

# Второе задание

Хохлов Алексей

1 апреля 2019 г.

## 1 Проекция на шар

Если  $\|\mathbf{a}\|_2 \leq 1$ , то, очевидно, проекция  $\mathbf{p} = \mathbf{a}$ .

Если же  $\|\mathbf{a}\|_2 > 1$ , очевидно, что искомая точка — точка пересечения прямой, соединяющей центр сферы и  $\mathbf{a}$ , с поверхностью единичной сферы. Другими словами, искомая точка  $\mathbf{p} = \frac{\mathbf{a}}{\|\mathbf{a}\|_2}$ . Чтобы доказать это, покажем, что  $\forall \mathbf{x} \in \mathbf{X}$  верно  $\|\mathbf{a} - \mathbf{x}\|_2 \geq \|\mathbf{a} - \mathbf{p}\|_2$ .

Для правой стороны неравенства

$$\|\mathbf{a} - \mathbf{p}\|_2 = \left\| \mathbf{a} - \frac{\mathbf{a}}{\|\mathbf{a}\|_2} \right\|_2 = \left(1 - \frac{1}{\|\mathbf{a}\|_2}\right) \|\mathbf{a}\|_2 = \|\mathbf{a}\|_2 - 1 \quad (1.1)$$

Для левой стороны неравенства

$$\|\mathbf{a} - \mathbf{x}\|_2 \geq \|\mathbf{a}\|_2 - \|\mathbf{x}\|_2 \geq \|\mathbf{a}\|_2 - 1 = \|\mathbf{a} - \mathbf{p}\|_2 \quad (1.2)$$

Видим, что для любой точки в единичном шаре точка  $\mathbf{p} = \frac{\mathbf{a}}{\|\mathbf{a}\|_2}$  является решением задачи

$$\|\mathbf{a} - \mathbf{x}\|_2 \rightarrow \min \quad (1.3)$$

## 2 Минимизация скалярного произведения

Искомая точка

$$p_i = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases} \quad (2.1)$$

$$(2.2)$$

где  $k$  — такая, что  $c_k = \min(c_1, c_2, \dots, c_n)$ . Докажем это.

Поскольку  $c_k \leq c_i$  для любого  $i$ , то  $c_k x_i \leq c_i x_i$ . Значит, что

$$\langle \mathbf{c}, \mathbf{p} \rangle = c_k = c_k \sum_{i=1}^n x_i = \sum_{i=1}^n c_k x_i \leq \sum_{i=1}^n c_i x_i = \langle \mathbf{c}, \mathbf{x} \rangle \quad (2.3)$$

## 3 Максимизация функции правдоподобия

В решении использованы формулы

$$\frac{\partial}{\partial \mathbf{A}} (\text{tr}(\mathbf{A}^{-1} \mathbf{B})) = -(\mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1})^T \quad (3.1)$$

$$\frac{\partial}{\partial \mathbf{A}} \det(\mathbf{A}) = \det(\mathbf{A}) \mathbf{A}^{-T} \quad (3.2)$$

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) \quad (3.3)$$

### 3.1

Разложим функцию на слагаемые

$$f = \frac{m}{2} \log \det \Sigma + \frac{1}{2} \sum_{i=1}^m (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) = f_1 + f_2 \quad (3.4)$$

Найдем производные слагаемых по  $\Sigma$ .

Для первого слагаемого воспользуемся (3.2)

$$\frac{\partial}{\partial \Sigma} f_1 = \frac{m}{2} \frac{1}{\det \Sigma} \det(\Sigma) \Sigma^{-T} = \frac{m}{2} \Sigma^{-T} \quad (3.5)$$

Для второго слагаемого воспользуемся (3.3) и (3.1).

$$\begin{aligned} \frac{\partial}{\partial \Sigma} f_2 &= \frac{1}{2} \frac{\partial}{\partial \Sigma} \text{tr} \left( \sum_{i=1}^m (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) \right) = \frac{1}{2} \frac{\partial}{\partial \Sigma} \text{tr} \left( \Sigma^{-1} \sum_{i=1}^m (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \right) = \\ &= -\frac{1}{2} \left( \Sigma^{-1} \left( \sum_{i=1}^m (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \right) \Sigma^{-1} \right)^T = -\frac{1}{2} \Sigma^{-T} \left( \sum_{i=1}^m (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \right)^T \Sigma^{-T} \end{aligned} \quad (3.6)$$

Приравняем производную функции к нулю

$$\frac{m}{2} \Sigma^{-T} - \frac{1}{2} \Sigma^{-T} \left( \sum_{i=1}^m (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \right)^T \Sigma^{-T} = 0 \quad (3.7)$$

Умножив на  $\Sigma$  и транспонировав, получим

$$\Sigma = \frac{1}{m} \sum_{i=1}^m (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \quad (3.8)$$

### 3.2

Докажем формулу (3.1). Во-первых, из  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{E}$  следует

$$\frac{\partial}{\partial a_{pq}} \sum_{j=1}^n \tilde{a}_{ij} a_{jk} = \sum_{j=1}^n \frac{\partial \tilde{a}_{ij}}{\partial a_{pq}} a_{jk} + \sum_{j=1}^n \tilde{a}_{ij} \frac{\partial a_{jk}}{\partial a_{pq}} = 0 \quad (3.9)$$

Домножая справа на обратную матрицу, получим

$$\frac{\partial \tilde{a}_{ij}}{\partial a_{pq}} = - \sum_{k=1}^n \sum_{m=1}^n \tilde{a}_{im} \frac{\partial a_{mk}}{\partial a_{pq}} \tilde{a}_{kj} = -\tilde{a}_{ip} \tilde{a}_{qj} \quad (3.10)$$

Раскроем теперь производную следа

$$\frac{\partial}{\partial a_{pq}} (\text{tr}(\mathbf{A}^{-1} \mathbf{B})) = \frac{\partial}{\partial a_{pq}} \sum_{i=1}^n \sum_{j=1}^n \tilde{a}_{ij} b_{ji} = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial \tilde{a}_{ij}}{\partial a_{pq}} b_{ji} \quad (3.11)$$

Подставляя значение производной из (3.10)

$$\frac{\partial}{\partial a_{pq}} (\text{tr}(\mathbf{A}^{-1} \mathbf{B})) = - \sum_{i=1}^n \sum_{j=1}^n \tilde{a}_{ip} \tilde{a}_{qj} b_{ji} = - \sum_{i=1}^n \sum_{j=1}^n \tilde{a}_{pi}^T b_{ij}^T \tilde{a}_{jq}^T \quad (3.12)$$

получим в матричном виде искомую формулу (3.1).

## 4 Парабола

$(x - 3)(x - 1) \leq 0$  — задача оптимизации решается на множестве  $Q = [1; 3]$

### 4.1 а

$\frac{\partial}{\partial x}(x^2 + 1) = 2x > 0$ ,  $\frac{\partial^2}{\partial x^2}(x^2 + 1) = 1 > 0$  — функция строго возрастающая на отрезке  $[1; 3]$ . Значит, минимальное значение достигается в точке  $x = 1$ .

$$\min_{x \in Q}(x^2 + 1) = x^2 + 1|_{x=1} = 2 \quad (4.1)$$

### 4.2 б

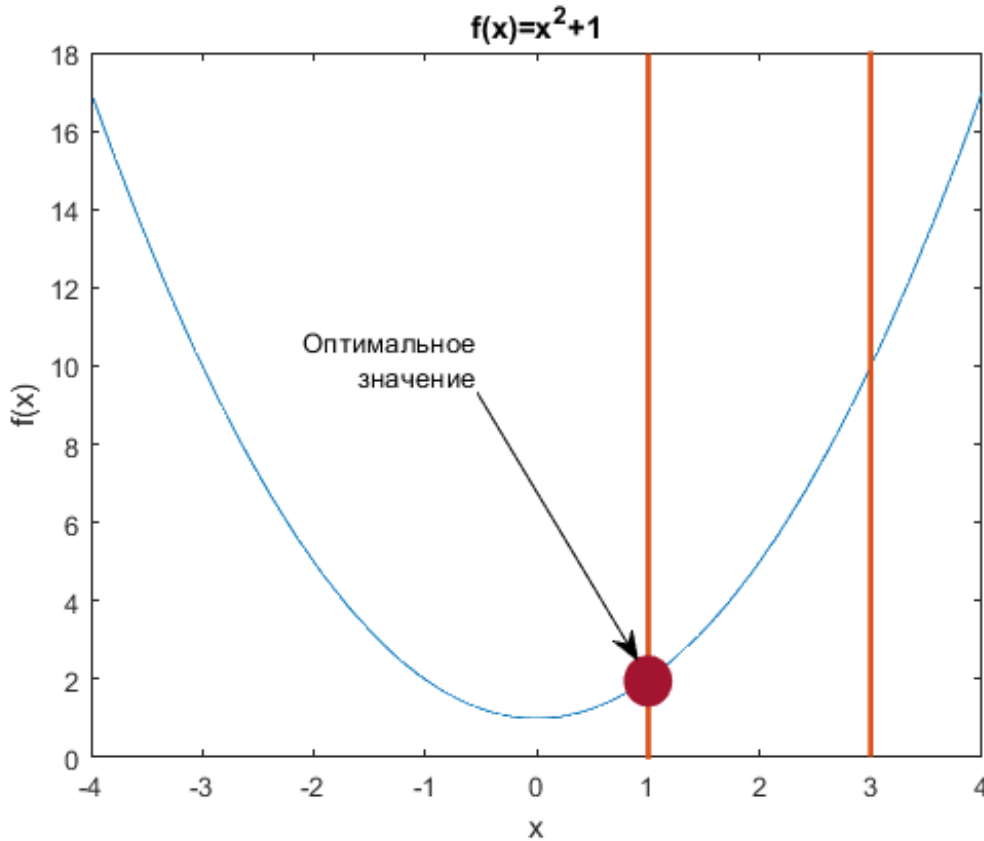


Рис. 1: Целевая функция, допустимое множество и оптимальное значение.

### 4.3 с

$$L(x, \mu) = x^2 + 1 + \mu(x - 1)(x - 3) = (1 + \mu)x^2 - 4\mu x + (1 + 3\mu) \quad (4.2)$$

Лагранжиан — параболическая функция. При  $\mu \leq -1$  инфимум  $\inf L = -\infty$ . При  $\mu > -1$  минимум параболического лагранжиана достигается в точке  $x = \frac{2\mu}{(1 + \mu)}$ . Двойственная функция запишется так:

$$g(\mu) = \inf L(x, \mu) = \frac{4\mu^2}{1 + \mu} - \frac{8\mu^2}{1 + \mu} + \frac{(1 + 3\mu)(1 + \mu)}{1 + \mu} = \frac{-\mu^2 + 4\mu + 1}{1 + \mu} = -\mu + 5 - \frac{4}{1 + \mu} \quad (4.3)$$

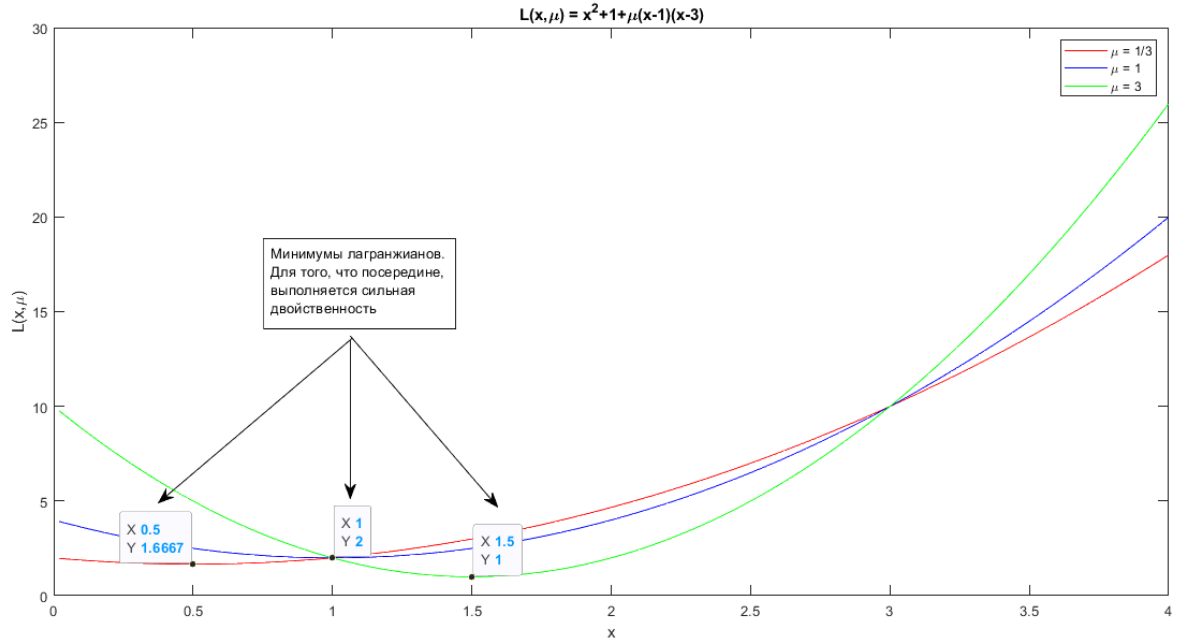


Рис. 2: График лагранжиана для нескольких  $\mu$ .  $p^* \geq \inf_x L(x, \mu)$

## 4.4 d

Формулировка двойственной задачи:

$$\begin{aligned} \max g(\mu) &= \max\left(-\mu + 5 - \frac{4}{1 + \mu}\right) \\ \text{s.t. } \mu &\geq 0 \end{aligned} \quad (4.4)$$

Найдем первую и вторую производные двойственной функции:

$$\frac{\partial}{\partial \mu} g(\mu) = -1 + \frac{4}{(1 + \mu)^2} \quad (4.5)$$

$$\frac{\partial^2}{\partial \mu^2} g(\mu) = -\frac{8}{(1 + \mu)^3} \quad (4.6)$$

При  $\mu \geq 0$ , как видно, двойственная функция является вогнутой, и поэтому ее экстремум при  $\mu = 1$  является максимумом.

$$\max_{\mu \geq 0} g(\mu) = g(\mu)|_{\mu=1} = 2 = (x^2 + 1)|_{x=1} = \min_{x \in Q} (x^2 + 1) \quad (4.7)$$

Сильная двойственность выполняется.

## 5 Задача бинарного линейного программирования

### 5.1

Задачу можно переписать в виде

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ & x_i(x_i - 1) = 0, i = 1, \dots, n \\ & \mathbf{Ax} \leq \mathbf{b} \end{aligned} \quad (5.1)$$

Лагранжиан запишется в виде

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_{i=1}^n (c_i x_i + \lambda_i x_i (x_i - 1)) + \sum_j \mu_j \left( \sum_{i=1}^n a_{ji} x_i - b_j \right) = \quad (5.2)$$

$$= \sum_{i=1}^n \left( \lambda_i x_i^2 + \left[ c_i - \lambda_i + \sum_j \mu_j a_{ji} \right] x_i \right) + \sum_j \mu_j b_j \quad (5.3)$$

Лагранжиан — параболическая функция. Рассмотрим лагранжиан при разных параметрах.

Если найдется такая  $i$ , что  $\lambda_i < 0$ , то инфимум  $\inf L = -\infty$ .

Если найдется такая  $i$ , что  $\lambda_i = 0$ ,  $c_i + \sum_j \mu_j a_{ji} \neq 0$ , то инфимум  $\inf L = -\infty$ .

Пусть теперь либо  $\lambda_i = 0$ ,  $c_i + \sum_j \mu_j a_{ji} = 0$ , либо  $\lambda_i > 0$ . Если для  $i$  выполнено первое условие, то вклад в инфимум функции от слагаемых, соответствующих этим  $i$ , равен нулю. Слагаемые, для которых выполнено второе условие, достигают минимума в точке  $x_i = -\frac{c_i - \lambda_i + \sum_j \mu_j a_{ji}}{2\lambda_i}$

Инфимум запишется в виде

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_x L = - \sum_i \frac{\left( c_i - \lambda_i + \sum_j \mu_j a_{ji} \right)^2}{4\lambda_i} + \sum_j \mu_j b_j \quad (5.4)$$

Суммирование ведется по таким  $i$ , что  $\lambda_i > 0$ .

## 5.2

Задача для непрерывной релаксации переписывается следующим образом:

$$\begin{aligned} \min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} \\ x_i(x_i - 1) \leq 0, i = 1, \dots, n \\ \mathbf{Ax} \leq \mathbf{b} \end{aligned} \quad (5.5)$$

Видно, что двойственная функция  $g(\boldsymbol{\mu}, \tilde{\boldsymbol{\mu}})$  запишется в таком же виде, что и в (5.4), только  $\tilde{\mu}_i$  вместо  $\lambda_i$ , причём для  $\tilde{\mu}_i$  условие  $\tilde{\mu}_i \geq 0$  будет автоматически выполнено при поиске решения двойственной задачи. Соответственно, при  $\lambda_i \geq 0$  максимумы двойственных функций совпадут друг с другом, и, иными словами, нижняя оценка релаксации Лагранжа совпадет с оценкой непрерывной релаксации.

Поскольку выполняются условия ККТ, то решение этой задачи действительно дает оценку снизу.

## 6 Задача наименьших квадратов

### 6.1

Лагранжиан для задачи запишется так:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \boldsymbol{\lambda}^T (\mathbf{Gx} - \mathbf{h}) \quad (6.1)$$

Производная лагранжиана по  $\mathbf{x}$  такова

$$\frac{\partial}{\partial \mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{A}^T (\mathbf{Ax} - \mathbf{b}) + \mathbf{G}^T \boldsymbol{\lambda} = \mathbf{A}^T \mathbf{Ax} - (\mathbf{A}^T \mathbf{b} - \mathbf{G}^T \boldsymbol{\lambda}) \quad (6.2)$$

С одной стороны, по теореме о ранге произведения матриц, ранг матрицы  $\mathbf{A}^T \mathbf{A}$  не превосходит  $n$ . С другой стороны, по неравенству Сильвестра  $2n = \text{rank}(\mathbf{A}^T) + \text{rank}(\mathbf{A}) \leq \text{rank}(\mathbf{A}^T \mathbf{A}) + n$ . Итак, ранг матрицы  $(n \times n)$   $\mathbf{A}^T \mathbf{A}$  равен  $n$  — значит, для неё существует обратная.

Тогда  $\frac{\partial}{\partial \mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}) = 0$  при  $\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^{-1}(\mathbf{A}^T \mathbf{b} - \mathbf{G}^T \boldsymbol{\lambda})$ . Отсюда можно получить двойственную функцию.

Двойственная задача запишется так:

$$\max_{\boldsymbol{\lambda}} \frac{1}{2} \|\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1}(\mathbf{A}^T \mathbf{b} - \mathbf{G}^T \boldsymbol{\lambda}) - \mathbf{b}\|_2^2 + \boldsymbol{\lambda}^T (\mathbf{G}(\mathbf{A}^T \mathbf{A})^{-1}(\mathbf{A}^T \mathbf{b} - \mathbf{G}^T \boldsymbol{\lambda}) - \mathbf{h}) \quad (6.3)$$

Для исходной (выпуклой) задачи выполнены условия ККТ — значит, выполняется сильная двойственность, т.е. решение двойственной задачи совпадает с решением исходной.

## 6.2

Найдём решение исходной задачи. Подставим найденное  $\mathbf{x}^*$  в равенство  $\mathbf{G}\mathbf{x} = \mathbf{h}$ . Получим

$$\mathbf{G}(\mathbf{A}^T \mathbf{A})^{-1}(\mathbf{A}^T \mathbf{b} - \mathbf{G}^T \boldsymbol{\lambda}^*) = \mathbf{h} \quad (6.4)$$

$$\mathbf{G}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} - \mathbf{h} = \mathbf{G}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{G}^T \boldsymbol{\lambda}^* \quad (6.5)$$

Матрица перед  $\boldsymbol{\lambda}^*$  представляет собой произведение трех матриц:  $\mathbf{G}$  размера  $p \times n$  и ранга  $p$ ,  $(\mathbf{A}^T \mathbf{A})^{-1}$  размера  $n \times n$  и ранга  $n$ ,  $\mathbf{G}^T$  размера  $n \times p$  и ранга  $p$ . Итоговая матрица имеет размер  $p \times p$ . Если  $p \leq n$ , то применяя теорему о ранге произведения и неравенство Сильвестра, получим, что ранг итоговой матрицы  $(p \times p)$  равен  $p$  — она обратима. Тогда решение исходной задачи запишется как

$$\boldsymbol{\lambda}^* = [\mathbf{G}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{G}^T]^{-1} [\mathbf{G}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} - \mathbf{h}] \quad (6.6)$$

$$\mathbf{x}^* = (\mathbf{A}^T \mathbf{A})^{-1}(\mathbf{A}^T \mathbf{b} - \mathbf{G}^T \boldsymbol{\lambda}^*) \quad (6.7)$$

$$\min_{\mathbf{x}} f(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x}^* - \mathbf{b}\|_2^2 = g(\boldsymbol{\lambda}^*) = \max_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}) \quad (6.8)$$

## 7 Условие на матрицу

### 7.1

Из критерия Сильвестра (все главные миноры неотрицательны) равносильная запись задачи

$$\begin{aligned} & \min(y_1) \\ & \text{s.t. } y_2 \geq 0 \\ & y_1 + 1 \geq 0 \\ & y_2(y_1 + 1) \geq 0 \\ & -y_1^2(y_1 + 1) \geq 0 \\ & -y_1^2 \geq 0 \end{aligned} \quad (7.1)$$

Или в каноничной записи

$$\begin{aligned}
& \min(y_1) \\
& \text{s.t. } -y_2 \leq 0 \\
& -y_1 - 1 \leq 0 \\
& -y_2(y_1 + 1) \leq 0 \\
& y_1^2(y_1 + 1) \leq 0 \\
& y_1^2 \leq 0
\end{aligned} \tag{7.2}$$

В такой записи задача не является выпуклой, т.к. функция  $y_1^2(y_1 + 1)$  является выпуклой не при всех  $y$ .

Упростим условия:

$$\begin{aligned}
& \min(y_1) \\
& \text{s.t. } y_2 \geq 0 \\
& y_1 = 0
\end{aligned} \tag{7.3}$$

Решение -  $y_1 = 0, y_2 \geq 0$

## 7.2

Лагранжиан можно записать как

$$\begin{aligned}
L(y_1, y_2, \Lambda) &= y_1 - \text{tr}(\mathbf{Y}^T \mathbf{\Lambda}) = y_1 - y_1 \lambda_{21} - y_1 \lambda_{12} - y_2 \lambda_{22} - (y_1 + 1) \lambda_{33} = \\
&= y_1(1 - \lambda_{21} - \lambda_{12} - \lambda_{33}) - \lambda_{22} y_2 - \lambda_{33}
\end{aligned} \tag{7.4}$$

$$\mathbf{Y}^T \mathbf{\Lambda} = \begin{pmatrix} 0 & y_1 & 0 \\ y_1 & y_2 & 0 \\ 0 & 0 & y_1 + 1 \end{pmatrix} \begin{pmatrix} \lambda_{11} & \lambda_{12} & \lambda_{13} \\ \lambda_{21} & \lambda_{22} & \lambda_{23} \\ \lambda_{31} & \lambda_{32} & \lambda_{33} \end{pmatrix} = \begin{pmatrix} y_1 \lambda_{21} & \dots & \dots \\ \dots & y_1 \lambda_{12} + y_2 \lambda_{22} & \dots \\ \dots & \dots & (y_1 + 1) \lambda_{33} \end{pmatrix}$$

$$\begin{aligned}
\frac{\partial}{\partial y_1} L(y_1, y_2, \Lambda) &= 1 - \lambda_{21} - \lambda_{12} - \lambda_{33} = 0 \\
\frac{\partial}{\partial y_2} L(y_1, y_2, \Lambda) &= -\lambda_{22} = 0
\end{aligned} \tag{7.5}$$

Отсюда

$$L(y_1, y_2, \Lambda) = -\lambda_{33} \tag{7.6}$$

Для двойственной задачи

$$\inf_{y_1, y_2} L = \begin{cases} -\infty, \lambda_{22} \neq 0, 1 - \lambda_{21} - \lambda_{12} - \lambda_{33} \neq 0 \\ -\lambda_{33}, \lambda_{22} = 0, 1 - \lambda_{21} - \lambda_{12} - \lambda_{33} = 0 \end{cases} \tag{7.7}$$

$$\tag{7.8}$$

$$\max g(\Lambda) = \max(-\lambda_{33}) = +\infty \tag{7.9}$$

Зазор двойственности —  $+\infty$ . Сильная двойственность не выполняется, т.к. не выполняется условие неотрицательности  $\Lambda$  при оптимальных двойственных переменных.