Методы оптимизации

Домашнее задание 2

Арутюнян Саро, Б05-924

Выпуклые функции.

Задача 1. Докажите, что функция $f(X) = \operatorname{tr}(X^{-1})$, $X \in S_n^{++}$ выпуклая, а функция $g(X) = (\det(X))^{\frac{1}{n}}$ вогнутая.

Решение. Имеем

$$df = \operatorname{tr}\big(d(X^{-1})\big) = \operatorname{tr}(-X^{-1}(dX)X^{-1})$$

$$d^2 f = \operatorname{tr}\big(d(-X^{-1}(dX)X^{-1})\big) = \operatorname{tr}(X^{-1}(dX_1)X^{-1}(dX_2)X^{-1}) + \operatorname{tr}(X^{-1}(dX_2)X^{-1}(dX_1)X^{-1}).$$

Покажем, что второй дифференциал f неотрицательно определен, т.е. $\forall H \in S_n$

$$D^2 f(X)[H,H] = 2\operatorname{tr}(X^{-1}HX^{-1}HX^{-1}).$$

Но это верно, поскольку

$$X^{-1}HX^{-1}HX^{-1} = \left(X^{-1}HX^{-\frac{1}{2}}\right)\left(X^{-\frac{1}{2}}HX^{-1}\right) = \left(X^{-\frac{1}{2}}HX^{-1}\right)^T\left(X^{-\frac{1}{2}}HX^{-1}\right) = \left(X^{-\frac{1}{2}}HX^{-1}\right)^2.$$

Теперь вторая часть задачи:

$$dg = \frac{1}{n}(\det(X))^{\frac{1}{n}-1}\cdot\det(X)\cdot\langle X^{-T}, dX\rangle = \frac{1}{n}(\det X)^{\frac{1}{n}}\cdot\langle X^{-T}, dX\rangle$$

$$d^2 g = \frac{1}{n}\cdot\frac{1}{n}(\det X)^{\frac{1}{n}}\cdot\langle X^{-T}, dX_2\rangle\cdot\langle X^{-T}, dX_1\rangle + \frac{1}{n}(\det X)^{\frac{1}{n}}\cdot\langle -X^{-T}(dX_2)^T X^{-T}, dX_1\rangle =$$

$$= \frac{1}{n^2}(\det X)^{\frac{1}{n}}\big[\operatorname{tr}(X^{-1}dX_1)\operatorname{tr}(X^{-1}dX_2) - n\cdot\operatorname{tr}\big(X^{-1}(dX_2)X^{-1}(dX_1)\big)\big].$$

Пусть $\lambda_1, \dots, \lambda_n$ есть собственные числа матрицы $X^{-1}H$. Тогда

$$D^2 g(X)[H,H] = \frac{1}{n^2}(\det X)^{\frac{1}{n}}[\operatorname{tr}(X^{-1}H)^2 - n\cdot\operatorname{tr}((X^{-1}H)^2)] =$$

$$= \frac{1}{n^2}(\det X)^{\frac{1}{n}}\left[\left(\sum_{i=1}^n \lambda_i\right)^2 - n\left(\sum_{i=1}^n \lambda^2\right)\right] \le 0.$$

Получили, что второй дифференциал g неположительно определен, что и требовалось доказать.

Задача 2. Дивергенция Кульбака-Лейблера между $p, q \in \mathbb{R}_n^{++}$ определяется следующим образом:

$$D(p,q) = \sum_{i=1}^n \left(p_i \ln\frac{p_i}{q_i} - p_i + q_i\right).$$

Докажите, что $D(p,q) \geq 0\ \forall p,q \in \mathbb{R}_n^{++}$ и $D(p,q) = 0$ тогда и только тогда, когда $p = q$.

Решение. Рассмотрим функцию $g\colon (0,+\infty) \to \mathbb{R}$

$$g(x) = \ln x - 1 + \frac{1}{x}.$$

Имеем $g'(x) = \frac{1}{x} - \frac{1}{x^2} = \frac{x-1}{x^2}$, значит функция убывает в $(0,1)$ и возрастает в $(1,+\infty)$. Отсюда минимум функции достигается в 1 и $g(x) \geq g(0) = 0$, причем $g(x) = 0 \Leftrightarrow x = 1$. Это означает, что

$$D(p,q) = \sum_{i=1}^{n}\left(p_i \ln\frac{p_i}{q_i} - p_i + q_i\right) = \sum_{i=1}^{n} p_i\left(\ln\frac{p_i}{q_i} - 1 + \frac{q_i}{p_i}\right) = \sum_{i=1}^{n} p_i g\left(\frac{p_i}{q_i}\right) \geq 0,$$

причем равенство достигается в том и только в том случае, когда $p_i = q_i\ \forall i \in \{1,\dots,n\}$, что и требовалось доказать.

Задача 3. Пусть $x \in \mathbb{R}^n$ случайная величина с заданным распределением вероятностей $\mathbb{P}(x_i = a_i) = p_i$, где $i = 1,\dots,n$ и $a_1 < \cdots < a_n$. Говорят, что вектор вероятностей исходов $p \in \mathbb{R}^n$ принадлежит вероятностному симплексу, т.е.

$$P = \{p\colon 1^T p = 1, p \succcurlyeq 0\}.$$

Определите выпуклость и вогнутость следующих функций на вероятностном симплексе.

(a) $\mathbb{E}x$
(b) $\mathbb{P}(x \geq \alpha)$
(c) $\mathbb{P}(\alpha \leq x \leq \beta)$
(d) $\sum_{i=1}^{n} p_i \ln p_i$
(e) $\mathbb{V}x \geq \alpha$
(f) $\mathrm{quartile}(x) = \inf\{\beta\colon \mathbb{P}(x \leq \beta) \geq 0.25\}$

Решение. Заметим, что $\forall x,y \in P$, $\forall \theta \in [0,1]$ $1^T(\theta x + (1-\theta)y) = \theta + (1-\theta) = 1$ и $\theta x + (1-\theta)y \succcurlyeq 0$, следовательно $\theta x + (1-\theta)y \in P$, т.е. P выпукло.

(a) Поскольку $\forall x,y \in P$, $\forall \theta \in [0,1]$ $\theta x + (1-\theta)y \in P$ и $\mathbb{E}(\theta x + (1-\theta)y) = \theta\mathbb{E}x + (1-\theta)\mathbb{E}y$, то функция $\mathbb{E}x$ выпукла на P.

(b) Покажем, что $\mathbb{P}(x \geq \alpha)$ выпуклая функция на P. Пусть $x,y \in P$, $\theta \in [0,1]$. Тогда

$$\theta\mathbb{P}(x \geq \alpha) + (1-\theta)\mathbb{P}(y \geq \alpha) = \mathbb{P}(x \geq \alpha) = \prod_{i=1}^{n}\mathbb{P}(x_i \geq \alpha_i) = \prod_{i=1}^{n}\left(p_n + \cdots + p_{\sigma(i)}\right),$$

где $\sigma(i) = \min\limits_{j \in \{1,\dots,n\}}\{j\colon a_j \geq \alpha_i\}$ (если $\nexists j\colon a_j \geq \alpha_i$, то тогда $\sigma(i) = 0$ и $p_0 = 0$). С другой стороны

$$\mathbb{P}(\theta x + (1-\theta)y \geq \alpha) = \prod_{i=1}^{n}\mathbb{P}(\theta x_i + (1-\theta)y_i \geq \alpha_i) \geq \prod_{i=1}^{n}\mathbb{P}\left(\theta x_i + (1-\theta)y_i \geq a_{\sigma(i)}\right).$$

Пусть $j = \sigma(i)$ и $A = \left\{\theta x_i + (1-\theta)y_i \geq a_{\sigma(i)}\right\}$. Заметим, что

$$\mathbb{P}(A) = \sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{P}(A|x_i = a_k)\mathbb{P}(A|y_i = a_l) \geq \sum_{k=j}^{n}\sum_{l=j}^{n}\mathbb{P}(A|x_i = a_k)\mathbb{P}(A|y_i = a_l) =$$

$= \sum_{k=j}^{n}\sum_{l=j}^{n} p_k p_l = \left(p_j + \cdots + p_n\right)^2 \dots$

(c) …

(d) Для функции $f(p) = \sum_{i=1}^{n} p_i \ln p_i$ определенной на $(0, +\infty)^n$ имеем $\frac{\partial^2}{\partial p_i \partial p_j} f(p) = \begin{cases} \frac{1}{p_i}, i = j \\ 0, i \neq j \end{cases}$. Значит $\nabla_{pp}^2 f = \operatorname{diag}\{1/p_i\} \succcurlyeq 0$ и f выпукла.

Задача 4. Являются ли следующие функции выпуклыми на $\mathbb{R}^n$:

- $f(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$
- $g(x) = \prod_{i=1}^{n} (x_i)^{\frac{1}{n}}$.

Решение. Поскольку второй дифференциал f тождественная ноль, то f является как выпуклой, так и вогнутой на $\mathbb{R}^n$. Заметим, что g определена только на том октанте $\mathbb{R}^n$, где все координаты положительны. Посчитаем второй дифференциал g:

$$\frac{\partial}{\partial x_i} g(x) = \frac{1}{n} x_i^{-1} g(x), \qquad \frac{\partial^2}{\partial x_i^2} g(x) = \frac{1}{n}\left(1 - \frac{1}{n}\right) x_i^{-2} g(x),$$

$$\frac{\partial^2}{\partial x_i \partial x_j} g(x) = |i \neq j| = \frac{1}{n^2} x_i^{-1} x_j^{-1} g(x).$$

$$\nabla^2 g(x) = \frac{g(x)}{n^2}\{x_i^{-1} x_j^{-1}\} - \frac{g(x)}{n} \operatorname{Diag}\{x_i^{-2}\} = \frac{g(x)}{n^2}\left(x^{-1}(x^{-1})^T - n \operatorname{Diag}\{x_i^{-2}\}\right)$$

Для любого $y \in \mathbb{R}^n$ имеем

$$y^T\left(\nabla^2 g(x)\right) y = \frac{g(x)}{n^2}\left[y^T x^{-1} (y^T x^{-1})^T - n y^T \operatorname{Diag}\{x_i^{-2}\} y\right] =$$

$$= \frac{g(x)}{n^2}\left[\left(\sum_{i=1}^{n} y_i x_i^{-1}\right)^2 - n \sum_{i=1}^{n} y_i^2 x_i^{-2}\right] \leq 0.$$

Значит g вогнута на своей области определения.

Задача 5. Докажите, что спектральная матричная норма выпуклая функция на множестве матриц

$$f(X) = \|X\|_2 = \sup_{y \neq 0} \frac{\|Xy\|_2}{\|y\|_2} = \sup_{\|y\|=1} \|Xy\|.$$

Решение. $\forall \theta \in [0, 1]$

$$\theta f(X_1) + (1-\theta) f(X_2) = \sup_{\|y\|=1} \|\theta X_1 y\| + \sup_{\|y\|=1} \|(1-\theta) X_2 y\| \geq$$

$$\geq \sup_{\|y\|=1} (\|\theta X_1 y\| + \|(1-\theta) X_2 y\|) \geq \sup_{\|y\|=1} (\|(\theta X_1 + (1-\theta) X_2) y\|) = f(\theta X_1 + (1-\theta) X_2).$$

Задача 6. Является ли функция $f(x) = -x \ln x - (1-x)\ln(1-x)$ выпуклой на $\mathbb{R}$?

Решение. Сначала заметим, что f определена на $(0, 1)$.

$$-f'(x) = \ln x + 1 - \ln(1-x) - 1 = \ln x - \ln(1-x)$$

$$-f''(x) = \frac{1}{x} + \frac{1}{1-x} > 0.$$

Из последнего неравенства следует вогнутость функции $f(x)$.

Условия оптимальности. Двойственность.

Задача 1. Найти решение задачи методом множителей Лагранжа

$$\min_{x\in\mathbb{R}^2}\{x_1^2 + x_2^2\}$$

$$x_1 \leq 0,$$

$$x_1 + x_2 = 0$$

Решение. Функцией Лагранжа будет

$$L(x_1, x_2, \lambda, \mu) = x_1^2 + x_2^2 + \lambda(x_1 + x_2) + \mu x_1$$

Нужно решить систему

$$\begin{cases} L_{x_1} = 2x_1 + \lambda + \mu = 0 \\ L_{x_2} = 2x_2 + \lambda = 0 \\ x_1 + x_2 = 0 \\ x_1 \leq 0 \\ \mu x_1 = 0 \\ \mu \geq 0 \\ \nabla_{xx} L = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succcurlyeq 0 \end{cases}.$$

Имеем $x_1 - x_2 = -\frac{\mu}{2}$, $x_1 + x_2 = 0$, следовательно $x_1 = -\frac{\mu}{4}$ и из $0 = \mu x_1 = -\frac{\mu^2}{4}$ следует, что $\mu = 0$, $x_1 = 0$, $x_2 = 0$, $\lambda = 0$. Минимум функции Лагранжа достигается в точке $(0,0,0,0)$: $L = 0$.

Задача 2. Найти решение задачи методом множителей Лагранжа

$$\min_{x\in\mathbb{R}^2}\{(x_1 - 4)^2 + (x_2 - 4)^2\}$$

$$x_1 + x_2 \leq 4,$$

$$x_1 + 3x_2 \leq 9.$$

Решение. Функция Лагранжа

$$L(x_1, x_2, \mu_1, \mu_2) = (x_1 - 4)^2 + (x_2 - 4)^2 + \mu_1(x_1 + x_2 - 4) + \mu_2(x_1 + 3x_2 - 9).$$

Составим и решим систему

$$
\begin{cases}
L_{x_1} = 2x_1 - 8 + \mu_1 + \mu_2 = 0 \\
L_{x_2} = 2x_2 - 8 + \mu_1 + 3\mu_2 = 0 \\
\mu_1 \geq 0, \qquad \mu_2 \geq 0 \\
\mu_1(x_1 + x_2 - 4) = 0 \\
\mu_2(x_1 + 3x_2 - 9) = 0 \\
x_1 + x_2 - 4 \leq 0 \\
x_1 + 3x_2 - 9 \leq 0 \\
\nabla_{xx} L = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succcurlyeq 0
\end{cases}.
$$

Имеем $x_1 = 4 - \frac{\mu_1 + \mu_2}{2}$, $x_2 = 4 - \frac{\mu_1 + 3\mu_2}{2}$, откуда

$$
\begin{cases}
\mu_1(4 - \mu_1 - 2\mu_2) = 0 \\
\mu_2(7 - 2\mu_1 - 5\mu_2) = 0
\end{cases}.
$$

Если $\mu_1 = 0$, то $\mu_2 = \frac{7}{5}$, $x_1 = \frac{33}{10}$, $x_2 = \frac{19}{10}$, но тогда $x_1 + x_2 - 4 > 0$ — противоречие. Если $\mu_2 = 0$, то $\mu_1 = 4$, $x_1 = 2$, $x_2 = -2$. Тогда $L = 4 + 36 = 40$. Если же $\mu_2 \neq 0$, то

$$
\begin{cases}
4 - \mu_1 - 2\mu_2 = 0 \\
7 - 2\mu_1 - 5\mu_2 = 0
\end{cases}
\Leftrightarrow
\begin{cases}
\mu_1 = 6 \\
\mu_2 = -1
\end{cases},
\qquad
\begin{cases}
x_1 = 1.5 \\
x_2 = 2.5
\end{cases}.
$$

Тогда $L = 2.5^2 + 1.5^2 = 8.5$. Итак, искомый минимум равен 8.5 и достигается при $x = (1.5 \quad 2.5)^T$.

Задача 3. Найти решение задачи методом множителей Лагранжа

$$
\min_{x \in \mathbb{R}^n} \left\{ (1 - x_1)^2 + \alpha \sum_{i=2}^{n} (x_i - x_{i-1})^2 \right\}
$$

$$
\alpha \geq 0.
$$

Решение. Функция Лагранжа

$$
L(x_1, \dots, x_n) = (1 - x_1)^2 + \alpha[(x_1 - x_2)^2 + (x_2 - x_3)^2 + \dots + (x_{n-1} - x_n)^2].
$$

$$
\begin{cases}
L_{x_1} = 2(x_1 - 1) + 2\alpha(x_1 - x_2) = 0 \\
L_{x_2} = 2\alpha(-x_1 + x_2 + x_2 - x_3) = 0 \\
\dots \\
L_{x_{n-1}} = 2\alpha(-x_{n-2} + x_{n-1} + x_{n-1} + x_n) = 0 \\
L_{x_n} = 2\alpha(-x_{n-1} + x_n) = 0
\end{cases}
\Rightarrow x_n = x_{n-1} = \dots x_2 = x_1 = 1.
$$

Итак, искомый минимум достигается в $x = (1 \quad \dots \quad 1)$ и равен 0.

Задача 4. Записать двойственную задачу к задаче

$$
\min_{x \in \mathbb{R}^n} x^T W x
$$

$$
x_i^2 = 1, \qquad i = 1, \dots, n
$$

где $W \in S^n$.

Решение. Функция Лагранжа

$$L(x,\lambda) = x^T W x + \sum_{i=1}^{n} \lambda_i (x_i^2 - 1)$$

$$0 = 2\langle Wx, dx\rangle + \sum_{i=1}^{n} 2\lambda_i x_i dx_i = 2\left\langle Wx + \begin{pmatrix} \lambda_1 x_1 \\ \vdots \\ \lambda_n x_n \end{pmatrix}, dx \right\rangle$$

Если $W \neq \begin{pmatrix} -\lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -\lambda_n \end{pmatrix}$, то L не имеет минимума по x. Значит двойственная задача:

$$\max_{\lambda \in \mathbb{R}^n} x^T W x + \sum_{i=1}^{n} \lambda_i (x_i^2 - 1)$$

$$W = \begin{pmatrix} -\lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -\lambda_n \end{pmatrix}.$$

Задача 5. Записать двойственную задачу к задаче

$$\min_{x \in \mathbb{R}^n} f_0(x) = \sum_{i=1}^{n} f_i(x_i)$$

$$a^T x = b$$

где $a \in \mathbb{R}^n$, $b \in \mathbb{R}$, $f_i \colon \mathbb{R} \to \mathbb{R}$ дифференцируемая и строго выпуклая.

Решение. Пусть $u_i = \frac{\partial}{\partial x_i} f_i$. Тогда u_i строго монотонная и существует обратная функция u^{-1}. Функция Лагранжа

$$L(x,\lambda) = \sum_{i=1}^{n} f_i(x_i) + \lambda(a^T x - b)$$

$$0 = \frac{\partial}{\partial x_i} L(x,\lambda) = u_i(x_i) + \lambda a_i \Longrightarrow x_i^* = u_i^{-1}(-\lambda a_i)$$

Двойственная задача:

$$\max_{\lambda \in \mathbb{R}} \sum_{i=1}^{n} f_i\left(u_i^{-1}(-\lambda a_i)\right) + \lambda(a^T u^{-1}(-\lambda a) - b)$$

(здесь $u^{-1}(-\lambda a) = (u_1^{-1}(-\lambda a_1) \quad \dots \quad u_n^{-1}(-\lambda a_n))$).

Задача 6. Записать двойственную задачу к задаче

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T P_0 x + q_0^T x + r_0$$

$$\frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0, \qquad i = 1, \dots, m$$

где $P_0 \in S_{++}^n$, $P_i \in S_+^n$, $a_i \in \mathbb{R}^n$, $q_i \in \mathbb{R}^n$, $r_i \in \mathbb{R}$. Выполняется ли условие сильной двойственности?

Решение. Функция Лагранжа

$$L(x_1, \dots, x_n, \mu_1, \dots, \mu_m) = \frac{1}{2}x^T P_0 x + q_0^T x + r_0 + \sum_{i=1}^{m} \mu_i \left(\frac{1}{2}x^T P_i x + q_i^T x + r_i\right) =$$

$$= \sum_{i=0}^{m} \mu_i \left(\frac{1}{2}x^T P_i x + q_i^T x + r_i\right) = \frac{1}{2}x^T \sum_{i=0}^{m} \mu_i P_i\, x + \sum_{i=0}^{m} \mu_i q_i^T\, x + \sum_{i=0}^{m} \mu_i r_i = \frac{1}{2}x^T P x + q^T x + r,$$

где $\mu_0 = 1$. Заметим, что $P \in S_{++}^n$, следовательно существует P^{-1}. Найдем $g(\mu) = \min\limits_{x \in \mathbb{R}^n} L(x, \mu)$:

$$0 = dL = \langle Px, dx \rangle + \langle q, dx \rangle = \langle Px + q, dx \rangle \Longrightarrow x^* = -P^{-1}q$$

$$d^2 L = \langle P dx_2, dx_1 \rangle = \langle P dx_1, dx_2 \rangle \Longrightarrow \nabla_{xx}^2 L = P \succcurlyeq 0$$

$$g(\mu) = L(x^*, \mu) = \frac{1}{2}q^T P^{-1} P P^{-1} q - q^T P^{-1} q + r = -\frac{1}{2}q^T P^{-1} q + r$$

Двойственная задача — $\max\limits_{\mu \in \mathbb{R}^m} g(\mu)$. Условие сильной двойственности не всегда выполняется: например если $\forall i\ \forall x \in \mathbb{R}^n$ выполнено $\frac{1}{2}x^T P_i x + q_i^T x + r_i \leq 0$.

Задача 8. Записать двойственную задачу к задаче

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

$$x^T x \leq 1$$

где $A \in S_{++}^n$, $b \in \mathbb{R}^n$. Выполняется ли условие сильной двойственности?

Решение. Функция Лагранжа

$$L(x, \mu) = x^T A x + 2b^T x + \mu(x^T x - 1).$$

$$0 = d_x L = 2\langle Ax, dx \rangle + \langle b, dx \rangle + 2\mu \langle x, dx \rangle = 2\langle Ax + b + \mu x, dx \rangle$$

$$Ax^* + b + \mu x^* = 0 \Longleftrightarrow x^* = -(A + \mu I_n)^{-1} b$$

$$\nabla_{xx}^2 L = A + \mu I_n \succcurlyeq 0$$

$$g(\mu) = \min_{x \in \mathbb{R}^n} L(x, \mu) = L(x^*, \mu) =$$

$$= b^T (A + \mu I_n)^{-1} A (A + \mu I_n)^{-1} b - 2b^T (A + \mu I_n)^{-1} b + \mu (b^T (A + \mu I_n)^{-2} b - 1)$$

Двойственная задача: $\max\limits_{\mu \in \mathbb{R}} g(\mu)$. Очевидно функция $f(x) = x^T A x + 2b^T x$ выпукла ($\nabla_{xx}^2 f = A \succcurlyeq 0$). Также например для $x = 0$ верно $x^T x < 1$. Значит задача сильно двойственная.