Методы оптимизации, ВМК, осень 2018

**Семинар 9: Эквивалентные преобразования задач. Стандартные классы выпуклых задач.**

# 1 Эквивалентные преобразования задач

**Терминология:** *Выпуклой задачей оптимизации* называется любая задача минимизации выпуклой функции $f : E \to \mathbb{R}$ на некотором выпуклом множестве $Q \subseteq E$: $\min_{x \in Q} f(x)$.

Одно из полезных применений условных задач: если исходная задача негладкая, то на практике довольно часто такую задачу можно переписать в эквивалентном гладком виде, с дополнительно введёнными переменными и условиями:

$$\min_{x \in \mathbb{R}^n, Ax=b} \|x\|_1 \qquad \Longleftrightarrow \qquad \min_{y,z \in \mathbb{R}^n_+, Ay-Az=b} \sum_i y_i + \sum_i z_i.$$

**Пример 1.1.** Рассмотрим задачу минимизации гладкой функции с $l^1$-регуляризатором:

$$\min_x f(x) + \|x\|_1.$$

Эта задача является выпуклой безусловной, но негладкой из-за присутствия $l^1$-нормы $\|x\|_1$. Превратим ее в выпуклую гладкую условную. Для этого введем дополнительные переменные $x^+$ и $x^-$, что

$$x_i^+ = \max\{x_i, 0\}, \qquad x_i^- = \max\{-x_i, 0\}.$$

Заметим, что из определения $x^+, x^- \succeq 0$. Тогда в терминах новых переменных

$$x_i = x_i^+ - x_i^-, \qquad |x_i| = x_i^+ + x_i^-.$$

и задача переписывается следующим образом:

$$\min_{x^+, x^-} f(x^+ - x^-) + \langle 1_n, x^+ \rangle + \langle 1_n, x^- \rangle \qquad \text{s. t.} \quad x^+, x^- \succeq 0.$$

(Почему эта задача эквивалентна исходной? Почему она выпуклая?)

Особенно полезным оказывается следующий прием, который называется *переформулировкой через надграфик*:

$$\min_x f(x) \quad \text{s. t. } x \in E \qquad \Longleftrightarrow \qquad \min_{x,t} t \quad \text{s. t. } t \geq f(x),\ x \in E,$$

где $f : E \to \mathbb{R}$ — функция. Такое преобразование задачи позволяет опустить целевую функцию в ограничения; при этом новая задача эквивалентна старой, а если старая является выпуклой, то и новая тоже будет выпуклой.

**Замечание 1.2.** Таким образом, целевую функцию всегда можно перенести в ограничения и получить эквивалентную задачу с линейной целевой функцией.

**Пример 1.3.** Пусть $f_1, \dots, f_m : \mathbb{R}^n \to \mathbb{R}$ — гладкие выпуклые функции. Рассмотрим задачу

$$\min_{x \in \mathbb{R}^n} \max_{1 \leq i \leq m} f_i(x).$$

Эта задача является выпуклой безусловной, но негладкой (из-за присутствия максимума). Используя переформулировку через надграфик, получаем эквивалентную задачу

$$\min_{(x,t) \in \mathbb{R}^n \times \mathbb{R}} t \qquad \text{s. t. } t \geq \max_{1 \leq i \leq m} f_i(x).$$

Эта задача имеет гладкую целевую функцию, но негладкое ограничение. Однако, поскольку

$$t \geq \max_{1 \leq i \leq m} f_i(x) \qquad \Longleftrightarrow \qquad t \geq f_i(x) \text{ для всех } 1 \leq i \leq m,$$

получаем, что одно негладкое ограничение можно эквивалентно заменить на $m$ гладких. Таким образом, исходная задача является эквивалентной следующей выпуклой условной гладкой задаче:

$$\min_{x,t} t \qquad \text{s.t.} \quad t \geq f_i(x), \quad 1 \leq i \leq m.$$

(Почему задача выпуклая?)

Иногда также полезно следующее обобщение переформулировки через надграфик:

$$\min_{x \in E} g(x) + \sum_{i=1}^{m} f_i(x) \qquad \Longleftrightarrow \qquad \min_{x,t} g(x) + \sum_{i=1}^{m} t_i \qquad \text{s.t. } x \in E, \quad t_i \geq f_i(x), \quad 1 \leq i \leq m,$$

где $g, f_1, \ldots, f_m : E \to \mathbb{R}$ — функции. (Здесь допускается случай $g = 0$.)

**Пример 1.4.** Снова рассмотрим задачу минимизации гладкой функции с $l^1$-регуляризатором:

$$\min_{x \in \mathbb{R}^n} f(x) + \|x\|_1,$$

где $f : \mathbb{R}^n \to \mathbb{R}$ — гладкая функция. Поскольку $\|x\|_1 = \sum_{i=1}^n |x_i|$, то, использую обобщенную переформулировку через надграфик, задачу можно эквивалентно переписать в виде

$$\min_{(x,t) \in \mathbb{R}^n \times \mathbb{R}^m} f(x) + \sum_{i=1}^{m} t_i \qquad \text{s.t. } t_i \geq |x_i|, \quad 1 \leq i \leq m.$$

Используя $t_i \geq |x_i| \Longleftrightarrow t_i \geq x_i,\ t_i \geq -x_i$, получаем эквивалентную гладкую задачу:

$$\min_{(x,t) \in \mathbb{R}^n \times \mathbb{R}^m} f(x) + \sum_{i=1}^{m} t_i \qquad \text{s.t. } t_i \geq x_i,\ t_i \geq -x_i \quad 1 \leq i \leq m.$$

# 2 Стандартные классы выпуклых задач

## 2.1 Линейное программирование (LP)

**Стандартная форма LP:**

$$\min_x \langle c, x \rangle \qquad \text{s.t.} \quad Ax \preceq b, \quad Gx = h.$$

где $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $G \in \mathbb{R}^{p \times n}$, $h \in \mathbb{R}^p$.

**Пример 2.1** (Максимум из аффинных функций)**.** Пусть $a_1, \ldots, a_m \in \mathbb{R}^n$ и $b_1, \ldots, b_m \in \mathbb{R}$. Рассмотрим задачу

$$\min_{x \in \mathbb{R}^n} \max_{1 \leq i \leq m} \{\langle a_i, x \rangle + b_i\}.$$

Эта задача является задачей *негладкой* безусловной минимизации (из-за присутствия максимума). Тем не менее, эта задача эквивалентна следующей задаче линейного программирования:

$$\min_{x,t} t \qquad \text{s.t.} \quad t \geq \langle a_i, x \rangle + b_i, \quad 1 \leq i \leq m.$$

В стандартной форме это соответствует тому, что $\tilde{n} = n + 1$, $\tilde{m} = m$, $\tilde{x} = (x, t) \in \mathbb{R}^{n+1}$, $\tilde{c} := (0_n, 1) \in \mathbb{R}^{n+1}$, $\tilde{b} := b$ и $\tilde{A} := (A, 1_n) \in \mathbb{R}^{m \times (n+1)}$, где $A \in \mathbb{R}^{m \times n}$ — матрица со строками $a_1, \ldots, a_m$.

## 2.2 Квадратичное программирование (QP)

| **Стандартная форма QP:** |
| --- |
| $$\min_x \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle \qquad \text{s.t.} \quad Cx \preceq d, \quad Gx = h.$$ где $x, b \in \mathbb{R}^n$, $A \in \mathbb{S}^n_+$, $C \in \mathbb{R}^{m\times n}$, $d \in \mathbb{R}^m$, $G \in \mathbb{R}^{p\times n}$, $h \in \mathbb{R}^p$. |

**Пример:** линейная регрессия $\|Ax - b\|_2^2$ на положительном ортанте: $x \succeq 0$.
**Пример:** расстояние между двумя выпуклыми многогранниками: $\min \|x - y\|_2^2$, $Ax \preceq b$, $Cy \preceq d$.

## 2.3 Квадратичное программирование с квадратичными ограничениями (QCQP)

| **Стандартная форма QCQP:** |
| --- |
| $$\min_x \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle \qquad \text{s.t.} \quad \frac{1}{2}\langle P_i x, x\rangle + \langle q_i, x\rangle + r_i \le 0, \quad 1 \le i \le m, \quad Gx = h.$$ где $x, b, q_1, \ldots, q_m \in \mathbb{R}^n$, $A, P_1, \ldots, P_m \in \mathbb{S}^n_+$, $r_1, \ldots, r_m \in \mathbb{R}$, $G \in \mathbb{R}^{p\times n}$, $h \in \mathbb{R}^p$. |

**Пример 2.2** (Квадратичная функция на единичном шаре)**.** Рассмотрим задачу

$$\min_x \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle \qquad \text{s.t.} \quad \|x\|_2 \le 1,$$

где $x, b \in \mathbb{R}^n$ и $A \in \mathbb{S}^n_+$. Эта задача эквивалентна следующей задаче QCQP:

$$\min_x \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle \qquad \text{s.t.} \quad \langle x, x\rangle \le 1.$$

В этом случае $m = 1$, $P_1 = I_n$, $q_1 = 0$ и $c_1 = -1$.

## 2.4 Коническое программирование второго порядка (SOCP)

| **Стандартная форма SOCP:** |
| --- |
| $$\min_x \langle q, x\rangle \qquad \text{s.t.} \quad \|A_i x + b_i\|_2 \le \langle c_i, x\rangle + d_i, \quad 1 \le i \le m, \quad Gx = h.$$ где $x, q, c_1, \ldots, c_m \in \mathbb{R}^n$, $A_1 \in \mathbb{R}^{s_1\times n}$, $b_1 \in \mathbb{R}^{s_1}$, $\ldots$, $A_m \in \mathbb{R}^{s_m\times n}$, $b_m \in \mathbb{R}^{s_m}$, $d_1, \ldots, d_m \in \mathbb{R}$, $G \in \mathbb{R}^{p\times n}$, $h \in \mathbb{R}^p$. |

**Пример 2.3.** Пусть $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Рассмотрим задачу

$$\min_x \|Ax - b\|_2 + \|x\|_2. \tag{2.1}$$

Эта задача является негладкой из-за присутствия слагаемого $\|x\|_2$. Эта задача является эквивалентной следующей задаче SOCP:

$$\min_{x, t_1, t_2} t_1 + t_2 \qquad \text{s.t.} \quad \|Ax - b\|_2 \le t_1, \qquad \|x\|_2 \le t_2.$$

(Почему эту задачу (2.1) нельзя было похожим образом переформулировать как QCQP?)

## 2.5 Полуопределённое программирование (SDP)

| **Стандартная форма SDP:** |
|---|
| $$\min_x \langle c, x\rangle \qquad \text{s.t.} \quad A_0 + x_1 A_1 + \cdots + x_n A_n \succeq 0, \quad Cx = d, Gx = h.$$ где $x, c \in \mathbb{R}^n$, $A_0, \ldots, A_n \in \mathbb{S}^d$, $C \in \mathbb{R}^{m\times n}$, $d \in \mathbb{R}^m$, $G \in \mathbb{R}^{p\times n}$, $h \in \mathbb{R}^p$ |

При преобразовании задач оптимизации к SDP чрезвычайно полезно следующее утверждение из линейной алгебры, представляющее собой самостоятельный интерес:

**Утверждение 2.4** (Лемма о дополнении Шура)**.** *Пусть $A \in \mathbb{S}^n$, $C \in \mathbb{S}^m$ — симметричные матрицы, $B \in \mathbb{R}^{m\times n}$, и пусть*

$$M := \begin{pmatrix} A & B^T \\ B & C \end{pmatrix}.$$

*Пусть $A \succ 0$. Тогда $M \succeq 0$, если и только если $BA^{-1}B^T \preceq C$.*

**Замечание 2.5.** Матрица $C - BA^{-1}B^T$ называется *дополнением Шура для матрицы $A$*. Таким образом, лемма о дополнении Шура утверждает, что при условии $A \succ 0$ блочная матрица $M$ неотрицательно определена, если и только если неотрицательно определено дополнение Шура для матрицы $A$. Лемму о дополнении Шура также можно сформулировать и относительно матрицы $C$: если $C \succ 0$, то $M \succeq 0$ тогда и только тогда, когда $B^T C^{-1} B \preceq A$; матрица $A - B^T C^{-1} B$ называется *дополнением Шура для матрицы $C$*.

*Доказательство.* По определению, $M \succeq 0$, если и только если

$$\left\langle M \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle = \langle Ax, x\rangle + 2\langle B^T y, x\rangle + \langle Cy, y\rangle \geq 0$$

для всех $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, что эквивалентно

$$\inf_{x\in\mathbb{R}^n} \{\langle Ax, x\rangle + 2\langle B^T y, x\rangle + \langle Cy, y\rangle\} \geq 0 \tag{2.2}$$

для всех $y \in \mathbb{R}^m$. Заметим, что для $y \in \mathbb{R}^m$ функция $x \mapsto \langle Ax, x\rangle + 2\langle B^T y, x\rangle$ является строго выпуклой квадратичной функцией с минимумом в точке $\bar{x} := -A^{-1}B^T y$. Таким образом, (2.2) принимает вид

$$\langle (C - BA^{-1}B^T)y, y\rangle \geq 0$$

для всех $y \in \mathbb{R}^m$. Но это, по определению, означает $BA^{-1}B^T \preceq C$. □

**Пример 2.6** (Минимизация операторной нормы)**.** Рассмотрим задачу

$$\min_{x\in\mathbb{R}^n} \|B(x)\|_{\mathrm{op}},$$

где $B : \mathbb{R}^n \to \mathbb{R}$ — преобразование $B(x) := B_0 + x_1 B_1 + \cdots + x_n B_n$ для $B_0, \ldots, B_n \in \mathbb{R}^{p\times n}$. Представим эту задачу в виде SDP. Для этого сначала перейдем эквивалентной формулировке через надграфик:

$$\min_{x,t} t \qquad \text{s.t.} \quad t \geq \|B(x)\|_{\mathrm{op}}.$$

Заметим, что

$$\|B(x)\|_{\mathrm{op}} \leq t \qquad \Longleftrightarrow \qquad \lambda_{\max}(B(x)B(x)^T) \leq t^2, \quad t \geq 0 \qquad \Longleftrightarrow \qquad B(x)B(x)^T \preceq t^2 I_p, \quad t \geq 0.$$

По лемме о дополнении Шура, получаем

$$B(x)B(x)^T \preceq t^2 I_p, \quad t \geq 0 \qquad \Longleftrightarrow \qquad \begin{pmatrix} tI_q & B(x)^T \\ B(x) & tI_p \end{pmatrix} \succeq 0.$$

Таким образом, исходная задача эквивалентна следующей SDP:

$$\min_{x,t} t \qquad \text{s.t.} \quad \begin{pmatrix} tI_q & B(x)^T \\ B(x) & tI_p \end{pmatrix} \succeq 0.$$

**Замечание 2.7.** На самом деле, ответ пока еще не записан в стандартной форме SDP. Чтобы это сделать, необходимо переписать матрицу

$$\begin{pmatrix} tI_q & B(x)^T \\ B(x) & tI_p \end{pmatrix}$$

в виде аффинного преобразования от переменных $x$ и $t$. Нетрудно видеть, что это возможно сделать, поскольку каждый блок указанной $2 \times 2$ блочной матрицы представляет собой аффинное преобразование от переменных $x$ и $t$. В данном случае явная форма записи такая:

$$\begin{pmatrix} tI_q & B(x)^T \\ B(x) & tI_p \end{pmatrix} = \begin{pmatrix} 0 & B_0 \\ B_0 & 0 \end{pmatrix} + \sum_{i=1}^n x_i \begin{pmatrix} 0 & B_i^T \\ B_i & 0 \end{pmatrix} + t \begin{pmatrix} I_q & 0 \\ 0 & I_p \end{pmatrix}.$$

**Замечание 2.8.** Прокомментируем немного аккуратнее эквивалентность

$$BB^T \preceq t^2 I_p, \quad t \geq 0 \qquad \Longleftrightarrow \qquad \begin{pmatrix} tI_q & B^T \\ B & tI_p \end{pmatrix} \succeq 0$$

для $B \in \mathbb{R}^{p \times n}$, $t \in \mathbb{R}$, и почему здесь не возникает никаких проблем при $t = 0$. Для этого воспользуемся стандартным регуляризационным аргументом: если $t \geq 0$, то

$$BB^T \preceq t^2 I_p \qquad \Longleftrightarrow \qquad BB^T \preceq (t+\varepsilon)^2 I_p \quad \text{для всех } \varepsilon > 0.$$

(Переход слева направо следует из положительности $\varepsilon$, а справа налево из предельного перехода.) Поскольку $t + \varepsilon > 0$, то по лемме о дополнении Шура

$$BB^T \preceq (t+\varepsilon)^2 I_p \qquad \Longleftrightarrow \qquad B((t+\varepsilon)^{-1} I_q) B^T \preceq (t+\varepsilon) I_p \qquad \Longleftrightarrow \qquad \begin{pmatrix} (t+\varepsilon)I_q & B^T \\ B & (t+\varepsilon)I_p \end{pmatrix} \succeq 0.$$

В итоге, показано, что, если $t \geq 0$, то

$$BB^T \preceq t^2 I_p \qquad \Longleftrightarrow \qquad \begin{pmatrix} (t+\varepsilon)I_q & B^T \\ B & (t+\varepsilon)I_p \end{pmatrix} = \begin{pmatrix} tI_q & B^T \\ B & tI_p \end{pmatrix} + \varepsilon I_{p+q} \succeq 0 \quad \text{для всех } \varepsilon > 0.$$

Но

$$\begin{pmatrix} tI_q & B^T \\ B & tI_p \end{pmatrix} + \varepsilon I_{p+q} \succeq 0 \quad \text{для всех } \varepsilon > 0 \qquad \Longleftrightarrow \qquad \begin{pmatrix} tI_q & B^T \\ B & tI_p \end{pmatrix} \succeq 0.$$

Таким образом, показано, что

$$BB^T \preceq t^2 I_p, \quad t \geq 0 \qquad \Longleftrightarrow \qquad \begin{pmatrix} tI_q & B^T \\ B & tI_p \end{pmatrix} \succeq 0, \quad t \geq 0.$$

Остается заметить, что условие $t \geq 0$ справа излишне, так как из положительной полуопределенности блочной матрицы следует $tI_q \succeq 0$ (все главные подматрицы также должны быть положительно полуопределены); но последнее эквивалентно $t \geq 0$.

## 2.6 Вложенность классов друг в друга

Можно показать, что LP $\subset$ QP $\subset$ QCQP $\subset$ SOCP $\subset$ SDP. Таким образом, SDP являются наиболее общей постановкой выпуклых задач.

Вложения LP $\subset$ QP $\subset$ QCQP очевидные.

Покажем вложение QCQP $\subset$ SOCP. Опуская целевую функцию в ограничения (замечание 1.2), получаем задачу с линейной целевой функцией и квадратичным ограничением. Таким образом, достаточно показать, что любое квадратичное ограничение

$$\frac{1}{2}\langle Px, x\rangle + \langle q, x\rangle + r \le 0,$$

где $P \in \mathbb{S}^n_+$, $q \in \mathbb{R}^n$, $r \in \mathbb{R}$, может быть эквивалентным образом переписано в виде конического ограничения

$$\|Ax + b\| \le \langle c, x\rangle + d.$$

Поскольку $P \in \mathbb{S}^n_+$, то $P = 2L^T L$ для некоторого $L \in \mathbb{R}^{n\times n}$. Отсюда

$$\frac{1}{2}\langle Px, x\rangle + \langle q, x\rangle + r = \|Lx\|^2 + \langle q, x\rangle + r.$$

Используя тождество $t = \left(\frac{t+1}{2}\right)^2 - \left(\frac{t-1}{2}\right)^2$, получаем

$$\langle q, x\rangle + r = \left(\frac{\langle q, x\rangle + r + 1}{2}\right)^2 - \left(\frac{\langle q, x\rangle + r - 1}{2}\right)^2.$$

Таким образом,

$$\frac{1}{2}\langle Px, x\rangle + \langle q, x\rangle + r \le 0 \qquad \Longleftrightarrow \qquad \|Lx\|^2 + \left(\frac{\langle q, x\rangle + r + 1}{2}\right)^2 \le \left(\frac{\langle q, x\rangle + r - 1}{2}\right)^2.$$

Извлекая квадратный корень и используя, что $\frac{1}{2}\langle Px, x\rangle + \langle q, x\rangle + r \le 0$ возможно только в случае $\langle q, x\rangle + r \le 0$, в итоге получаем

$$\frac{1}{2}\langle Px, x\rangle + \langle q, x\rangle + r \le 0 \qquad \Longleftrightarrow \qquad \left(\|Lx\|^2 + \left(\frac{\langle q, x\rangle + r + 1}{2}\right)^2\right)^{1/2} \le \frac{1 - \langle q, x\rangle - r}{2}, \quad \langle q, x\rangle + r \ge 0.$$

Заметим, что второе условие, на самом деле, избыточное, поскольку справедлива импликация

$$\left(\|Lx\|^2 + \left(\frac{\langle q, x\rangle + r + 1}{2}\right)^2\right)^{1/2} \le \frac{1 - \langle q, x\rangle - r}{2} \qquad \Longrightarrow \qquad \langle q, x\rangle + r \ge 0.$$

(Действительно, это следует из того, что левая часть $\ge \frac{|\langle q, x\rangle + r + 1|}{2} \ge \frac{\langle q, x\rangle + r + 1}{2}$.) Таким образом,

$$\frac{1}{2}\langle Px, x\rangle + \langle q, x\rangle + r \le 0 \qquad \Longleftrightarrow \qquad \left(\|Lx\|^2 + \left(\frac{\langle q, x\rangle + r + 1}{2}\right)^2\right)^{1/2} \le \frac{1 - \langle q, x\rangle - r}{2}.$$

Поскольку преобразования $x \mapsto (Lx, \frac{\langle q, x\rangle + r + 1}{2})$ и $x \mapsto \frac{1 - \langle q, x\rangle - r}{2}$ аффинные, то найдутся $A$, $b$, $c$ и $d$, такие, что

$$Ax + b = \left(Lx, \frac{\langle q, x\rangle + r + 1}{2}\right), \qquad \langle c, x\rangle + d = \frac{1 - \langle q, x\rangle - r}{2}$$

для всех $x \in \mathbb{R}^n$. Таким образом, каждое квадратичное ограничение можно представить в виде конического ограничения.

Вложение SOCP $\subset$ SDP следует из леммы о дополнении Шура (см. пример 2.6 и замечания 2.7 и 2.8 для более подробного описания):

$$\|Ax+b\| \le \langle c, x\rangle + d \qquad \Longleftrightarrow \qquad \begin{cases} \|Ax + b\|^2 \le (\langle c, x\rangle + d)^2, \\ \langle c, x\rangle + d \ge 0 \end{cases} \qquad \Longleftrightarrow \qquad \begin{pmatrix} \langle c, x\rangle + d & Ax + b \\ (Ax + b)^T & \langle c, x\rangle + d \end{pmatrix} \succeq 0.$$