1 Выпуклые множества

1. $f(x) = \max\limits_{i=\overline{1,p}}(a_i x + b_i)$, где x, a_i, $b_i \in \mathbb{R}$

   По определению сопряженная функция вводится следующим образом:

   $$f^*(y) = \sup_x \; [xy - f(x)] = \sup_x \; \left[xy - \max_{i=\overline{1,p}}(a_i x + b_i)\right].$$

   $f(x)$ — поточечный максимум линейных функций, поэтому эта функция кусочно-линейная и является выпуклой. Найдем точки «переключения» между линейными функциями. Без ограничения общности предположим, что $a_i \leqslant a_{i+1}$, $i = \overline{1,p-1}$, всегда можно переупорядочить функции таким образом. Для определенности предположим также, что каждая из p линейных функций $a_i x + b_i$ в каких-то точках равна $f(x)$, то есть является максимумом, в противном случае уберем такие функции из рассмотрения.

   Используя описанные предположения, найдем точки «переключения» между линейными функциями в $f(x)$, а именно точки «переключения» между i и $i+1$ линейной функцией $\tilde{x}_i$, $i = \overline{1,p-1}$:

   $$a_i \tilde{x}_i + b_i = a_{i+1}\tilde{x}_i + b_{i+1}; \quad \tilde{x}_i = \frac{b_i - b_{i+1}}{a_{i+1} - a_i}.$$

   Из определения $f^*(y)$ можно видеть, что сопряженная функция будет ограничена только при $y \in [a_1; a_p]$, при этом максимум будет достигаться в точках «переключения» $\tilde{x}_i$ между функциями. Это можно понять, расписав $f^*(y)$ через индикаторные функции $\mathbb{1}$:

   $$f^*(y) = \sup_x \Bigg[x \left(y - a_1 \mathbb{1}(x \leqslant \tilde{x}_1) - a_p \mathbb{1}(\tilde{x}_{p-1} < x) - \sum_{i=2}^{p-1} a_i \mathbb{1}(\tilde{x}_{i-1} < x \leqslant \tilde{x}_i) \right) - \\ b_1 \mathbb{1}(x \leqslant \tilde{x}_1) - b_p \mathbb{1}(\tilde{x}_{p-1} < x) - \sum_{i=2}^{p-1} b_i \mathbb{1}(\tilde{x}_{i-1} < x \leqslant \tilde{x}_i) \Bigg].$$

   Действительно, для $y \in [a_1; a_p]$ максимум достигается в точках $\tilde{x}_i$, y вне этого интервала приводят к ненулевому линейному коэффициенту при x, поэтому функция будет неограничена сверху. Также этот факт виден из геометрической интерпретации сопряженной функции. Тогда общая формула $f^*(y)$ записывается в виде:

   $$f^*(y) = \begin{cases} \dfrac{b_i - b_{i+1}}{a_{i+1} - a_i}(y - a_i) - b_i, & y \in [a_i; a_{i+1}],\ i = \overline{1,p-1} \\ \infty, & \text{else.} \end{cases}$$

   Из итоговой формулы можно видеть, что сопряженная функция на своей области определения $\mathrm{dom}(f^*) = [a_1; a_p]$ также является кусочно-линейной.

2. $g(\mathbf{x}) = \inf_{\mathbf{x}_1 + \ldots + \mathbf{x}_k = \mathbf{x}}(f_1(\mathbf{x}_1) + \ldots + f_k(\mathbf{x}_k))$, где f_i — выпуклые функции.

   Распишем $g^*(\mathbf{x})$ по определению:

   $$g^*(\mathbf{y}) = \sup_{\mathbf{x}} \; \left[\mathbf{x}^{\mathsf{T}}\mathbf{y} - g(\mathbf{x})\right] = \sup_{\substack{\mathbf{x_1}, \ldots, \mathbf{x_k}, \\ \mathbf{x}_1 + \ldots + \mathbf{x}_k = \mathbf{x}}} \; \left[\left(\sum_{i=1}^{k} \mathbf{x_i}^{\mathsf{T}} \right) \mathbf{y} - \inf_{\mathbf{x}_1 + \ldots + \mathbf{x}_k = \mathbf{x}} (f_1(\mathbf{x}_1) + \ldots + f_k(\mathbf{x}_k)) \right].$$

Заметим, что $\inf_{\mathbf{x}}(h(\mathbf{x})) = -\sup_{\mathbf{x}}(-h(\mathbf{x}))$. Поэтому можно переписать формулу в более простой форме:

$$g^*(\mathbf{y}) = \sup_{\mathbf{x}_1,\ldots,\mathbf{x}_k} \left[\sum_{i=1}^{k} \mathbf{x_i}^\mathsf{T}\mathbf{y} - \sum_{i=1}^{k} f_i(\mathbf{x_i})\right] = \sum_{i=1}^{k} f_i^*(\mathbf{x}_i),$$

здесь было использовано определение сопряженных к f_i функций. Таким образом, сопряженная функция к инфимальной свертке функций — это сумма сопряженных функций.

3. $f(\mathbf{X}) = \mathrm{tr}(\mathbf{X}^{-1})$, $\mathrm{dom}(f) = \mathbf{S}^n_{++}$.

   Запишем определение сопряженной функции:

$$f^*(\mathbf{Y}) = \sup_{\mathbf{X}} \left[\mathrm{tr}(\mathbf{X}^\mathsf{T}\mathbf{Y}) - \mathrm{tr}(\mathbf{X}^{-1})\right].$$

   Найдем область определения сопряженной функции. Докажем, что $\mathrm{dom}(f^*) = -\mathbf{S}^n_{+}$. Запишем разложение $\mathbf{Y}$ по собственным векторам и приведем к диагональному виду $\mathbf{X}$ в базисе из собственных векторов $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{Q\Lambda Q}^\mathsf{T} = \sum_{i=1}^{n} \lambda_i \mathbf{q}_i \mathbf{q}_i^\mathsf{T}; \quad \mathbf{X} = \mathbf{QAQ}^\mathsf{T} = \sum_{i=1}^{n} \alpha_i \mathbf{q}_i \mathbf{q}_i^\mathsf{T},\ \alpha_i > 0.$$

   В этом случае под супремумом находится следующая функция:

$$\mathrm{tr}(\mathbf{X}^\mathsf{T}\mathbf{Y}) - \mathrm{tr}(\mathbf{X}^{-1}) = \mathrm{tr}(\mathbf{\Lambda A} - \mathbf{A}^{-1}) = \sum_{i=1}^{n} \lambda_i \alpha_i - \sum_{i=1}^{n} 1/\alpha_i.$$

   Из формулы можно видеть, что если хотя бы для одного i выполнено $\lambda_i > 0$ (то есть в спектре $\mathbf{Y}$ есть положительное собственное значение), то данная функция неограничена сверху, поэтому областью определения сопряженной функции является $\mathbf{Y} \in -\mathbf{S}^n_{+}$,

   В номере 3.1(f) первого задания было доказано, что $\mathrm{tr}(\mathbf{X}^{-1})$ является выпуклой функцией на $\mathbf{S}^n_{++}$. Кроме того, $\mathrm{tr}(\mathbf{X}^\mathsf{T}\mathbf{Y})$ — аффинная по $\mathbf{X}$ как линейная функция. Таким образом, $\mathrm{tr}(\mathbf{X}^\mathsf{T}\mathbf{Y}) - \mathrm{tr}(\mathbf{X}^{-1})$ является вогнутой функцией.

   Найдем супремум функции аналитически. Из вогнутости и дифференцируемости функции следует, что можно найти ее максимум, приравняв градиент функции нулю:

$$\frac{\partial(\mathrm{tr}(\mathbf{X}^\mathsf{T}\mathbf{Y}) - \mathrm{tr}(\mathbf{X}^{-1}))}{\partial \mathbf{X}} = \mathbf{Y} + \mathbf{X}^{-2} = \mathbf{0}, \quad \mathbf{X} = (-\mathbf{Y})^{-1/2}.$$

   Таким образом, сопряженная функция имеет вид:

$$f^*(\mathbf{Y}) = -2\ \mathrm{tr}((-\mathbf{Y})^{1/2}).$$

2 Условия оптимальности

1. Требуется решить следующую задачу:

$$\begin{gathered}\min_{\mathbf{x}}\ \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s.t. } \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \leqslant 1,\end{gathered}$$

где $\mathbf{A} \in \mathbf{S}^n_{++}$.

Так как $\mathbf{A} \in \mathbf{S}^n_{++}$, сделаем замену $\mathbf{z} = \mathbf{A}^{1/2}\mathbf{x}$ и обозначим $\mathbf{d} = -\mathbf{A}^{-1/2}\mathbf{c}$. Получаем эквивалентную задачу максимизации линейной функции на единичном шаре:

$$\begin{aligned} &\max_{\mathbf{z}} \ \mathbf{d}^\mathsf{T}\mathbf{z} \\ &\text{s.t. } \|\mathbf{z}\|_2^2 \leqslant 1. \end{aligned}$$

Данная задача является выпуклой и для нее можно было бы использовать условия ККТ, но получим решение $\mathbf{z}^*$ более простым способом, используя неравенство КБШ:

$$\mathbf{d}^\mathsf{T}\mathbf{z} \leqslant \|\mathbf{d}\|_2\|\mathbf{z}\|_2 \leqslant \|\mathbf{d}\|_2, \quad \mathbf{d}^\mathsf{T}\mathbf{z}^* = \|\mathbf{d}\|_2, \quad \mathbf{z}^* = \mathbf{d}/\|\mathbf{d}\|_2.$$

Таким образом, после обратной замены координат, получаем точку минимума исходной задачи $\mathbf{x}^*$ и значение целевой функции $\mathbf{c}^\mathsf{T}\mathbf{x}^*$ в ней:

$$\mathbf{x}^* = \frac{-1}{\|\mathbf{A}^{-1/2}\mathbf{c}\|_2}\mathbf{A}^{-1}\mathbf{c} = \frac{-1}{\sqrt{\mathbf{c}^\mathsf{T}\mathbf{A}^{-1}\mathbf{c}}}\mathbf{A}^{-1}\mathbf{c}; \quad \mathbf{c}^\mathsf{T}\mathbf{x}^* = -\sqrt{\mathbf{c}^\mathsf{T}\mathbf{A}^{-1}\mathbf{c}}.$$

2. Требуется решить следующую оптимизационную задачу:

$$\begin{aligned} &\min_{\mathbf{X}\in\mathbf{S}^n_{++}} \ \mathrm{tr}(\mathbf{X}) - \log(\det(\mathbf{X})) \\ &\text{s.t. } \mathbf{X}\mathbf{z} = \mathbf{y}, \end{aligned}$$

где $\mathbf{y}$,$\mathbf{z} \in \mathbb{R}^n$ и $\mathbf{y}^\mathsf{T}\mathbf{z} = 1$.

Данная задача выпуклая. Действительно, целевая функция является суммой аффинной функции (след) и вогнутой функции (на лекции было доказана выпуклость логарифма определителя матрицы, здесь перед ним стоит минус). Функциональные ограничения в задача линейные и область определения является выпуклым множеством ($\mathbf{S}^n_{++}$). Запишем функцию Лагранжа для данной задачи:

$$L(\mathbf{X}, \lambda) = \mathrm{tr}(\mathbf{X}) - \log(\det(\mathbf{X})) + \lambda^\mathsf{T}(\mathbf{X}\mathbf{z} - \mathbf{y}),\ \lambda \in \mathbb{R}^n$$

Выпишем условия ККТ:

$$\begin{aligned} &\frac{\partial L(\mathbf{X}, \lambda)}{\partial \mathbf{X}} = \mathbf{0}; \\ &\mathbf{X}\mathbf{z} = \mathbf{y}; \\ &\mathbf{X} \succ 0. \end{aligned}$$

Найдем градиенты каждого из слагаемых в функции Лагранжа:

$$\begin{aligned} &\frac{\partial}{\partial \mathbf{X}}\log(\det(\mathbf{X})) = \frac{1}{\det \mathbf{X}}\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} = \frac{1}{\det \mathbf{X}}\mathrm{adj}(\mathbf{X})^\mathsf{T} = \mathbf{X}^{-\mathsf{T}} \\ &\frac{\partial}{\partial \mathbf{X}}(\lambda^\mathsf{T}(\mathbf{X}\mathbf{z} - \mathbf{y})) = \frac{1}{2}(\lambda\mathbf{z}^\mathsf{T} + \mathbf{z}\lambda^\mathsf{T}) \\ &\frac{\partial}{\partial \mathbf{X}}\mathrm{tr}(\mathbf{X}) = \mathbf{I}. \end{aligned}$$

Таким образом, условия на стационарную точку функции Лагранжа переписывается в виде:

$$\frac{\partial L(\mathbf{X}, \lambda)}{\partial \mathbf{X}} = \mathbf{I} - \mathbf{X}^{-\mathsf{T}} + \frac{1}{2}(\lambda\mathbf{z}^\mathsf{T} + \mathbf{z}\lambda^\mathsf{T}) = \mathbf{0}; \quad \mathbf{X}^{-1} = \mathbf{I} + \frac{1}{2}(\lambda\mathbf{z}^\mathsf{T} + \mathbf{z}\lambda^\mathsf{T}).$$

Из ограничения равенства можно выразить $\mathbf{z} = \mathbf{X}^{-1}\mathbf{y}$. Используем это выражение, домножив формулу $\mathbf{X}^{-1}$ справа на $\mathbf{y}$. Также учтем условия равенства единице скалярного произведения векторов $\mathbf{z}$ и $\mathbf{y}$, домножив после этого полученное равенство слева на $\mathbf{y}^\mathsf{T}$:

$$\mathbf{z} = \mathbf{y} + \frac{1}{2}\lambda + \frac{1}{2}\mathbf{z}\lambda^\mathsf{T}\mathbf{y}; \quad 1 = \mathbf{y}^\mathsf{T}\mathbf{z} = \|\mathbf{y}\|_2^2 + \lambda^\mathsf{T}\mathbf{y}; \quad \lambda^\mathsf{T}\mathbf{y} = 1 - \|\mathbf{y}\|_2^2.$$

Подставим формулу для $\lambda^\mathsf{T}\mathbf{y}$ в исходное равенство на $\mathbf{z}$ и выразим из него множитель Лагранжа λ:

$$\mathbf{z} = \mathbf{y} + \frac{1}{2}\lambda + \frac{1}{2}(1 - \|\mathbf{y}\|_2^2)\mathbf{z}; \quad \lambda = -2\mathbf{y} + (1 + \|\mathbf{y}\|_2^2)\mathbf{z}.$$

Таким образом, подставив λ в выражение на $\mathbf{X}^{-1}$, получаем явный вид обратной матрицы предполагаемого решения задачи:

$$\mathbf{X}^{-1} = \mathbf{I} + \frac{1}{2}\left(1 + \|\mathbf{y}\|_2^2\right)\mathbf{z}\mathbf{z}^\mathsf{T} - \mathbf{y}\mathbf{z}^\mathsf{T} - \mathbf{z}\mathbf{y}^\mathsf{T}.$$

Теперь требуется найти исходную матрицу $\mathbf{X}$. Догадаемся до решения $\tilde{\mathbf{X}}$ и докажем, что оно действительно является $\mathbf{X}$ проверкой $\mathbf{X}\mathbf{X}^{-1} = \mathbf{I}$.

$$\begin{aligned}
\tilde{\mathbf{X}} &= \mathbf{I} + \mathbf{y}\mathbf{y}^\mathsf{T} - \frac{1}{\|\mathbf{z}\|_2^2}\mathbf{z}\mathbf{z}^\mathsf{T}; \\
\tilde{\mathbf{X}}\mathbf{X}^{-1} &= \left(\mathbf{I} + \mathbf{y}\mathbf{y}^\mathsf{T} - \frac{1}{\|\mathbf{z}\|_2^2}\mathbf{z}\mathbf{z}^\mathsf{T}\right)\left(\mathbf{I} + \frac{1}{2}\left(1 + \|\mathbf{y}\|_2^2\right)\mathbf{z}\mathbf{z}^\mathsf{T} - \mathbf{y}\mathbf{z}^\mathsf{T} - \mathbf{z}\mathbf{y}^\mathsf{T}\right) \\
&= \mathbf{I} + \frac{1}{2}\left(1 + \|\mathbf{y}\|_2^2\right)\mathbf{z}\mathbf{z}^\mathsf{T} - \mathbf{y}\mathbf{z}^\mathsf{T} - \mathbf{z}\mathbf{y}^\mathsf{T} + \\
&\quad \mathbf{y}\mathbf{y}^\mathsf{T} + \frac{1}{2}\left(1 + \|\mathbf{y}\|_2^2\right)\mathbf{y}\mathbf{y}^\mathsf{T}\mathbf{z}\mathbf{z}^\mathsf{T} - \mathbf{y}\mathbf{y}^\mathsf{T}\mathbf{y}\mathbf{z}^\mathsf{T} - \mathbf{y}\mathbf{y}^\mathsf{T}\mathbf{z}\mathbf{y}^\mathsf{T} - \\
&\quad \frac{1}{\|\mathbf{z}\|_2^2}\mathbf{z}\mathbf{z}^\mathsf{T} - \frac{1}{2}\left(1 + \|\mathbf{y}\|_2^2\right)\frac{1}{\|\mathbf{z}\|_2^2}\mathbf{z}\mathbf{z}^\mathsf{T}\mathbf{z}\mathbf{z}^\mathsf{T} + \frac{1}{\|\mathbf{z}\|_2^2}\mathbf{z}\mathbf{z}^\mathsf{T}\mathbf{y}\mathbf{z}^\mathsf{T} + \frac{1}{\|\mathbf{z}\|_2^2}\mathbf{z}\mathbf{z}^\mathsf{T}\mathbf{z}\mathbf{y}^\mathsf{T} \\
&= \mathbf{I},
\end{aligned}$$

что и требовалось доказать. Таким образом, $\tilde{\mathbf{X}} = \mathbf{X}$. Осталось показать, что $\mathbf{X} \in \mathbf{S}^n_{++}$ (удовлетворяет рассматриваемому множеству симметричных положительно определенных матриц). Действительно, используем $\mathbf{z}^\mathsf{T}\mathbf{y} = 1$ и введем матрицу $\mathbf{B} = \mathbf{I} + \mathbf{y}\mathbf{z}^\mathsf{T}/\|\mathbf{z}\|_2 - \mathbf{z}\mathbf{z}^\mathsf{T}/\|\mathbf{z}\|_2^2$. Для нее справедливо выражение: $\mathbf{X} = \mathbf{B}\mathbf{B}^\mathsf{T}$ то есть $\mathbf{X} \succeq 0$ (заметим, что область определения сопряженных функций замкнута по определению). Таким образом, выполнены условия ККТ и условия регулярности (нашли допустимую точку из относительной внутренности), то есть $\mathbf{X}$ является решением: $\mathbf{X}^* = \mathbf{I} + \mathbf{y}\mathbf{y}^\mathsf{T} - \frac{1}{\|\mathbf{z}\|_2^2}\mathbf{z}\mathbf{z}^\mathsf{T}$.

3. Требуется привести алгоритм получения решения с оценкой его асимптотической сложности ждя следующей задачи:

$$\begin{aligned}
&\min_{\mathbf{x}} \ \|\mathbf{x} - \mathbf{y}\|_2^2 \\
&\text{s.t.} \ \sum_{i=1}^n x_i = 1; \\
&\quad\ x_i \geqslant 0.
\end{aligned}$$

Иными словами, требуется найти евклидову проекцию точки $\mathbf{y} \in \mathbb{R}^n$ на вероятностный симплекс. Запишем функцию Лагранжа:

$$L(\mathbf{x},\lambda,\mu) = \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda(\mathbf{1}^\mathsf{T}\mathbf{x} - 1) - \mu^\mathsf{T}\mathbf{x}, \quad \mu \geqslant 0.$$

Задача, очевидно, является выпуклой. В литературе предлагается решать задачу, используя часть функции Лагранжа (без неравенств), выделяя отдельно условие $\mathbf{x} \geqslant 0$:

$$\tilde{L}(\mathbf{x},\lambda) = \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda(\mathbf{1}^\mathsf{T}\mathbf{x} - 1), \quad \mathbf{x} \geqslant 0.$$

Запишем двойственную функцию, учитывая явное условие неотрицательности $\mathbf{x} \geqslant 0$:

$$g(\lambda) = \inf_{\mathbf{x}\geqslant 0} \tilde{L}(\mathbf{x},\lambda) = \inf_{\mathbf{x}\geqslant 0} \sum_{i=1}^{n} \left(\frac{1}{2}(x_i - y_i)^2 + \lambda x_i \right) - \lambda.$$

Получили, что оптимизируемая функция разложилась на сумму функций одного аргумента x_i: $g_i(\lambda) = \inf\limits_{x_i \geqslant 0} \frac{1}{2}(x_i - y_i)^2 + \lambda x_i$. Таким образом, инфимум достигается в точке $x^* \in \mathbb{R}^n$, равной:

$$x_i^* = (y_i - \lambda)_+.$$

Подставим x^* в «усеченную» функцию Лагранжа $\tilde{L}(\mathbf{x},\lambda)$, получим явный вид двойственной функции:

$$g(\lambda) = \sum_{i=1}^{n} \left(\frac{1}{2}((y_i - \lambda)_+ - y_i)^2 + \lambda(y_i - \lambda)_+ \right) - \lambda.$$

Для решения двойственной задачи требуется максимизировать $g(\lambda)$ по скалярному аргументу λ: $\max_\lambda \ g(\lambda)$ при условии равенства из прямой задачи. Известно, что двойственная функция всегда выпуклая. Однако наличие функции $(\cdot)_+$ делает задачу недифференцируемой. Поэтому ее требуется решать безградиентыми методами, например, методом секущих. Пусть решением является λ_i. Решение прямой задачи (выполняется условие регулярности Слейтера в выпуклой задаче, имеет место сильная двойственность) будет равным $x_i^* = (y_i - \lambda^*)_+$, $i = \overline{1,n}$ (проекция на шар). Сложность такого решения будет зависеть не менее чем линейно от размерности n (сепарабельность функций, линейно растет число операций) и логарифмически от точности решения (как метод секущих).

3 Двойственные задачи

1. Бинарная задача линейного программирования имеет вид:

$$\begin{aligned} &\min_{\mathbf{x}} \ \mathbf{c}^\mathsf{T}\mathbf{x} \\ \text{s.t. } &\mathbf{A}\mathbf{x} \leqslant \mathbf{b}; \\ &x_i \in \{0; 1\}, \, i = \overline{1,n}. \end{aligned}$$

Данная задача не является выпуклой, имеет дискретную область определения, поэтому требует перебора всех возможных наборов $\mathbf{x}$, их число равно 2^n, то есть растет экспоненциально с ростом размерности. Рассмотрим две релаксации данной задачи.

Данную задачу можно эквивалентно переписать в виде:

$$\begin{aligned}&\min_{\mathbf{x}}\ \mathbf{c}^\mathsf{T}\mathbf{x}\\ \text{s.t. }&\mathbf{A}\mathbf{x}\leqslant\mathbf{b};\\ &x_i(1-x_i)=0,\ i=\overline{1,n}.\end{aligned}$$

Рассмотрим (1) релаксацию Лагранжа для данной задачи дискретной оптимизации: построение двойственной задачи к исходной в непрерывном пространстве. Докажем, что ее решение совпадает с решением непрерывной релаксации (2) исходной задачи:

$$\begin{aligned}&\min_{\mathbf{x}}\ \mathbf{c}^\mathsf{T}\mathbf{x}\\ \text{s.t. }&\mathbf{A}\mathbf{x}\leqslant\mathbf{b};\\ &x_i\in[0;1],\ i=\overline{1,n}.\end{aligned}$$

Также докажем, что обе задачи дают оценку снизу на решение исходной.

Запишем функцию лагранжа исходной задачи:

$$L(\mathbf{x},\lambda,\mu)=\mathbf{c}^\mathsf{T}\mathbf{x}+\mu^\mathsf{T}(\mathbf{A}\mathbf{x}-\mathbf{b})+\sum_{i=1}^n\lambda_i(x_i^2-x_i)=(\mathbf{c}+\mathbf{A}^\mathsf{T}\mu-\lambda)^\mathsf{T}\mathbf{x}-\mathbf{b}^\mathsf{T}\mu+\sum_{i=1}^n\lambda_i x_i^2,\quad \mu\geqslant 0.$$

Найдем двойственную функцию $g(\lambda,\mu)=\inf\limits_{\mathbf{x}} L(\mathbf{x},\lambda,\mu)$:

$$g(\lambda,\mu)=\inf_{\mathbf{x}}\left((\mathbf{c}+\mathbf{A}^\mathsf{T}\mu-\lambda)^\mathsf{T}\mathbf{x}-\mathbf{b}^\mathsf{T}\mu+\sum_{i=1}^n\lambda_i x_i^2\right),$$

Данная функция неограничена снизу, если $\exists i\colon \lambda_i<0$ (квадратичная форма будет неограничена снизу). Иначе получаем задачу минимизации квадратичной функции с положительно определенной матрицей: $\mathbf{x}^\mathsf{T}\mathbf{D}\mathbf{x}+\mathbf{h}^\mathsf{T}\mathbf{x}+\mathbf{v}$, где $\mathbf{D}=\operatorname{diag}(\lambda)$, $\mathbf{h}=\mathbf{c}+\mathbf{A}^\mathsf{T}\mu-\lambda$, $\mathbf{v}=-\mathbf{b}^\mathsf{T}\mu$. Данная функция дифференцируемая и сильно выпуклая (было доказано в 1.1 первого задания), ее решение может быть найдено по критерию первого порядка, точка минимума равна: $\tilde{\mathbf{x}}=-\frac{1}{2}\mathbf{D}^{-1}\mathbf{h}$, после подстановки находим значение в минимуме: $-\frac{1}{4}\mathbf{h}^\mathsf{T}\mathbf{D}^{-1}\mathbf{h}+\mathbf{v}$. Тогда мы можем сразу записать явный вид двойственной функции:

$$g(\lambda,\mu)=\begin{cases}-\frac{1}{4}\sum\limits_{i=1}^n\frac{1}{\lambda_i}(\mathbf{a}_i^\mathsf{T}\mu+c_i-\lambda_i)^2-\mathbf{b}^\mathsf{T}\mu, & \lambda\geqslant 0;\\ -\infty, & \text{else.}\end{cases}$$

запишем двойственную задачу:

$$\sup_{\mu\geqslant 0}\ g(\lambda,\mu)=\sup_{\mu\geqslant 0}\left(-\frac{1}{4}\sum_{i=1}^n\frac{1}{\lambda_i}(\mathbf{a}_i^\mathsf{T}\mu+c_i-\lambda_i)^2-\mathbf{b}^\mathsf{T}\mu\right),$$

где $\mathbf{a}_i$ — i-ый столбец матрицы A. Можно видеть, что данная задача распадается в сумму n функций, зависящих только от λ_i и μ_i, $i=\overline{1,n}$ соответственно. Относительно множителя Лагранжа $\lambda\geqslant 0$ можно записать аналитическое решение — это покоординатная минимизация $-\frac{1}{4}(\mathbf{a}_i^\mathsf{T}\mu+c_i-\lambda_i)^2/\lambda_i$ относительно $\lambda_i\geqslant 0$. Данная функция является выпуклой для неотрицательных λ_i, при $\mathbf{a}_i^\mathsf{T}\mu + c_i \leqslant 0$ точкой минимума является $\lambda_i = -\mathbf{a}_i^\mathsf{T}\mu - c_i$, иначе $\lambda_i = \mathbf{a}_i^\mathsf{T}\mu + c_i$. Получаем, что максимумом является $\min(0; \mathbf{a}_i^\mathsf{T}\mu + c_i)$, $i = \overline{1,n}$. Таким образом, двойственная задача может быть записана в виде:

$$\sup_{\mu \geqslant 0} \sum_{i=1}^{n} \min(0; \mathbf{a}_i^\mathsf{T}\mu + c_i) - \mathbf{b}^\mathsf{T}\mu.$$

Теперь рассмотрим (2) — непрерывную релаксацию задачи. Запишем для нее функцию Лагранжа (имеем три ограничения-неравенства):

$$\tilde{L}(\mathbf{x}, \alpha, \beta, \gamma) = \mathbf{c}^\mathsf{T}\mathbf{x} + \alpha^\mathsf{T}(\mathbf{A}\mathbf{x} - \mathbf{b}) - \beta^\mathsf{T}\mathbf{x} + \gamma^\mathsf{T}(\mathbf{x} - \mathbf{1}) = (\mathbf{A}^\mathsf{T}\alpha - \beta + \gamma + \mathbf{c})^\mathsf{T}\mathbf{x} - \mathbf{b}^\mathsf{T}\alpha - \sum_{i=1}^{n} \gamma_i.$$

Данная функция является аффинной, поэтому ее инфимум по $\mathbf{x}$ равен $-\infty$ всюду, кроме $\mathbf{A}^\mathsf{T}\alpha + \mathbf{c} - \beta + \gamma = 0$, здесь двойственная функция имеет вид: $\tilde{g}(\alpha, \beta, \gamma) = -\mathbf{b}^\mathsf{T}\alpha - \sum\limits_{i=1}^{n} \gamma_i$. Двойственная задача записывается следующим образом:

$$\begin{aligned} &\max_{\alpha, \beta, \gamma} \; -\mathbf{b}^\mathsf{T}\alpha - \sum_{i=1}^{n} \gamma_i \\ \text{s.t. } &\mathbf{A}^\mathsf{T}\alpha - \beta + \gamma + \mathbf{c} = 0; \\ &\alpha \geqslant 0; \\ &\beta \geqslant 0; \\ &\gamma \geqslant 0. \end{aligned}$$

Можно видеть, что две данные задачи являются эквивалентными с точностью до переобозначений: $\alpha = \mu \geqslant 0$, $\gamma_i = -\min(0; \mathbf{a}_i^\mathsf{T}\mu + c_i) \geqslant 0$, $\beta = (\mathbf{a}_i^\mathsf{T}\mu + c_i)_+ \geqslant 0$. Решение данных двойственных задач является оценкой снизу для решения прямой задачи бинарного линейного программирования из свойств двойственных задач ($\mathbf{p}^* \geqslant \mathbf{d}^*$).

2. Требуется построить двойственную задачу для следующей задачи:

$$\min_{\mathbf{x}} \max_{i=\overline{1,k}} (\mathbf{a}_i^\mathsf{T}\mathbf{x} + b_i), \quad \mathbf{x} \in \mathbb{R}^n.$$

Запишем эквивалентную оптимизационную задачу, перенеся аффинные функции из целевой функции в ограничения-равенства:

$$\begin{aligned} &\min_{\mathbf{z}, \mathbf{x}} \; \max_{i=\overline{1,k}} \; z_i \\ \text{s.t. } &\mathbf{a}_i^\mathsf{T}\mathbf{x} + b_i = z_i, \quad i = \overline{1,k}. \end{aligned}$$

Запишем для полученной переформулированной задачи двойственную функцию $g(\lambda)$:

$$g(\lambda) = \inf_{\mathbf{x}, \mathbf{z}} \left(\max_{i=\overline{1,k}} \; z_i + \sum_{i=1}^{k} \lambda_i (\mathbf{a}_i^\mathsf{T}\mathbf{x} + b_i - z_i) \right) = \inf_{\mathbf{x}, \mathbf{z}} \left(\max_{i=\overline{1,k}} \; z_i - \lambda^\mathsf{T}\mathbf{z} + \left(\sum_{i=1}^{k} \lambda_i \mathbf{a}_i \right)^\mathsf{T} \mathbf{x} + \lambda^\mathsf{T}\mathbf{b} \right).$$

По переменной $\mathbf{x}$ данная функция является аффинной, поэтому она ограничена снизу по $\mathbf{x}$ только для $\sum\limits_{i=1}^{k} \lambda_i \mathbf{a}_i = 0$.

По переменной $\mathbf{z}$ минимизируется функция $\max\limits_{i=\overline{1,k}} \ z_i - \lambda^\mathsf{T} \mathbf{z}$. Она ограничена снизу только в случае $\sum_{i=1}^k \lambda_i = 1$ и $\lambda_i \geqslant 0$, $i = \overline{1,k}$. Действительно, при $\sum_{i=1}^k \lambda_i \neq 1$ и выборе $\lambda_i = \alpha$, функция $\alpha(1 - \sum_{i=1}^k \lambda_i)$ неограничена снизу. При выборе же $\lambda_j < 0$ для какого-то j можно выбрать y_j самым большим по модулю отрицательным числом и неограниченно уменьшать его, поэтому двойственная функция также будет неограниченной снизу. При выполнении условий $\sum_{i=1}^k \lambda_i = 1$ и $\lambda_i \geqslant 0$, $i = \overline{1,k}$ минимум функции, очевидно, равен нулю (достигается, например, при $\mathbf{z} = \mathbf{0}$). Таким образмо было получено, что двойственная функция ограничена снизу при $\sum\limits_{i=1}^{k} \lambda_i \mathbf{a}_i = 0$, $\sum_{i=1}^k \lambda_i = 1$ и $\lambda_i \geqslant 0$, $i = \overline{1,k}$ и равна $\lambda^\mathsf{T} \mathbf{b}$.

Теперь все готово для того, чтобы записать двойственную задачу:

$$\begin{aligned} &\max_{\lambda} \ \lambda^\mathsf{T} \mathbf{b} \\ \text{s.t. } &\sum_{i=1}^{k} \lambda_i \mathbf{a}_i = 0; \\ &\sum_{i=1}^{k} \lambda_i = 1; \\ &\lambda \geqslant 0. \end{aligned}$$

3. Требуется построить двойственную задачу для следующей задачи:

$$\begin{aligned} &\min_{\mathbf{X} \in \mathbf{S}^n_{++}} \ -\log\det \mathbf{X} \\ \text{s.t. } &\mathbf{a}_i^\mathsf{T} \mathbf{X} \mathbf{a}_i \leqslant 1, \quad i = \overline{1,m}. \end{aligned}$$

Заметим, что ограничения-неравенства являются аффинными относительно $\mathbf{X}$: $\mathbf{a}_i^\mathsf{T} \mathbf{X} \mathbf{a}_i = \mathrm{tr}(\mathbf{a}_i^\mathsf{T} \mathbf{X} \mathbf{a}_i) = \mathrm{tr}(\mathbf{a}_i \mathbf{a}_i^\mathsf{T} \mathbf{X}) \leqslant 1$. Запишем определение двойственной функции:

$$g(\lambda) = \inf_{\mathbf{X}} \ -\log\det \mathbf{X} + \sum_{i=1}^{m} \lambda_i (\mathrm{tr}(\mathbf{a}_i \mathbf{a}_i^\mathsf{T} \mathbf{X}) - 1).$$

На лекции было доказано, что для задач с аффинными ограничениями двойственная функция выражается через сопряженную целевую функцию. Приведем здесь краткое доказательство. Запишем прямую задачу:

$$\begin{aligned} &\min_{\mathbf{x}} \ f_0(\mathbf{x}) \\ \text{s.t. } &\mathbf{C}\mathbf{x} \leqslant \mathbf{d}; \end{aligned}$$

Двойственная функция для этой задачи будет равна: $g(\lambda) = \inf\limits_{\mathbf{x}} \big((\mathbf{C}^\mathsf{T} \mu)^\mathsf{T} \mathbf{x} - \mu^\mathsf{T} \mathbf{b} + f_0(\mathbf{x}) \big) = -\sup\limits_{\mathbf{x}} \big((-\mathbf{C}^\mathsf{T} \mu)^\mathsf{T} \mathbf{x} - f_0(\mathbf{x}) \big) - \mu^\mathsf{T} \mathbf{b} = -f_0^*(-\mathbf{C}^\mathsf{T} \mu) - \mu^\mathsf{T} \mathbf{b}$, где f_0^* — сопряженная к f_0 функция.

Найдем сопряженную к $f(\mathbf{X}) = \log\det\mathbf{X}^{-1}$ функцию на области определения $\mathbf{X} \in \mathbf{S}^n_{++}$. Запишем определение:

$$f^*(\mathbf{Y}) = \sup_{\mathbf{X}\in\mathbf{S}^n_{++}} \left(\mathrm{tr}(\mathbf{Y}^\mathsf{T}\mathbf{X}) - \log\det\mathbf{X}^{-1}\right) = \sup_{\mathbf{X}\in\mathbf{S}^n_{++}} \left(\mathrm{tr}(\mathbf{Y}^\mathsf{T}\mathbf{X}) + \log\det\mathbf{X}\right).$$

Докажем, что данная функция ограничена сверху только на $\mathbf{Y} \in -\mathbf{S}^n_{++}$. Действительно, допустим, что существует собственный вектор $\mathbf{q}_i$ единичной длины, которому соответствует неотрицательное собственное значение λ_i. Рассмотрим матрицу вида $\mathbf{X} = \mathbf{I} + \alpha\mathbf{q}_i\mathbf{q}_i^\mathsf{T}$, используя лемму об определителе можем получить:

$$\mathrm{tr}(\mathbf{Y}^\mathsf{T}\mathbf{X}) + \log\det\mathbf{X} = \mathrm{tr}\mathbf{Y} + \alpha\lambda_i + \log(1+\alpha) \to \infty, \quad \alpha \to \infty.$$

На лекции было доказано, что $\log\det\mathbf{X}$ является выпуклой функцией, $\mathrm{tr}(\mathbf{Y}^\mathsf{T}\mathbf{X})$ — аффинная функция, поэтому функция под супремумом является выпуклой, к тому же она дифференцируемая. Найдем явный вид сопряженной функции, используя условие оптимальности первого порядка (заметим, что по условию $\mathbf{X} \in \mathbf{S}^n_{++}$):

$$\begin{aligned}
&\frac{\partial}{\partial\mathbf{X}}\left(\mathrm{tr}(\mathbf{Y}^\mathsf{T}\mathbf{X}) + \log\det\mathbf{X}\right) = \mathbf{Y} + \mathbf{X}^{-1} = 0; \quad \mathbf{X} = \mathbf{Y}^{-1};\\
&f^*(\mathbf{Y}) = -\mathrm{tr}(\mathbf{I}) + \log\det((-\mathbf{Y})^{-1}) = \log\det((-\mathbf{Y})^{-1}) - n, \quad \mathbf{Y} \in -\mathbf{S}^n_{++}.
\end{aligned}$$

Используя только что доказанные утверждения, можем сразу записать двойственную функцию $g(\lambda)$:

$$g(\lambda) = \begin{cases} \log\det\left(\sum_{i=1}^m \lambda_i\mathbf{a}_i\mathbf{a}_i^\mathsf{T}\right) - \sum_{i=1}^m \lambda_i + n, & \sum\limits_{i=1}^m \lambda_i\mathbf{a}_i\mathbf{a}_i \in \mathbf{S}^n_{++};\\ -\infty, \quad \text{else.} \end{cases}$$

Таким образом, получена двойственная задача:

$$\begin{aligned}
&\max_\lambda \ \log\det\left(\sum_{i=1}^m \lambda_i\mathbf{a}_i\mathbf{a}_i^\mathsf{T}\right) - \sum_{i=1}^m \lambda_i + n\\
\text{s.t. } &\sum_{i=1}^m \lambda_i\mathbf{a}_i\mathbf{a}_i \succ 0;\\
&\lambda \geqslant 0.
\end{aligned}$$

4. -

5. Требуется построить двойственную задачу для следующей задачи:

$$\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\|_2^2 + \sum_{i=1}^p \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2, \quad \mathbf{x}, \mathbf{x}_0 \in \mathbb{R}^n, \mathbf{b}_i \in \mathbb{R}^{m_i}, \mathbf{A} \in \mathbb{R}^{m_i\times n}$$

Введем переменные $\mathbf{z}_i = \mathbf{A}_i\mathbf{x} + \mathbf{b}_i$ и рассмотрим эквивалентную задачу:

$$\begin{aligned}
&\min_{\mathbf{x}} \ \frac{1}{2}\|\mathbf{x} - \mathbf{x}_0\|_2^2 + \sum_{i=1}^p \|\mathbf{z}_i\|_2\\
&\text{s.t. } \mathbf{z}_i - \mathbf{A}_i\mathbf{x} - \mathbf{b}_i = 0.
\end{aligned}$$

Для данной задачи запишем функцию Лагранжа:

$$L(\mathbf{x},\mathbf{z}_1,\ldots,\mathbf{z}_p,\lambda_1,\ldots,\lambda_p) = \frac{1}{2}\|\mathbf{x}-\mathbf{x}_0\|_2^2 + \sum_{i=1}^{p}\|\mathbf{z}_i\|_2 + \sum_{i=1}^{p}\lambda_i^{\mathsf{T}}(\mathbf{z}_i - \mathbf{A}_i\mathbf{x} - \mathbf{b}_i).$$

Найдем двойственную функцию $g(\lambda_1,\ldots,\lambda_p)$. Можно заметить, что лагранжиан распадается на сумму функций, которые можно оптимизировать отдельно.

Найдем минимальное значение по $\mathbf{x}$. Так как функция Лагранжа является выпуклой относительно $\mathbf{x}$ (как сумма аффинной и второй нормы вектора) и дифференцируемой, воспользуемся свойством оптимальности первого порядка и получим $\mathbf{x} = \mathbf{x}_0 + \sum_{i=1}^{p}\mathbf{A}_i^{\mathsf{T}}\lambda_i$. Минимальное значение по $\mathbf{z}_i$ можно найти следующим образом: каждое слагаемое имеет вид $\|\mathbf{z}_i\|_2 + \lambda_i^{\mathsf{T}}\mathbf{z}_i$. При $\|\lambda_i\|_2 \leqslant 1$ из КБШ следует, что $\|\mathbf{z}_i\|_2 + \lambda_i^{\mathsf{T}}\mathbf{z}_i \geqslant \|\mathbf{z}_i\|_2 \cdot \|\lambda_i\|_2 + \lambda_i^{\mathsf{T}}\mathbf{z}_i \geqslant 0$, то есть минимальное значение равно 0 и достигается в $\mathbf{z}_i = 0$. При $\|\lambda_i\|_2 > 1$ функция Лагранжа неограничена снизу (при выборе $\mathbf{z}_i = \alpha\lambda_i$, $\alpha \to -\infty$ значение функции стремится к $-\infty$).

Таким образом, можем записать двойственную задачу:

$$\begin{aligned}
&\max_{\lambda_1,\ldots,\lambda_p} \; -\frac{1}{2}\|\sum_{i=1}^{p}\mathbf{A}_i^{\mathsf{T}}\lambda_i\|_2^2 - \sum_{i=1}^{p}(\mathbf{A}_i\mathbf{x}_0 + \mathbf{b}_i)^{\mathsf{T}}\lambda_i \\
&\text{s.t. } \|\lambda_i\|_2 \leqslant 1,\ i = \overline{1,p}
\end{aligned}$$