) = -\infty$, то $\Omega = \emptyset$, и наоборот.

- **Иногда проще решить двойственную задачу, чем прямую.**

  В этом случае, если выполняется сильная двойственность $g(y^*) = f_0(x^*)$, мы ничего не теряем.

- **Получение нижней границы на остаток функции.**

  $f_0(x) - f_0^* \leq f_0(x) - g(y)$ для произвольного $y \in \Omega$ (сертификат субоптимальности). Более того, $p^* \in [g(y), f_0(x)],\ d^* \in [g(y), f_0(x)]$

Полезные свойства двойственности

- **Построение нижней границы для решения прямой задачи.**

  Иногда исходную задачу решить крайне сложно. Но если у нас есть двойственная задача, можно взять произвольное $y \in \Omega$ и подставить его в $g(y)$ — мы сразу получим некоторую нижнюю границу.

- **Проверка разрешимости задачи и достижимости решения.**

  Из неравенства $\max\limits_{y \in \Omega} g(y) \leq \min\limits_{x \in S} f_0(x)$ следует: если $\min\limits_{x \in S} f_0(x) = -\infty$, то $\Omega = \emptyset$, и наоборот.

- **Иногда проще решить двойственную задачу, чем прямую.**

  В этом случае, если выполняется сильная двойственность $g(y^*) = f_0(x^*)$, мы ничего не теряем.

- **Получение нижней границы на остаток функции.**

  $f_0(x) - f_0^* \leq f_0(x) - g(y)$ для произвольного $y \in \Omega$ (сертификат субоптимальности). Более того, $p^* \in [g(y), f_0(x)], \;\; d^* \in [g(y), f_0(x)]$

- **Двойственная функция всегда вогнута**

  Так как она является поточечным инфимумом аффинных функций.

Условие Слейтера

Theorem

Если для выпуклой задачи оптимизации (предполагаем минимизацию, функции f_0, f_i выпуклые, а h_i аффинные) существует точка x такая, что $h(x) = 0$ и $f_i(x) < 0$ (существует строго допустимая точка), то зазор двойственности равен нулю, а условия ККТ становятся необходимыми и достаточными.

Пример выпуклой задачи, когда условие Слейтера не выполняется

Example

$$
\min\{f_0(x) = x \mid f_1(x) = \frac{x^2}{2} \leq 0\},
$$

Пример выпуклой задачи, когда условие Слейтера не выполняется

Example

$$
\min\{f_0(x) = x \mid f_1(x) = \frac{x^2}{2} \leq 0\},
$$

Единственная точка допустимого множества: $x^* = 0$. Однако невозможно найти неотрицательное $\lambda^* \geq 0$ такое, что

$$
\nabla f_0(0) + \lambda^* \nabla f_1(0) = 1 + \lambda^* \cdot 0 = 1 \neq 0.
$$

Невыпуклая квадратичная задача с сильной двойственностью

Иногда сильная двойственность выполняется даже для невыпуклой задачи. В качестве важного примера рассмотрим задачу минимизации невыпуклой квадратичной функции на единичном шаре:

Невыпуклая квадратичная задача с сильной двойственностью

Иногда сильная двойственность выполняется даже для невыпуклой задачи. В качестве важного примера рассмотрим задачу минимизации невыпуклой квадратичной функции на единичном шаре:

$$
\begin{aligned}
& x^T A x + 2b^T x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^T x \leq 1
\end{aligned}
$$

Невыпуклая квадратичная задача с сильной двойственностью

Иногда сильная двойственность выполняется даже для невыпуклой задачи. В качестве важного примера рассмотрим задачу минимизации невыпуклой квадратичной функции на единичном шаре:

$$
\begin{split}
& x^T A x + 2b^T x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^T x \leq 1
\end{split}
$$

где $A \in \mathbb{S}^n$, $A \nsucceq 0$ и $b \in \mathbb{R}^n$. Поскольку $A \nsucceq 0$, задача невыпуклая. Эту задачу иногда называют *trust region problem* — задачей доверительной области; она возникает при минимизации квадратичной аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Невыпуклая квадратичная задача с сильной двойственностью

Иногда сильная двойственность выполняется даже для невыпуклой задачи. В качестве важного примера рассмотрим задачу минимизации невыпуклой квадратичной функции на единичном шаре:

$$
\begin{aligned}
& x^T A x + 2b^T x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^T x \leq 1
\end{aligned}
$$

где $A \in \mathbb{S}^n$, $A \nsucceq 0$ и $b \in \mathbb{R}^n$. Поскольку $A \nsucceq 0$, задача невыпуклая. Эту задачу иногда называют *trust region problem* — задачей доверительной области; она возникает при минимизации квадратичной аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Решение

Невыпуклая квадратичная задача с сильной двойственностью

Иногда сильная двойственность выполняется даже для невыпуклой задачи. В качестве важного примера рассмотрим задачу минимизации невыпуклой квадратичной функции на единичном шаре:

$$
\begin{aligned}
& x^T A x + 2b^T x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^T x \leq 1
\end{aligned}
$$

где $A \in \mathbb{S}^n$, $A \nsucceq 0$ и $b \in \mathbb{R}^n$. Поскольку $A \nsucceq 0$, задача невыпуклая. Эту задачу иногда называют *trust region problem* — задачей доверительной области; она возникает при минимизации квадратичной аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Решение

Лагранжиан и двойственная функция:

$$
L(x, \lambda) = x^T A x + 2b^T x + \lambda(x^T x - 1) = x^T (A + \lambda I) x + 2b^T x - \lambda
$$

Невыпуклая квадратичная задача с сильной двойственностью

Иногда сильная двойственность выполняется даже для невыпуклой задачи. В качестве важного примера рассмотрим задачу минимизации невыпуклой квадратичной функции на единичном шаре:

$$
\begin{aligned}
&x^T A x + 2b^T x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } &x^T x \leq 1
\end{aligned}
$$

где $A \in \mathbb{S}^n$, $A \nsucceq 0$ и $b \in \mathbb{R}^n$. Поскольку $A \nsucceq 0$, задача невыпуклая. Эту задачу иногда называют *trust region problem* — задачей доверительной области; она возникает при минимизации квадратичной аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Решение

Лагранжиан и двойственная функция:

$$
L(x, \lambda) = x^T A x + 2b^T x + \lambda(x^T x - 1) = x^T (A + \lambda I) x + 2b^T x - \lambda
$$

$$
g(\lambda) = \begin{cases} -b^T (A + \lambda I)^{\dagger} b - \lambda, & \text{если } A + \lambda I \succeq 0 \\ -\infty, & \text{иначе} \end{cases}
$$

Невыпуклая квадратичная задача с сильной двойственностью

Иногда сильная двойственность выполняется даже для невыпуклой задачи. В качестве важного примера рассмотрим задачу минимизации невыпуклой квадратичной функции на единичном шаре:

$$
\begin{aligned}
& x^T A x + 2b^T x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^T x \leq 1
\end{aligned}
$$

где $A \in \mathbb{S}^n, \ A \nsucceq 0$ и $b \in \mathbb{R}^n$. Поскольку $A \nsucceq 0$, задача невыпуклая. Эту задачу иногда называют *trust region problem* — задачей доверительной области; она возникает при минимизации квадратичной аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Решение

Лагранжиан и двойственная функция:

$$
L(x, \lambda) = x^T A x + 2b^T x + \lambda(x^T x - 1) = x^T (A + \lambda I) x + 2b^T x - \lambda
$$

$$
g(\lambda) = \begin{cases} -b^T (A + \lambda I)^\dagger b - \lambda, & \text{если } A + \lambda I \succeq 0 \\ -\infty, & \text{иначе} \end{cases}
$$

Двойственная задача:

$$
\begin{aligned}
& -b^T (A + \lambda I)^\dagger b - \lambda \to \max_{\lambda \in \mathbb{R}} \\
\text{s.t. } & A + \lambda I \succeq 0
\end{aligned}
$$

Невыпуклая квадратичная задача с сильной двойственностью

Иногда сильная двойственность выполняется даже для невыпуклой задачи. В качестве важного примера рассмотрим задачу минимизации невыпуклой квадратичной функции на единичном шаре:

$$
\begin{aligned}
& x^T A x + 2b^T x \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & x^T x \leq 1
\end{aligned}
$$

где $A \in \mathbb{S}^n$, $A \nsucceq 0$ и $b \in \mathbb{R}^n$. Поскольку $A \nsucceq 0$, задача невыпуклая. Эту задачу иногда называют *trust region problem* — задачей доверительной области; она возникает при минимизации квадратичной аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Решение

Лагранжиан и двойственная функция:

$$
L(x, \lambda) = x^T A x + 2b^T x + \lambda(x^T x - 1) = x^T (A + \lambda I) x + 2b^T x - \lambda
$$

$$
g(\lambda) = \begin{cases} -b^T (A + \lambda I)^\dagger b - \lambda, & \text{если } A + \lambda I \succeq 0 \\ -\infty, & \text{иначе} \end{cases}
$$

Двойственная задача:

$$
\begin{aligned}
& -b^T (A + \lambda I)^\dagger b - \lambda \to \max_{\lambda \in \mathbb{R}} \\
\text{s.t. } & A + \lambda I \succeq 0
\end{aligned}
$$

$$
\begin{aligned}
& -\sum_{i=1}^n \frac{(q_i^T b)^2}{\lambda_i + \lambda} - \lambda \to \max_{\lambda \in \mathbb{R}} \\
\text{s.t. } & \lambda \geq -\lambda_{\min}(A)
\end{aligned}
$$

Приложения

Решение прямой задачи через двойственную

Важное следствие условия стационарности: при выполнении сильной двойственности, если известны решения двойственной задачи λ^*, ν^*, то x^* является также решением задачи

$$\min_{x \in \mathbb{R}^n} f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x)$$

Часто решение этой задачи можно записать **в явном виде**, что позволяет выразить x^* через λ^*, ν^*.

Более того, если решение этой задачи единственно, то оно и есть решение прямой задачи x^*.

Это бывает очень полезно, когда двойственную задачу решить проще, чем прямую.

Решение прямой задачи через двойственную

Рассмотрим, например:

$$
\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b
$$

Решение прямой задачи через двойственную

Рассмотрим, например:

$$
\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b
$$

где каждая $f_i(x_i) = \frac{1}{2}c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$
g(\nu) = \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x)
$$

Решение прямой задачи через двойственную

Рассмотрим, например:

$$
\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b
$$

где каждая $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$
\begin{aligned}
g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\
&= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\}
\end{aligned}
$$

Решение прямой задачи через двойственную

Рассмотрим, например:

$$
\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b
$$

где каждая $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$
\begin{aligned}
g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\
&= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\} \\
&= b\nu - \sum_{i=1}^n f_i^*(a_i \nu),
\end{aligned}
$$

Решение прямой задачи через двойственную

Рассмотрим, например:

$$
\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b
$$

где каждая $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$
\begin{aligned}
g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\
&= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\} \\
&= b\nu - \sum_{i=1}^n f_i^*(a_i \nu),
\end{aligned}
$$

Решение прямой задачи через двойственную

Рассмотрим, например:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b$$

где каждая $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

где каждая $f_i^*(y) = \frac{1}{2c_i} y^2$ — это сопряжённая функция к f_i.

Решение прямой задачи через двойственную

Рассмотрим, например:

$$
\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b
$$

где каждая $f_i(x_i) = \frac{1}{2}c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$
\begin{aligned}
g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\
&= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\} \\
&= b\nu - \sum_{i=1}^n f_i^*(a_i \nu),
\end{aligned}
$$

где каждая $f_i^*(y) = \frac{1}{2c_i} y^2$ — это сопряжённая функция к f_i.

Следовательно, двойственная задача имеет вид:

$$
\max_\nu b\nu - \sum_{i=1}^n f_i^*(a_i \nu) \quad \Longleftrightarrow \quad \min_\nu \sum_{i=1}^n f_i^*(a_i \nu) - b\nu
$$

Решение прямой задачи через двойственную

Рассмотрим, например:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b$$

где каждая $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

где каждая $f_i^*(y) = \frac{1}{2c_i} y^2$ — это сопряжённая функция к f_i.

Следовательно, двойственная задача имеет вид:

$$\max_\nu b\nu - \sum_{i=1}^n f_i^*(a_i \nu) \quad \Longleftrightarrow \quad \min_\nu \sum_{i=1}^n f_i^*(a_i \nu) - b\nu$$

Это задача выпуклой минимизации с одной скалярной переменной — значительно проще, чем прямая.

Решение прямой задачи через двойственную

Рассмотрим, например:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b$$

где каждая $f_i(x_i) = \frac{1}{2}c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

где каждая $f_i^*(y) = \frac{1}{2c_i} y^2$ — это сопряжённая функция к f_i.

Следовательно, двойственная задача имеет вид:

$$\max_\nu b\nu - \sum_{i=1}^n f_i^*(a_i \nu) \quad \Longleftrightarrow \quad \min_\nu \sum_{i=1}^n f_i^*(a_i \nu) - b\nu$$

Это задача выпуклой минимизации с одной скалярной переменной — значительно проще, чем прямая. Если известно ν^*, то решение прямой задачи x^* удовлетворяет:

$$\min_x \sum_{i=1}^n \left(f_i(x_i) - a_i \nu^* x_i\right)$$

Решение прямой задачи через двойственную

Рассмотрим, например:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b$$

где каждая $f_i(x_i) = \frac{1}{2} c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

где каждая $f_i^*(y) = \frac{1}{2c_i} y^2$ — это сопряжённая функция к f_i.

Следовательно, двойственная задача имеет вид:

$$\max_\nu b\nu - \sum_{i=1}^n f_i^*(a_i \nu) \quad \Longleftrightarrow \quad \min_\nu \sum_{i=1}^n f_i^*(a_i \nu) - b\nu$$

Это задача выпуклой минимизации с одной скалярной переменной — значительно проще, чем прямая. Если известно ν^*, то решение прямой задачи x^* удовлетворяет:

$$\min_x \sum_{i=1}^n (f_i(x_i) - a_i \nu^* x_i)$$

Строгая выпуклость каждой функции f_i гарантирует, что решение единственно, и его можно найти из условия $f_i'(x_i) = a_i \nu^*$ для каждого i.

Решение прямой задачи через двойственную

Рассмотрим, например:

$$\min_x \sum_{i=1}^n f_i(x_i) \quad \text{s.t.} \quad a^T x = b$$

где каждая $f_i(x_i) = \frac{1}{2}c_i x_i^2$ (гладкая и строго выпуклая). Двойственная функция:

$$\begin{aligned} g(\nu) &= \min_x \sum_{i=1}^n f_i(x_i) + \nu(b - a^T x) \\ &= b\nu + \sum_{i=1}^n \min_{x_i} \{f_i(x_i) - a_i \nu x_i\} \\ &= b\nu - \sum_{i=1}^n f_i^*(a_i \nu), \end{aligned}$$

где каждая $f_i^*(y) = \frac{1}{2c_i} y^2$ — это сопряжённая функция к f_i.

Следовательно, двойственная задача имеет вид:

$$\max_\nu b\nu - \sum_{i=1}^n f_i^*(a_i \nu) \quad \Longleftrightarrow \quad \min_\nu \sum_{i=1}^n f_i^*(a_i \nu) - b\nu$$

Это задача выпуклой минимизации с одной скалярной переменной — значительно проще, чем прямая. Если известно ν^*, то решение прямой задачи x^* удовлетворяет:

$$\min_x \sum_{i=1}^n \left(f_i(x_i) - a_i \nu^* x_i\right)$$

Строгая выпуклость каждой функции f_i гарантирует, что решение единственно, и его можно найти из условия $f_i'(x_i) = a_i \nu^*$ для каждого i. Отсюда получаем:

$$x_i^* = \frac{a_i \nu^*}{c_i}.$$

Смешанные стратегии для матричных игр

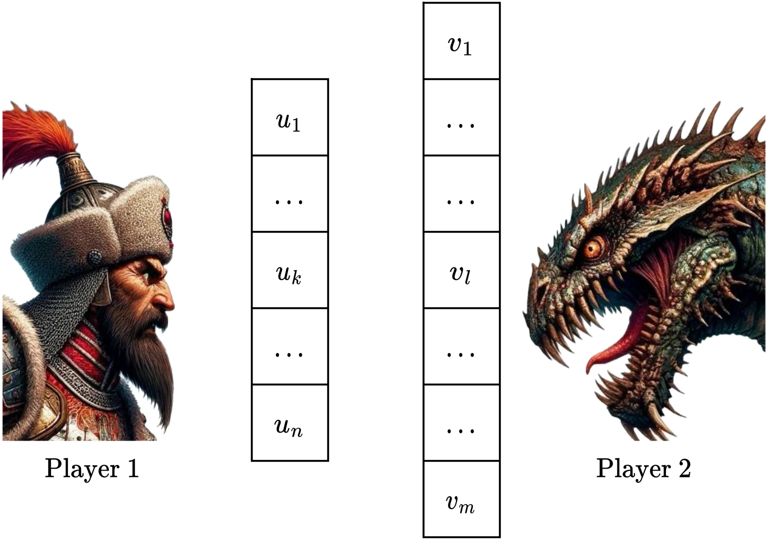


Figure 24: Схема смешанной стратегии в матричной игре

Смешанные стратегии для матричных игр

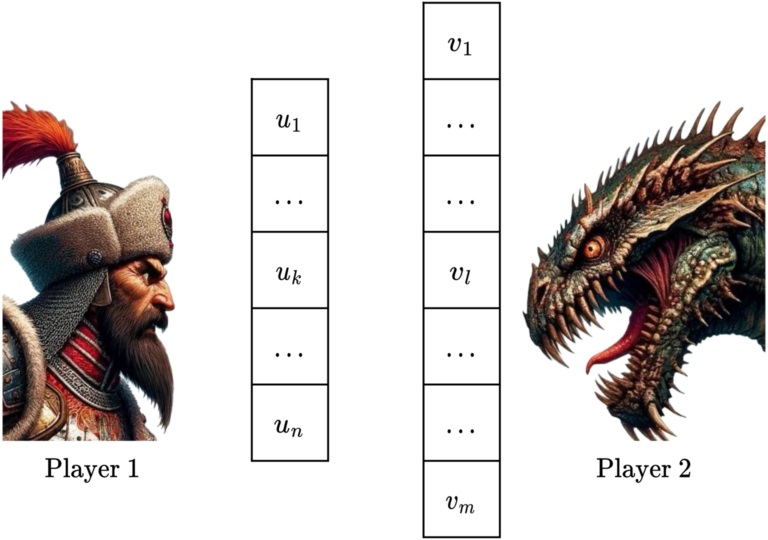


Figure 24: Схема смешанной стратегии в матричной игре

В матричной игре с нулевой суммой (zero-sum) игроки 1 и 2 выбирают действия из множеств $\{1, \dots, n\}$ и $\{1, \dots, m\}$ соответственно.

Результатом является выплата от игрока 1 игроку 2, определяемая платежной матрицей $P \in \mathbb{R}^{n \times m}$.

Каждый игрок использует **смешанные стратегии**, то есть выбирает действия согласно вероятностному распределению: игрок 1 — с вероятностями u_k для каждого действия k, а игрок 2 — с вероятностями v_l для каждого действия l.

Смешанные стратегии для матричных игр

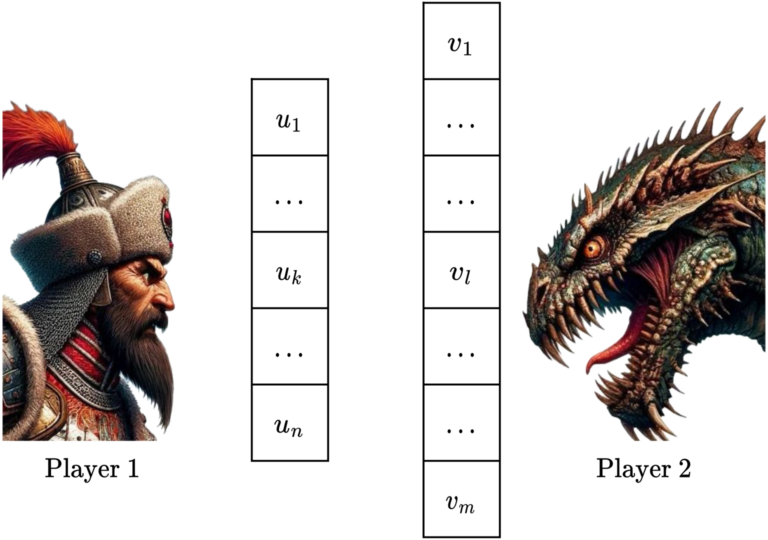


Figure 24: Схема смешанной стратегии в матричной игре

В матричной игре с нулевой суммой (zero-sum) игроки 1 и 2 выбирают действия из множеств $\{1, \dots, n\}$ и $\{1, \dots, m\}$ соответственно. Результатом является выплата от игрока 1 игроку 2, определяемая платежной матрицей $P \in \mathbb{R}^{n \times m}$. Каждый игрок использует **смешанные стратегии**, то есть выбирает действия согласно вероятностному распределению: игрок 1 — с вероятностями u_k для каждого действия k, а игрок 2 — с вероятностями v_l для каждого действия l. Математическое ожидание выигрыша (выплаты от игрока 1 игроку 2) задаётся выражением $\sum_{k=1}^{n} \sum_{l=1}^{m} u_k v_l P_{kl} = u^T P v$. Игрок 1 стремится **минимизировать** это ожидаемое значение, а игрок 2 — **максимизировать**

Смешанные стратегии для матричных игр. Перспектива игрока 1

Player 1

u_1
$\dots$
u_k
$\dots$
u_n

Пусть игрок 2 знает стратегию игрока 1 u. Тогда он выберет v, максимизирующее u^TPv. В наихудшем случае для игрока 1 ожидаемый выигрыш равен:

$$
\max_{v \geq 0, \mathbf{1}^T v = 1} u^T P v = \max_{i=1,\dots,m} (P^T u)_i
$$

Смешанные стратегии для матричных игр. Перспектива игрока 1

Player 1

$$\begin{array}{|c|}\hline u_1 \\ \hline \cdots \\ \hline u_k \\ \hline \cdots \\ \hline u_n \\ \hline \end{array}$$

Пусть игрок 2 знает стратегию игрока 1 u. Тогда он выберет v, максимизирующее u^TPv. В наихудшем случае для игрока 1 ожидаемый выигрыш равен:

$$\max_{v \geq 0, \mathbf{1}^T v = 1} u^T P v = \max_{i=1,\dots,m} (P^T u)_i$$

Оптимальная стратегия игрока 1 минимизирует это наихудшее значение, что приводит к следующей задаче оптимизации:

$$\begin{aligned} &\min \max_{i=1,\dots,m} (P^T u)_i \\ &\text{s.t. } u \geq 0 \\ &\mathbf{1}^T u = 1 \end{aligned} \tag{1}$$

Это выпуклая задача оптимизации, оптимальное значение которой обозначается как p_1^*.

Смешанные стратегии для матричных игр. Перспектива игрока 2

$$\begin{array}{|c|}\hline v_1 \\ \hline \dots \\ \hline \dots \\ \hline v_l \\ \hline \dots \\ \hline \dots \\ \hline v_m \\ \hline \end{array}$$

Player 2

Аналогично, если игрок 1 знает стратегию игрока 2 v, его цель — минимизировать $u^T P v$. Это приводит к задаче:

$$\min_{u \geq 0, \mathbf{1}^T u = 1} u^T P v = \min_{i=1,\dots,n} (Pv)_i$$

Смешанные стратегии для матричных игр. Перспектива игрока 2

$$\begin{array}{|c|}\hline v_1 \\ \hline \dots \\ \hline \dots \\ \hline v_l \\ \hline \dots \\ \hline \dots \\ \hline v_m \\ \hline \end{array}$$

Player 2

Аналогично, если игрок 1 знает стратегию игрока 2 v, его цель — минимизировать $u^T P v$. Это приводит к задаче:

$$\min_{u \geq 0, \mathbf{1}^T u = 1} u^T P v = \min_{i=1,\ldots,n} (Pv)_i$$

Игрок 2 затем максимизирует это значение, чтобы получить наибольшую гарантированную выплату, решая следующую задачу оптимизации:

$$\begin{aligned} &\max \min_{i=1,\ldots,n} (Pv)_i \\ &\text{s.t. } v \geq 0 \\ &\mathbf{1}^T v = 1 \end{aligned} \tag{2}$$

Оптимальное значение здесь обозначается как p_2^*.

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,
2. Сумма элементов вектора u равна 1 ($\mathbf{1}^T u = 1$),

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,
2. Сумма элементов вектора u равна 1 ($\mathbf{1}^T u = 1$),
3. $P^T u \leq t\mathbf{1}$ (вектор не превосходит t поэлементно).

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,
2. Сумма элементов вектора u равна 1 ($\mathbf{1}^T u = 1$),
3. $P^T u \leq t\mathbf{1}$ (вектор не превосходит t поэлементно).

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,
2. Сумма элементов вектора u равна 1 ($\mathbf{1}^T u = 1$),
3. $P^T u \leq t\mathbf{1}$ (вектор не превосходит t поэлементно).

Здесь t — дополнительная переменная, принимающая значения на множестве действительных чисел ($t \in \mathbb{R}$).

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,
2. Сумма элементов вектора u равна 1 ($\mathbf{1}^T u = 1$),
3. $P^T u \leq t\mathbf{1}$ (вектор не превосходит t поэлементно).

Здесь t — дополнительная переменная, принимающая значения на множестве действительных чисел ($t \in \mathbb{R}$).

Построение лагранжиана

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,
2. Сумма элементов вектора u равна 1 ($\mathbf{1}^T u = 1$),
3. $P^T u \leq t\mathbf{1}$ (вектор не превосходит t поэлементно).

Здесь t — дополнительная переменная, принимающая значения на множестве действительных чисел ($t \in \mathbb{R}$).

Построение лагранжиана

Введём множители для ограничений: λ — для $P^T u \leq t\mathbf{1}$, μ — для $u \geq 0$, и ν — для $\mathbf{1}^T u = 1$. Тогда лагранжиан имеет вид:

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,
2. Сумма элементов вектора u равна 1 ($\mathbf{1}^T u = 1$),
3. $P^T u \leq t\mathbf{1}$ (вектор не превосходит t поэлементно).

Здесь t — дополнительная переменная, принимающая значения на множестве действительных чисел ($t \in \mathbb{R}$).

Построение лагранжиана

Введём множители для ограничений: λ — для $P^T u \leq t\mathbf{1}$, μ — для $u \geq 0$, и ν — для $\mathbf{1}^T u = 1$. Тогда лагранжиан имеет вид:

$$L = t + \lambda^T (P^T u - t\mathbf{1}) - \mu^T u + \nu(1 - \mathbf{1}^T u) = \nu + (1 - \mathbf{1}^T \lambda)t + (P\lambda - \nu\mathbf{1} - \mu)^T u$$

Смешанные стратегии для матричных игр

Двойственность и эквивалентность

Обычно знание стратегии противника даёт преимущество, однако, что удивительно, в случае смешанных стратегий для матричных игр это преимущество исчезает. Ключ заключается в двойственности: приведённые выше задачи являются двойственными по Лагранжу. Если записать задачу игрока 1 как задачу линейного программирования и ввести множители Лагранжа, то двойственная задача совпадёт с задачей игрока 2. Благодаря сильной двойственности, выполняющейся для допустимых задач линейного программирования, имеем $p_1^* = p_2^*$, что показывает отсутствие преимущества в знании стратегии противника.

Постановка и решение двойственной задачи Лагранжа

Рассмотрим задачу (1), записав её в виде задачи линейного программирования (LP). Цель — минимизировать переменную t при следующих ограничениях:

1. $u \geq 0$,
2. Сумма элементов вектора u равна 1 ($\mathbf{1}^T u = 1$),
3. $P^T u \leq t\mathbf{1}$ (вектор не превосходит t поэлементно).

Здесь t — дополнительная переменная, принимающая значения на множестве действительных чисел ($t \in \mathbb{R}$).

Построение лагранжиана

Введём множители для ограничений: λ — для $P^T u \leq t\mathbf{1}$, μ — для $u \geq 0$, и ν — для $\mathbf{1}^T u = 1$. Тогда лагранжиан имеет вид:

$$L = t + \lambda^T(P^T u - t\mathbf{1}) - \mu^T u + \nu(1 - \mathbf{1}^T u) = \nu + (1 - \mathbf{1}^T \lambda)t + (P\lambda - \nu\mathbf{1} - \mu)^T u$$

Смешанные стратегии для матричных игр

Определение двойственной функции

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T\lambda = 1 \text{ и } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}$$

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T\lambda = 1 \text{ и } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}$$

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

1. $\lambda \geq 0$,

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T\lambda = 1 \text{ и } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}$$

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

1. $\lambda \geq 0$,
2. Сумма элементов вектора λ равна 1 ($\mathbf{1}^T\lambda = 1$),

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T \lambda = 1 \text{ и } P\lambda - \nu \mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}$$

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

1. $\lambda \geq 0$,
2. Сумма элементов вектора λ равна 1 ($\mathbf{1}^T \lambda = 1$),
3. $\mu \geq 0$,

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T\lambda = 1 \text{ и } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}$$

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

1. $\lambda \geq 0$,
2. Сумма элементов вектора λ равна 1 ($\mathbf{1}^T\lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T\lambda = 1 \text{ и } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}$$

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

1. $\lambda \geq 0$,
2. Сумма элементов вектора λ равна 1 ($\mathbf{1}^T\lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T \lambda = 1 \text{ и } P\lambda - \nu \mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}$$

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

1. $\lambda \geq 0$,
2. Сумма элементов вектора λ равна 1 $(\mathbf{1}^T \lambda = 1)$,
3. $\mu \geq 0$,
4. $P\lambda - \nu \mathbf{1} = \mu$.

После исключения μ получаем двойственную по Лагранжу задачу к (1):

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$
g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T\lambda = 1 \text{ и } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}
$$

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

1. $\lambda \geq 0$,
2. Сумма элементов вектора λ равна 1 ($\mathbf{1}^T\lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

После исключения μ получаем двойственную по Лагранжу задачу к (1):

$$
\begin{aligned}
&\max \nu \\
\text{s.t. } &\lambda \geq 0, \quad \mathbf{1}^T\lambda = 1, \quad P\lambda \geq \nu\mathbf{1}
\end{aligned}
$$

Смешанные стратегии для матричных игр

Определение двойственной функции

Двойственная функция $g(\lambda, \mu, \nu)$ определяется как:

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \text{если } \mathbf{1}^T\lambda = 1 \text{ и } P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{в противном случае} \end{cases}$$

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

1. $\lambda \geq 0$,
2. Сумма элементов вектора λ равна 1 ($\mathbf{1}^T\lambda = 1$),
3. $\mu \geq 0$,
4. $P\lambda - \nu\mathbf{1} = \mu$.

После исключения μ получаем двойственную по Лагранжу задачу к (1):

$$\begin{aligned} &\max \nu \\ \text{s.t. } &\lambda \geq 0, \quad \mathbf{1}^T\lambda = 1, \quad P\lambda \geq \nu\mathbf{1} \end{aligned}$$

Заключение

Эта постановка показывает, что двойственная по Лагранжу задача эквивалентна задаче (2). Так как оба этих линейных программирования допустимы, выполняется сильная двойственность, а значит, оптимальные значения задач (1) и (2) совпадают.