Второе задание

Хохлов Алексей

28 марта 2019 г.

1 Проекция на шар

Если $\|\mathbf{a}\|_2 \leqslant 1$, 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 $\mathbf{p} = \dfrac{\mathbf{a}}{\|\mathbf{a}\|_2}$. Чтобы доказать это, покажем, что $\forall \mathbf{x} \in \mathbf{X}$ верно $\|\mathbf{a} - \mathbf{x}\|_2 \geqslant \|\mathbf{a} - \mathbf{p}\|_2$.

Для правой стороны неравенства

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

Для левой стороны неравенства

$$\|\mathbf{a} - \mathbf{x}\|_2 \geqslant \|\mathbf{a}\|_2 - \|\mathbf{x}\|_2 \geqslant \|\mathbf{a}\|_2 - 1 = \|\mathbf{a} - \mathbf{p}\|_2 \tag{1.2}$$

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

$$\|\mathbf{a} - \mathbf{x}\|_2 \to \min \tag{1.3}$$

2 Минимизация скалярного произведения

Искомая точка

$$p_i = \begin{cases} 1,\, i = k \\ 0,\, i \neq k \end{cases} \tag{2.1}$$

$$\tag{2.2}$$

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

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

$$\langle \mathbf{c}, \mathbf{p} \rangle = c_k = c_k \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} c_k x_i \leqslant \sum_{i=1}^{n} c_i x_i = \langle \mathbf{c}, \mathbf{x} \rangle \tag{2.3}$$

3 Максимизация функции правдоподобия

В решении использованы формулы

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

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

$$\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{BCA}) \tag{3.3}$$

3.1

Разложим функцию на слагаемые

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

Найдем производные слагаемых по Σ.
Для первого слагаемого воспользуемся (3.2)

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

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

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

Приравняем производную функции к нулю

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

Умножив на Σ и транспонировав, получим

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

3.2

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

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

Домножая справа на обратную матрицу, получим

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

Раскроем теперь производную следа

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

Подставляя значение производной из (3.10)

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

получим в матричном виде искомую формулу (3.1).

4 Парабола

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

4.1 a

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

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

4.2 b

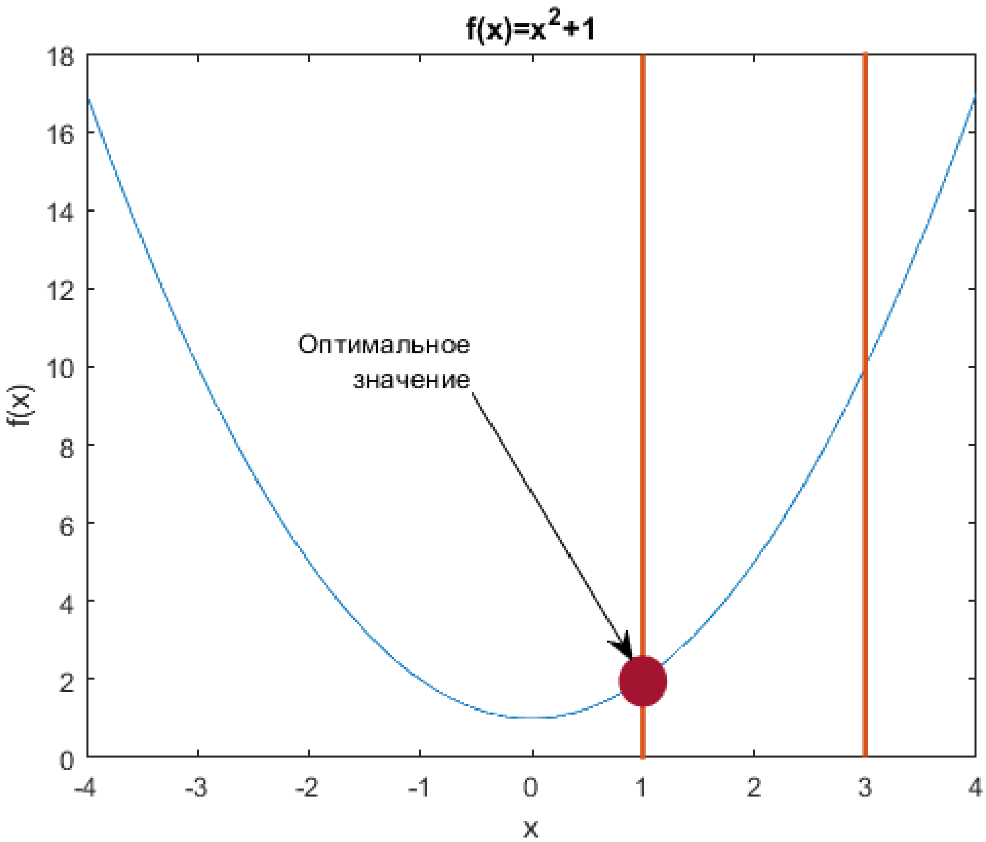

Рис. 1: Целевая функция, допустимое множество и оптимальное значение.

4.3 c

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

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

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

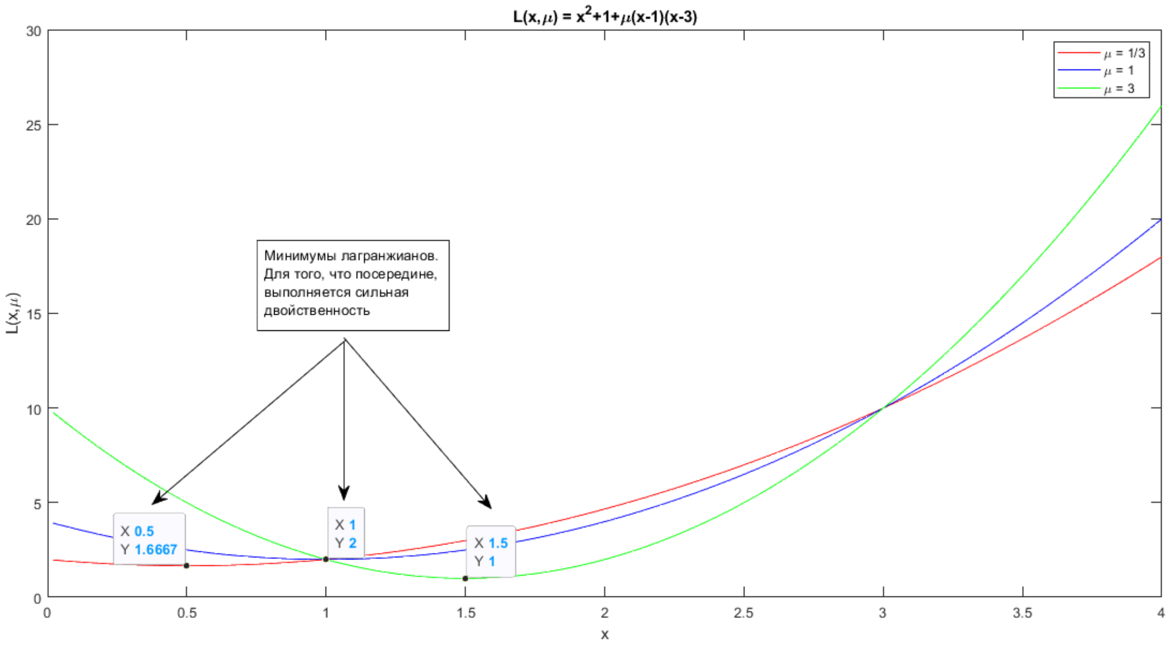

Рис. 2: График лагранжиана для нескольких μ. $p^* \geqslant \inf_x L(x,\mu)$

4.4 d

Формулировка двойственной задачи:

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

Найдем первую и вторую производные двойственной функции:

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

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

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

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

Сильная двойственность выполняется.

5 Задача бинарного линейного программирования

5.1

Задачу можно переписать в виде

$$\begin{aligned}&\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}}\mathbf{x}\\ &x_i(x_i - 1) = 0, i = 1, ..., n\\ &\mathbf{Ax} \leqslant \mathbf{b}\end{aligned} \tag{5.1}$$

Лагранжиан запишется в виде

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

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

Лагранжиан — параболическая функция. Рассмотрим лагранжиан при разных параметрах.

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

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

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 инфинум функции от слагаемых, соответвующих этим i, равен нулю. Слагаемые, для которых выполнено второе условие, достигают минимума в точке $x_i = -\dfrac{c_i - \lambda_i + \sum\limits_j \mu_j a_{ji}}{2\lambda_i}$

Инфинум запишется в виде

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

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

5.2

Задача для непрерывной релаксации перепишется следующим образом:

$$\begin{aligned} &\min_{\mathbf{x}} \mathbf{c}^{\mathrm{T}} \mathbf{x} \\ &x_i(x_i - 1) \leqslant 0, i = 1, ..., n \\ &\mathbf{A}\mathbf{x} \leqslant \mathbf{b} \end{aligned} \tag{5.5}$$

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 λ_i, причём для $\tilde{\mu}_i$ условие $\tilde{\mu}_i \geqslant 0$ будет автоматически выполнено при поиске решения двойственной задачи. Соответственно, при $\lambda_i \geqslant 0$ максимумы двойственных функций совпадут друг с другом, и, иными словами, нижняя оценка релаксации Лагранжа совпадет с оценкой непрерывной релаксации.

5.3

6 Задача наименьших квадратов

6.1

Лагранжиан для задачи запишется так:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \boldsymbol{\lambda}^{\mathrm{T}} (\mathbf{G}\mathbf{x} - \mathbf{h}) \tag{6.1}$$

Производная лагранжиана по $\mathbf{x}$ такова

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

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

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

Двойственная задача запишется так:

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

Для исходной задачи выполнено условие Слейтера — 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

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

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

Матрица перед $\boldsymbol{\lambda}^*$ представляет собой произведение трех матриц: $\mathbf{G}$ размера $p \times n$ и ранга p, $(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}$ размера $n \times n$ и ранга n, $\mathbf{G}^{\mathrm{T}}$ размера $n \times p$ и ранга p. Итоговая матрица имеет размер $p \times p$. Если $p \leqslant n$, 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

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

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

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

7

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

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

Или, избавившись от лишних условий:

$$\begin{aligned} &\min(y_1) \\ \text{s.t.}\, &y_2 \geqslant 0 \\ &y_1 = 0 \end{aligned} \tag{7.2}$$