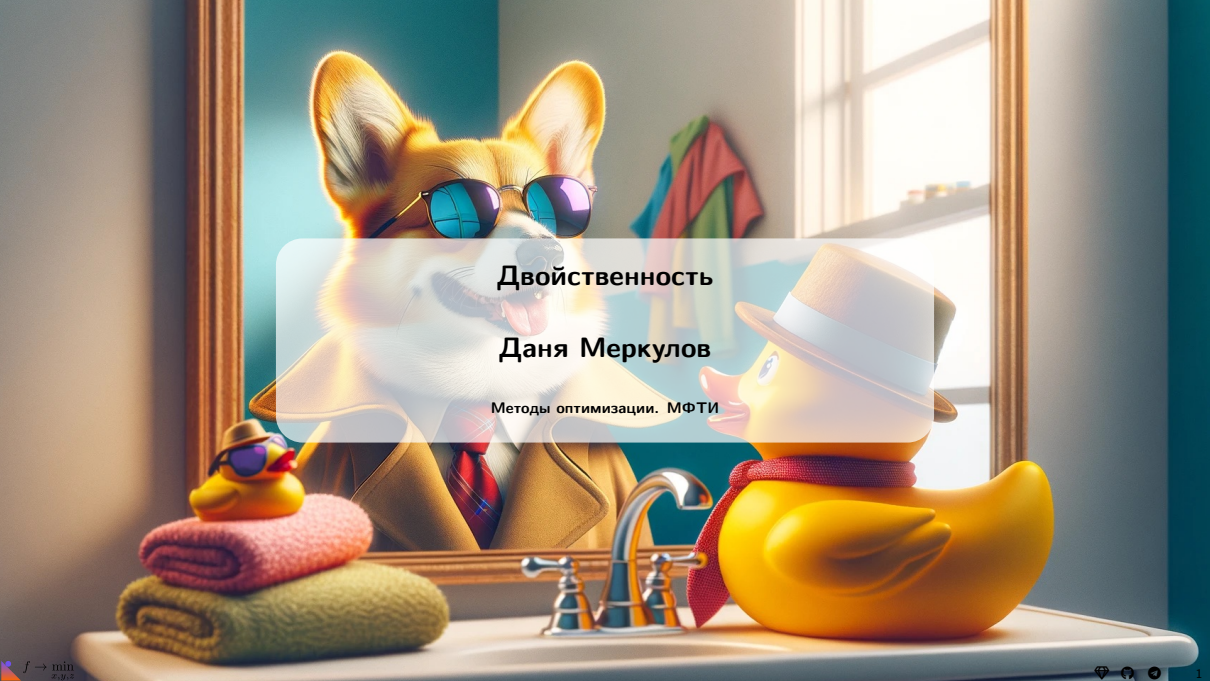




Двойственность

Даня Меркулов

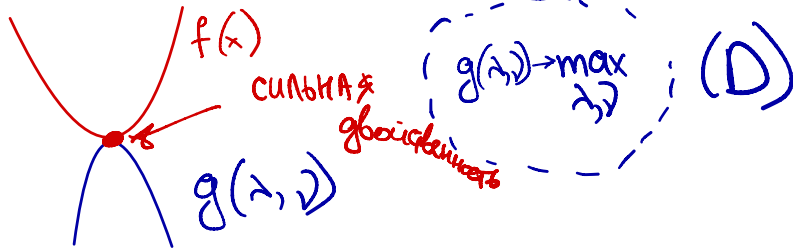
Методы оптимизации. МФТИ

Двойственность

Мотивация

Двойственность позволяет связать любую задачу оптимизации с ограничениями 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:

PRIMAL
(P) $\min f(x)$
 $x \in S$



Мотивация

Двойственность позволяет связать любую задачу оптимизации с ограничениями 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:

$$\text{Primal: } f(x) \rightarrow \min_{x \in S}$$

$$\text{Dual: } g(y) \rightarrow \max_{y \in \Omega}$$

Мотивация

Двойственность позволяет связать любую задачу оптимизации с ограничениями 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:

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

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

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

Мотивация

Двойственность позволяет связать любую задачу оптимизации с ограничениями 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:

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

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

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

В результате получаем:

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

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:

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

общая задача
мат. программирования

Лагранжева двойственность

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

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

И лагранжиан, соответствующий этой задаче:

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

Двойственная функция

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

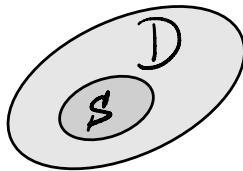
Определим **лагранжеву двойственную функцию** (или просто **двойственную функцию**)

$g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ как минимальное значение лагранжиана по x , для $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

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

инфимум

Двойственная функция



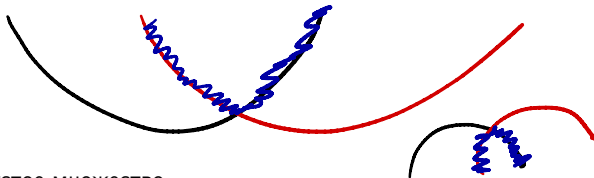
Handwritten notes: x with an arrow pointing to S , and $\delta \log x$ below it.

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

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

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

Двойственная функция



ВОГНУТАЯ



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

Определим **лагранжеву двойственную функцию** (или просто **двойственную функцию**) $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ как минимальное значение 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 $\lambda \succeq 0, \nu$.

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

$$\begin{aligned} \min f_0(x) \\ f_i(x) \leq 0 \\ h_i(x) = 0 \end{aligned}$$

$$\begin{aligned} (P) \text{ опт. } \\ \text{знач.} \\ \text{прямой} \\ \text{задачи} \\ p^* \\ \downarrow \\ f_0(x^*), x^* \in S \end{aligned}$$

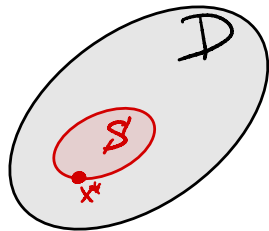
Двойственная функция как нижняя граница

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 $h_i(\hat{x}) = 0$, при $\lambda \succeq 0$. Тогда имеем:

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

пусть
 $\hat{x} \in S$



$$\Rightarrow L(x^*, \lambda, \nu) \leq f_0(x^*)$$

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 $h_i(\hat{x}) = 0$, при $\lambda \succeq 0$. Тогда имеем:

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

Следовательно, **опр. $g(\lambda, \nu)$**

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

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 $h_i(\hat{x}) = 0$, при $\lambda \succeq 0$. Тогда имеем:

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

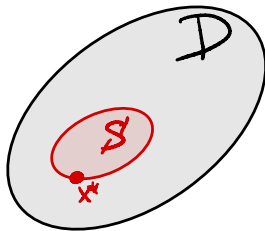
Следовательно,

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

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

$\forall \lambda, \nu$
 $\forall x \in \mathcal{S}$

Глобальная
оценка
снизу



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 $h_i(\hat{x}) = 0$, при $\lambda \succeq 0$. Тогда имеем:

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

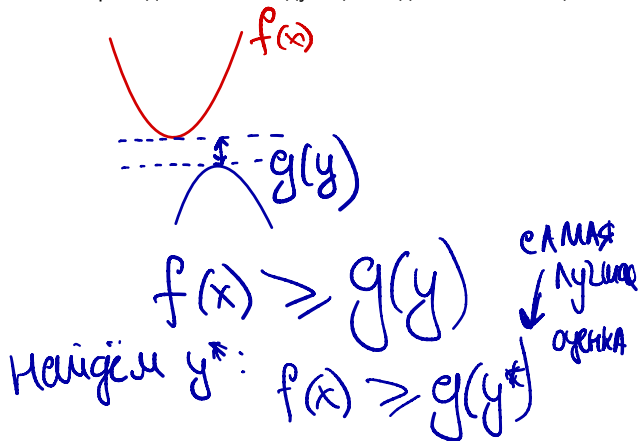
Следовательно,

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

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

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 $h_i(\hat{x}) = 0$, при $\lambda \succeq 0$. Тогда имеем:

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

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:

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

↑
двойственная
задача
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 $h_i(\hat{x}) = 0$, при $\lambda \succeq 0$. Тогда имеем:

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

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:

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

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

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**.

Саммари

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

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

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:

Построить
и решить
фэй-функцию

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

$$\|x\|_2^2 \quad \text{где} \quad m < n$$

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

$$\begin{aligned} \Rightarrow & \left\langle -\frac{1}{2}A^T, -\frac{1}{2}A^T \right\rangle + \\ & + \left\langle \lambda, -\frac{1}{2}AA^T - b \right\rangle = \end{aligned}$$

Решение: 1) $L(x, \lambda) = x^T x + \lambda^T (Ax - b)$
2) Запишем: $g(\lambda) = \inf_{x \in \mathbb{R}^n} [x^T x + \lambda^T (Ax - b)]$

$$\begin{aligned} \nabla_x L &= 0 \\ 2x + A^T \lambda &= 0 \\ x &= -\frac{1}{2}A^T \lambda \end{aligned}$$

$$\begin{aligned} \langle \lambda, Ax - b \rangle &= \\ = \langle \lambda, Ax \rangle - \langle \lambda, b \rangle \\ &= \langle A^T \lambda, x \rangle - \langle \lambda, b \rangle \end{aligned}$$

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

3) Запишем фэй-функцию : $g(\lambda) \rightarrow \max_{\lambda \in \mathbb{R}^m}$

Пример. Линейный МНК

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

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

$$x^T x \geq d^*$$

гешёво

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

$$4) -\frac{1}{4} v^T A A^T v - b^T v \rightarrow \max_{v \in \mathbb{R}^m}$$

$$d^* = ? \quad b^T (A A^T)^{-1} b$$

сколько стоит? $O(m^3)$

$$\nabla g = 0$$

$$-\frac{1}{4} \cdot 2 A A^T v - b = 0$$

$$v^* = -2 (A A^T)^{-1} b$$

$$O(m^2 n)$$

Пример. Линейный МНК

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

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

$$L(x, \lambda) = x^T x + \lambda^T (Ax - b)$$

$$\nabla_x L = 0$$

$$\begin{cases} 2x + A^T \lambda = 0 \\ Ax = b \end{cases} \quad x = -\frac{1}{2} A^T \lambda$$

$$-\frac{1}{2} A \cdot A^T \lambda = b$$

$$\lambda = -2(AA^T)^{-1} b$$

$$\begin{aligned} p^* \quad x^* x^* &= \left(A^T (AA^T)^{-1} b \right)^T A^T (AA^T)^{-1} b = \\ &= b^T (AA^T)^{-1} b = A^* \end{aligned}$$

А если найти p^* ?
 $O(nm^2)$

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

Пример. Линейный МНК

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

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

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:

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

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:

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

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:

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

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:

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

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 $\mathbb{R}^p$.

Согласно свойству нижней границы, для любого $\nu \in \mathbb{R}^p$ выполняется неравенство:

Пример. Линейный МНК

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

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

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 $\mathbb{R}^p$.

Согласно свойству нижней границы, для любого $\nu \in \mathbb{R}^p$ 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:

$$\min x^T W x$$

$$\text{s.t. } x_i^2 = 1, \quad i = 1, \dots, n,$$

Пример. Задача двустороннего разбиения

Мы рассматриваем (невыпуклую) задачу:

$$\min x^T W x$$

$$\text{s.t. } x_i^2 = 1, \quad i = 1, \dots, n,$$

$$x_i = \pm 1$$

всего 2^n точек

$$n = 100$$

$$2^{100} = (2^{10})^{10} = (10^3)^{10} = 10^{30}$$

6 разн.
- w_{ij}

$$\sum_{i,j} x_i \cdot w_{ij} \cdot x_j$$

стоимость
max.
i-ого
и
j-ого объектов
в группе

$$x_i = 1 \quad w_{ij}$$

$$x_j = 1$$

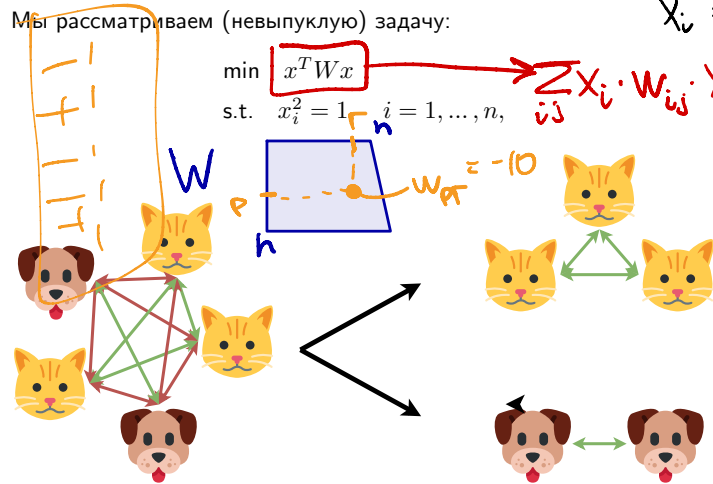


Рис. 1: Иллюстрация задачи двустороннего разбиения

Пример. Задача двустороннего разбиения

Мы рассматриваем (невыпуклую) задачу:

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

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

$$\{1, \dots, n\} = \{i | x_i = -1\} \cup \{i | x_i = 1\}.$$

Запишем $L(x, \mathcal{D})$:

$$\begin{aligned} L(x, \mathcal{D}) &= x^T W x + \sum_i \lambda_i (x_i^2 - 1) = \\ &= x^T W x + x^T \text{diag}(\mathcal{D}) x - \text{tr}(\text{diag}(\mathcal{D})) \end{aligned}$$

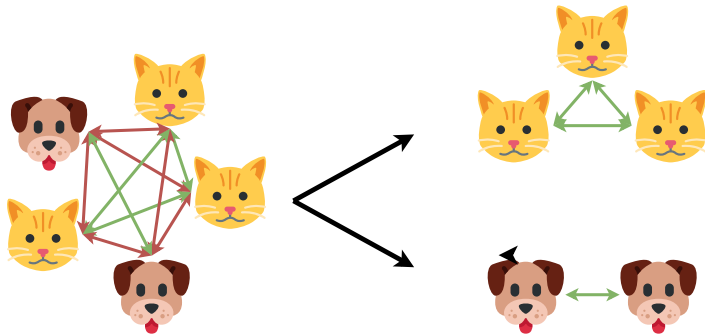


Рис. 1: Иллюстрация задачи двустороннего разбиения

Пример. Задача двустороннего разбиения

Мы рассматриваем (невыпуклую) задачу:

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

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

$$\{1, \dots, n\} = \{i | x_i = -1\} \cup \{i | x_i = 1\}.$$

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.

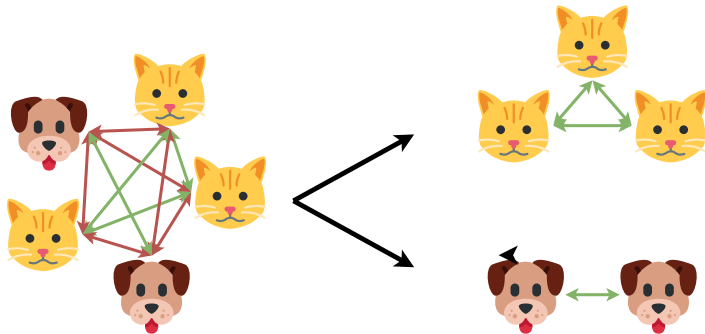


Рис. 1: Иллюстрация задачи двустороннего разбиения

Пример. Задача двустороннего разбиения

$$W \in \mathbb{S}^n$$

Теперь выведем двойственную функцию для этой задачи. Лагранжиан имеет вид

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

Усл 2:

$$g(\nu) = \inf_{x \in \mathbb{R}^n} \left[x^T (W + \text{diag}(\nu)) x - \mathbf{1}^T \nu \right]$$

$$W + \text{diag}(\nu) \succeq 0 \quad \leftarrow \begin{array}{l} \text{Требуем} \\ \text{(иначе} \\ L \rightarrow -\infty \end{array}$$

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:

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

нужно
 $\nu_i = \tilde{\nu} \in \mathbb{R}_-$

$\Rightarrow \nu_i = -\lambda_{\min}(W)$

КАНДИДАТ

нужно $\lambda(W)$:

~~КАНДИДАТ~~

$\lambda_{\min}(W) \quad 0 \quad \lambda_{\max}(W)$

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.

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:

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

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:

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

несколько $O(n^2)$

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:

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

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

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

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:

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

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

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

Код для этой задачи доступен здесь:  Open in Colab

$p^* = d^*$ СИЛЬНАЯ
двойственность.

$p \geq d$
и АБА двойственность

Сильная двойственность

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

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

Сильная двойственность

Связь 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

Связь 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 (которая всегда вогнутая, независимо от того, была ли исходная задача выпуклой), то мы получаем некоторую нижнюю границу. Удивительно, но существуют случаи, когда эти решения совпадают.

Сильная двойственность

Связь 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 (которая всегда вогнутая, 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

Связь 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 (которая всегда вогнутая, 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

Связь 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 (которая всегда вогнутая, 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

$$x^T x \rightarrow \min$$

$$Ax = b$$

$$m < n$$

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.

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

$$f_0(x) \geq d^*$$

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 $\max_{y \in \Omega} g(y) \leq \min_{x \in S} f_0(x)$ следует: если $\min_{x \in S} f_0(x) = -\infty$, то $\Omega = \emptyset$, и наоборот.

Полезные свойства двойственности

- Построение нижней границы для решения прямой задачи.

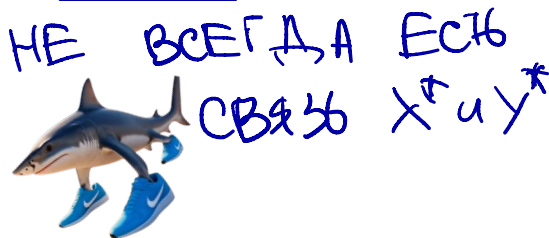
Иногда исходную задачу решить крайне сложно. Но если у нас есть двойственная задача, можно взять произвольное $y \in \Omega$ и подставить его в $g(y)$ — мы сразу получим некоторую нижнюю границу.

- Проверка разрешимости задачи и достижимости решения.

Из неравенства $\max_{y \in \Omega} g(y) \leq \min_{x \in S} f_0(x)$ следует: если $\min_{x \in S} f_0(x) = -\infty$, 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 $\max_{y \in \Omega} g(y) \leq \min_{x \in S} f_0(x)$ следует: если $\min_{x \in S} f_0(x) = -\infty$, 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,
 $p^* \in [g(y), f_0(x)], d^* \in [g(y), f_0(x)]$

ЗАЗОР ОТ
 $f_0^* \geq g(y)$
 $-f_0^* \leq -g(y)$

$$f_0(x_k) - f_0^* \leq \epsilon$$

$$f_0(x_k) - g(y) \leq \epsilon$$

СЕРТИФИКАТ
ОПТИМАЛЬНОСТИ

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 $\max_{y \in \Omega} g(y) \leq \min_{x \in S} f_0(x)$ следует: если $\min_{x \in S} f_0(x) = -\infty$, 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 минимумом аффинных функций.

Условие Слейтера

Theorem

Если для выпуклой задачи оптимизации (предполагаем $\min$, 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

i Example

KKT:

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

$$L = x + \lambda \frac{x^2}{2}$$

$$\nabla_x L = 1 + \lambda x = 0$$

$$\lambda \geq 0$$

$$\lambda \frac{x^2}{2} = 0$$

$$\lambda \geq 0$$

$$\frac{x^2}{2} \leq 0$$

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

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

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:

Стивен Бойд

$$x^\top Ax + 2b^\top x \rightarrow \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } x^\top x \leq 1$$

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:

$$x^\top Ax + 2b^\top x \rightarrow \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } x^\top x \leq 1$$

где $A \in \mathbb{S}^n$, $A \not\preceq 0$ и $b \in \mathbb{R}^n$.

Поскольку $A \not\preceq 0$, задача невыпуклая.

Эту задачу иногда называют *trust*

region problem — задачей

доверительной области; она возникает

при минимизации второй

аппроксимации функции на

единичном шаре, в пределах которого

эта аппроксимация считается

примерно корректной.

Невыпуклая квадратичная задача с сильной двойственностью

Решение

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

$$x^\top Ax + 2b^\top x \rightarrow \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } x^\top x \leq 1$$

где $A \in \mathbb{S}^n$, $A \not\preceq 0$ и $b \in \mathbb{R}^n$.

Поскольку $A \not\preceq 0$, задача невыпуклая.

Эту задачу иногда называют *trust region problem* — задачей

доверительной области; она возникает при минимизации второй аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Невыпуклая квадратичная задача с сильной двойственностью

Решение

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

Лагранжиан и двойственная функция:

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

$$x^\top Ax + 2b^\top x \rightarrow \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } x^\top x \leq 1$$

где $A \in \mathbb{S}^n$, $A \not\preceq 0$ и $b \in \mathbb{R}^n$.

Поскольку $A \not\preceq 0$, задача невыпуклая.

Эту задачу иногда называют *trust region problem* — задачей

доверительной области; она возникает при минимизации второй аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Невыпуклая квадратичная задача с сильной двойственностью

Решение

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

Лагранжиан и двойственная функция:

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

$$x^\top A x + 2b^\top x \rightarrow \min_{x \in \mathbb{R}^n}$$

$$\text{s.t. } x^\top x \leq 1$$

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

где $A \in \mathbb{S}^n$, $A \not\geq 0$ и $b \in \mathbb{R}^n$.

Поскольку $A \not\geq 0$, задача невыпуклая.

Эту задачу иногда называют *trust region problem* — задачей

доверительной области; она возникает при минимизации второй 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:

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

где $A \in \mathbb{S}^n$, $A \not\geq 0$ и $b \in \mathbb{R}^n$.
Поскольку $A \not\geq 0$, задача невыпуклая. Эту задачу иногда называют *trust region problem* — задачей доверительной области; она возникает при минимизации второй аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Решение

Лагранжиан и двойственная функция:

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

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

Двойственная задача:

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

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:

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

где $A \in \mathbb{S}^n$, $A \not\geq 0$ и $b \in \mathbb{R}^n$. Поскольку $A \not\geq 0$, задача невыпуклая. Эту задачу иногда называют *trust region problem* — задачей доверительной области; она возникает при минимизации второй аппроксимации функции на единичном шаре, в пределах которого эта аппроксимация считается примерно корректной.

Решение

Лагранжиан и двойственная функция:

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

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

Двойственная задача:

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

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

Приложения

Решение прямой задачи через двойственную

Важное следствие условия стационарности: при выполнении сильной двойственности, если известны решения двойственной задачи λ^*, ν^* , то любое решение прямой задачи x^* удовлетворяет

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

Часто решения этой задачи без ограничений можно записать **в явном виде**, что позволяет явно выразить решения прямой задачи через решения двойственной.

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:

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

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:

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

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:

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

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 .

Отсюда получаем:

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

Смешанные стратегии для матричных игр

Начало игры



Player 1

Петя

u_1
...
u_k
...
u_n

v_1
...
...
v_l
...
...
v_m



Player 2

Александр

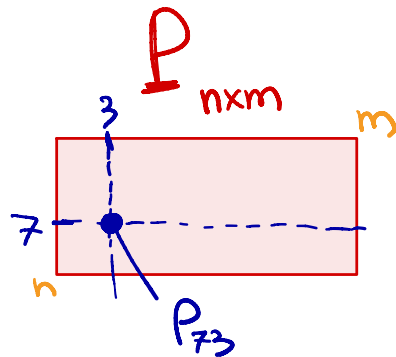


Рис. 2: Схема смешанной стратегии в матричной игре

Смешанные стратегии для матричных игр



Player 1

u_1
$\dots$
u_k
$\dots$
u_n

v_1
$\dots$
$\dots$
v_l
$\dots$
$\dots$
v_m



Player 2

В нулевой (zero-sum) матричной игре 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 i , а игрок 2 — с вероятностями v_l .

Рис. 2: Схема смешанной стратегии в матричной игре

Смешанные стратегии для матричных игр



Player 1

u_1
$\dots$
u_k
$\dots$
u_n

v_1
$\dots$
$\dots$
v_l
$\dots$
$\dots$
v_m



Player 2

В нулевой (zero-sum) матричной игре 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 i , а игрок 2 — с вероятностями v_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**.

Рис. 2: Схема смешанной стратегии в матричной игре

Смешанные стратегии для матричных игр. Перспектива игрока 1



Player 1

Петя

u_1
$\dots$
u_k
$\dots$
u_n

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

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

Смешанные стратегии для матричных игр. Перспектива игрока 1



Player 1

u_1
$\dots$
u_k
$\dots$
u_n

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

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

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

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

(1)

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

Смешанные стратегии для матричных игр. Перспектива игрока 2

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

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

u ↗

=

0
0
1
6
8



Player 2

Александр

v_1

...

...

v_l

...

...

v_m

Смешанные стратегии для матричных игр. Перспектива игрока 2

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

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

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

$$\begin{aligned} \max \quad & \min_{i=1, \dots, n} (P v)_i \\ \text{s.t.} \quad & v \geq 0 \\ & 1^T v = 1 \end{aligned}$$

(2)

Оптимальное значение здесь обозначается как p_2^* .



Player 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 ($1^T u = 1$),

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 ($1^T u = 1$),
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 ($1^T u = 1$),
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 ($1^T u = 1$),
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 ($1^T u = 1$),
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 ($1^T u = 1$),
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 $1^T u = 1$. 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 ($1^T u = 1$),
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 $1^T u = 1$. Тогда лагранжиан имеет вид:

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

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 ($1^T u = 1$),
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 $1^T u = 1$. Тогда лагранжиан имеет вид:

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

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:

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

Смешанные стратегии для матричных игр

Определение двойственной функции

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

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

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:

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

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

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

Смешанные стратегии для матричных игр

Определение двойственной функции

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

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

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

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

Смешанные стратегии для матричных игр

Определение двойственной функции

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

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

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

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

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

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

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

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

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

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

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

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

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

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

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

Смешанные стратегии для матричных игр

Определение двойственной функции

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

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

Решение двойственной задачи

Двойственная задача заключается в максимизации ν при следующих условиях:

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

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

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

Заключение

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

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints
- On Second Order Optimality Conditions in Nonlinear Optimization

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints
- On Second Order Optimality Conditions in Nonlinear Optimization
- Numerical Optimization by Jorge Nocedal and Stephen J. Wright.

- Lecture on KKT conditions (very intuitive explanation) in the course “Elements of Statistical Learning” @ KTH.
- One-line proof of KKT
- On the Second Order Optimality Conditions for Optimization Problems with Inequality Constraints
- On Second Order Optimality Conditions in Nonlinear Optimization
- Numerical Optimization by Jorge Nocedal and Stephen J. Wright.
- Duality Uses and Correspondences lecture by Ryan Tibshirani course.