ДЗ-2 по Методам Оптимизации
Попов Николай, 776 группа

# Matrix calculus

1)$f(x) = \| Ax \|_2$ . $\bigtriangledown f(x)-?$
$df(x) = < \bigtriangledown f(x), dx > = d\sqrt{< Ax, Ax >} = \dfrac{d < Ax, Ax >}{2\sqrt{< Ax, Ax >}} = \dfrac{< Ax, Adx >}{\| Ax \|_2} =$
$\dfrac{1}{\| Ax \|_2} < A^T Ax, dx > \Rightarrow \bigtriangledown f(x) = \dfrac{A^T Ax}{\| Ax \|_2}$

2)
$f(X) = \ln detX$
$df(X) = \dfrac{1}{detX} d(detX) = < X^{-T}, dX >$

$0 = d(I) = d(XX^{-1}) = dX \cdot X^{-1} + Xd(X^{-1}) \Rightarrow -X^{-1} dX \cdot X^{-1} = d(X^{-1}) \Rightarrow$
$d(X^{-T}) = -X^{-T} dX \cdot X^{-T}$

$d^2 f = < dX^{-T}, dX_1 > = < -X^{-T} dX_2 X^{-T}, dX_1) > = -tr(x^{-1} dX_2^T X^{-1} dX_1^T)$

$$\| f"(X) \|_{ij} = \frac{\partial}{\partial X_{ij}} \bigtriangledown f(X) = \frac{\partial}{\partial X_{ij}} X^{-T}$$

3)
$\dfrac{\partial}{\partial X} \| X \|_F = \dfrac{\partial}{\partial X} \left( \sum_{i=1..n;\ j=1..m} X_{ij}^2 \right) = \left( \dfrac{\partial}{\partial X_{kl}} \sum X_{ij}^2 \right)_{k=1..n;\ l=1..m} =$
$= (2X_{kl})_{kl} = 2X$

4)$f(x) = \ln < Ax, x >$ . $A \in S_{++}^n$

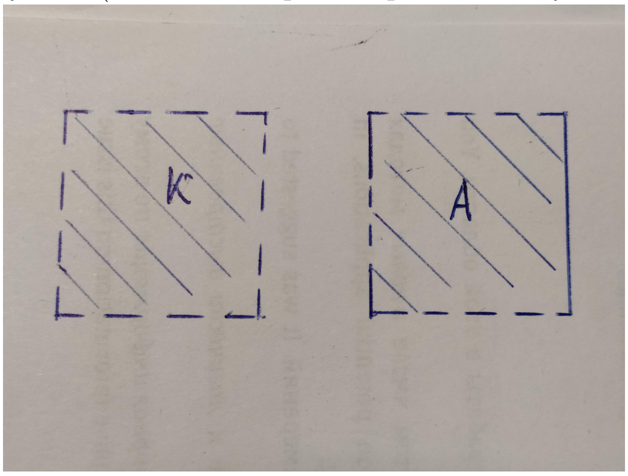
$df = \dfrac{d < Ax, x >}{< Ax, x >} = \dfrac{< (A^T + A)x, dx >}{< Ax, x >} = \dfrac{< 2Ax, dx >}{< Ax, x >} \Rightarrow \bigtriangledown f(x) =$
$= \dfrac{2Ax}{< Ax, x >}$

# Convex sets

1) Факт: $intA \subset A$. Пусть A - выпуклое множество. Докажем от противного, что $intA$ - выпуклое, т.е. предположим, что $intA$ - невыпуклое.$\Rightarrow$ $\exists x_1, x_2 \in intA\ \exists \lambda \in [0,1] : \tilde{x} = \lambda x_1 + (1-\lambda) x_2 \notin intA \Rightarrow \exists x_1, x_2 \in intA\ \exists \lambda \in [0,1] : \tilde{x} = \lambda x_1 + (1-\lambda) x_2 \notin A$. Противоречие с выпуклостью

множества А.

Контпримером покажем, что из выпуклости int A не следует выпуклость А. Множество К - квадрат без границ - выпуклое множество, т.к. оно открытое и $\forall x, y \in K\ \exists U_1(x) \subset K,\ \exists U_2(y) \subset K \Rightarrow$ отрезок $[x, y] \subset K$ (очевидно из рисунка ниже). При этом $K = intA$, но множество А - не выпуклое (на левой стороне отрезки между точками не лежат в множестве).




2) $\forall A, B \in S^n_{++} \forall \lambda \in [0, 1] \hookrightarrow \lambda A + (1-\lambda)B = M \in S^n_{++}$, поскольку $\lambda A$ и $(1-\lambda)B$ - это симметричные матрицы, то М - тоже симметричная, а также положительно определенная, т.к. $A \succ 0\ \&\ B \succ 0 \Leftrightarrow \forall x \in R^n \hookrightarrow x^T A x > 0\ \&\ x^T B x > 0 \Rightarrow \forall \lambda \in (0, 1) \hookrightarrow x^T(\lambda A + (1-\lambda)B)x = \lambda x^T A x + (1-\lambda)x^T B x > 0$ и при $\lambda = 0, 1$ это очевидно. В итоге, $S^n_{++}$ - выпуклое.

3)
$S = \{x \in R^n_+ | \prod_i x_i \geq 1\}$
$\forall x, y \in S\ \forall \lambda \in [0, 1] \hookrightarrow \tilde{x} = \lambda x + (1-\lambda)y \in S \Leftrightarrow \prod_i \tilde{x}_i \geq 1\ \&\ \tilde{x}_i \geq 0 \Leftrightarrow \lambda x_i + (1-\lambda)y_i \geq 0$ (т.к. $x, y \in R^n_+$) $\&\ \prod_i (\lambda x_i + (1-\lambda)y_i) \geq \prod_i x_i^\lambda y_i^{1-\lambda} = \left(\prod_i x_i\right)^\lambda \cdot \left(\prod_i y_i\right)^{1-\lambda} \geq 1$
Получили, что S - выпуклое множество.

4)
S - выпуклое $\Leftrightarrow \forall \alpha \geq 0, \beta \geq 0 \hookrightarrow (\alpha + \beta)S = \alpha S + \beta S$

$\Rightarrow$

S - выпуклое $\Leftrightarrow \forall \lambda \in [0, 1] \forall x, y \in S \hookrightarrow \lambda x + (1-\lambda)y \in S$. Пусть $\lambda = \dfrac{\alpha}{\alpha + \beta}$, где $\alpha > 0\ \&\ \beta > 0$ - произвольные. При этом получаем произвольное $\lambda \in (0,1)$. Тогда $\forall x, y \in S \hookrightarrow \lambda x + (1-\lambda)y \in S \Leftrightarrow \dfrac{\alpha}{\alpha+\beta}x + \dfrac{\beta}{\alpha+\beta}y \in S \Rightarrow \exists z \in S : (\alpha+\beta)z = \alpha x + \beta y \Rightarrow \alpha S + \beta S \subset (\alpha+\beta)S$. Второе вложение $(\alpha+\beta)S \subset \alpha S + \beta S$ тоже выполняется, поскольку $\forall x \in S \hookrightarrow (\alpha+\beta)x = \alpha x + \beta x$. Из двух вложений получаем равенство множеств. (при $\alpha = 0, \beta > 0 \hookrightarrow \beta S = \beta S$, при $\alpha = \beta = 0 \hookrightarrow 0 = 0 + 0$ - это два тождества)

$\Leftarrow$

$\forall \alpha, \beta > 0\, \forall x, y \in S \hookrightarrow \alpha x + \beta y \in (\alpha+\beta)S \Rightarrow \exists z \in S : (\alpha+\beta)z = \alpha x + \beta y \Leftrightarrow \dfrac{\alpha}{\alpha+\beta}x + \dfrac{\beta}{\alpha+\beta}y = z; \lambda = \dfrac{\alpha}{\alpha+\beta} \Rightarrow \forall \lambda \in (0,1) \hookrightarrow \lambda x + (1-\lambda)y = z \Rightarrow S-$ выпуклое (т.к. при $\lambda = 0, 1$ очевидно $x, y \in S$)

5) 1.

$S = \{p | P(x > \alpha) \leq \beta\} \Leftrightarrow \sum_{i:a_i>\alpha} p_i \leq \beta$ Так как индексы суммирования одни и те же для всех $p \in S$, то имеем
$\forall x, y \in S\ \forall \lambda \in [0,1] \hookrightarrow \sum_{i:a_i>\alpha} (\lambda x_i + (1-\lambda)y_i) = \lambda \sum_{i:a_i>\alpha} x_i +$
$(1-\lambda) \sum_{i:a_i>\alpha} y_i \leq \beta$, значит, множество выпуклое.

2.

$S = P \cap \{p | E|x^{201}| \leq \alpha E|x|\}$
$E|x^{201}| \leq \alpha E|x| \Leftrightarrow \sum_i^n p_i |a_i^{201}| - \alpha \sum_i^n p_i |a_i| \leq 0 \Leftrightarrow \sum_i^n p_i |a_i| (a_i^{200} - \alpha) \leq 0$

$\forall \lambda \in [0,1] \forall x, y \in S \hookrightarrow \lambda x + (1-\lambda)y \in S \Leftarrow \sum_i^n (\lambda x_i + (1-\lambda)y_i)|a_i|(a_i^{200} - \alpha) = \lambda \sum_i^n x_i |a_i| (a_i^{200} - \alpha) + (1-\lambda) \sum_i^n y_i |a_i| (a_i^{200} - \alpha) \leq 0 \Rightarrow S-$выпуклое.

3.

$Ex^2 \geq \alpha$
$M = \{p | \sum_i p_i a_i^2 \geq \alpha\ \&\ p \in P\}$
$\forall \lambda \in [0,1]\ \forall x, y \in M \hookrightarrow \lambda x + (1-\lambda)y \in M \Leftrightarrow \sum_i (\lambda x_i + (1-\lambda)y_i)a_i^2 = \lambda \sum_i x_i a_i^2 + (1-\lambda) \sum_i y_i a_i^2 \geq \lambda\alpha + (1-\lambda)\alpha = \alpha$
Значит, M - выпуклое.

4.-

# Projection

Пусть $M = \{X \in R^{m\times n} | rk(X) \leq k\}$, где $k \leq min\{m,n\}$, и $X_0 \in R^{m\times n}, rk(X_0) = r$.
Используем сингулярное разложение матрицы $X_0 = UDV = \sum_{i=1}^{r} q_i u_i v_i^T, q_i = \sqrt{\lambda_i(X_0^* X_0)}$.
Тогда ее проекцией на М по сингулярной норме $\parallel M \parallel_2 = q_{max}(M)$ будет матрица $Y = \sum_{i=1}^{min\{k,r\}} q_i u_i v_i^T$ так как очевидно, что $\parallel Y - X_0 \parallel_2 \leq \parallel A - X_0 \parallel_2 \forall A \in M$ так как максимальное число сингулярных значений занулится.
Ответ для нормы Фробениуса такой же поскольку матрицы U и V - ортогоналны и верно $\parallel UA \parallel_F = \parallel A \parallel_F$
Тогда $\parallel UDV - A \parallel_F = \parallel D - U^T A V^T \parallel_F, rk(D) = rk(X_0) \Rightarrow$
$Y = argmin_A(\parallel X_0 - A \parallel_F)$.

# Convex functions

1)
1.
$X \in S_{++}^n$
$tr(X^{-1}) =< X^{-T}, I >=< X^{-1}, I >$

$$d < X^{-1}, I >=< dX^{-1}, I >=< -X^{-1} \cdot dX \cdot X^{-1}, I >$$

$d^2(< X^{-1}, I >) =< X^{-1} \cdot dX_2 \cdot X^{-1} \cdot dX_1 \cdot X^{-1}, I > + < X^{-1} \cdot dX_1 \cdot X^{-1} \cdot dX_2 \cdot X^{-1}, I >$

Исследуем соответствующую $d^2[tr(X^{-1})]$ квадратичную форму на положительную полуопределённость (тут произвольная $H \in S_{++}^n$):

$$D^2[tr(X^{-1})][H,H] = 2 < X^{-1} \cdot H \cdot X^{-1} \cdot H \cdot X^{-1}, I >= 2 \cdot tr(X^{-1} \cdot H \cdot X^{-1} \cdot H \cdot X^{-1}) =$$

$$= 2 \cdot tr(X^{-1} \cdot H \cdot X^{-\frac{1}{2}} \cdot X^{-\frac{1}{2}} \cdot H \cdot X^{-1}) = 2 < X^{-\frac{1}{2}} \cdot H \cdot X^{-1}, X^{-\frac{1}{2}} \cdot H \cdot X^{-1} \geq 0$$

Значит, функция $tr(X^{-1})-$ выпуклая.(Здесь использовано свойство извлечения квадратного корня из симметричной полодительно определенной матрицы *wikipedia* $(*)$)

2.
Можем использовать свойство "Reduction to a line так как множество $S_{++}^n-$ выпуклое (показано в этой домашке) и свойтсво $(*)$. Пусть $V \in R_{++}^n-$ произвольная матрица. Тогда

$$f(X + Vt) = (det(X + Vt)^{\frac{1}{n}} = (det(X^{\frac{1}{2}} \cdot X^{\frac{1}{2}} + X^{\frac{1}{2}} X^{-\frac{1}{2}} V X^{-\frac{1}{2}} X^{\frac{1}{2}} t)^{\frac{1}{n}} =$$

$$= (det(X^{\frac{1}{2}}(I + X^{-\frac{1}{2}} V X^{-\frac{1}{2}} t) X^{\frac{1}{2}})^{\frac{1}{n}} = (det X \cdot det(I + X^{-\frac{1}{2}} V X^{-\frac{1}{2}} t))^{\frac{1}{n}} =$$

$$= (detX)^{\frac{1}{n}} \cdot \left( \prod_{i=1}^{n} \left( 1 + t\lambda_i (X^{-\frac{1}{2}} V X^{-\frac{1}{2}}) \right) \right)^{\frac{1}{n}}$$

Так как $X \in S^n_{++}$, то $detX > 0$. Матрица $X^{-\frac{1}{2}}VX^{-\frac{1}{2}} \in S^n_{++}$, поскольку $\forall x \in R^n \hookrightarrow x^T X^{-\frac{1}{2}} V X^{-\frac{1}{2}} x = (X^{-\frac{1}{2}}x)^T V (X^{-\frac{1}{2}}x) = u^T V u > 0$, поскольку $V \in S^n_{++}$. Тогда верно $\forall i = 1..n \hookrightarrow \lambda_i(X^{-\frac{1}{2}}VX^{-\frac{1}{2}}) > 0$. Тогда из вогнутости среднего геометрического (дальше показано) получаем, что данная функция аргумента $t$ вогнута на множестве $\{t | X + Vt \in S^n_{++}\}$.Значит, исходная функция от матрицы тоже вогнута.

2)

$$f(p) = \sum_{1}^{n} p_i \log p_i; \; \bigtriangledown f = (1 + \log p_1, \cdots, 1 + \log p_k, \cdots, 1 + \log p_n)^T$$

Т.к. $\forall i \hookrightarrow p_i > 0$ и $f"(p) = diag\left(\frac{1}{p_1}, \cdots, \frac{1}{p_i}, \cdots, \frac{1}{p_n}\right) \succ 0 \Rightarrow \forall p \neq q \in R^n_{++} \hookrightarrow f(p) > f(q) + \bigtriangledown^T f(q)(p-q)$ (строгая выпуклость). Этот факт доказывает пункты $\forall p, q \in R^n_{++} \hookrightarrow D(p,q) > 0$ и $D(p,q) = 0 \Rightarrow p = q$. Утверждение $p = q \Rightarrow D(p,q) = 0$ проверяется непосрелсственной подстановкой в выражение $D(p,q) = \sum_{i=1}^n p_i \log \frac{p_i}{q_i} - p_i + q_i$.

3)

1.

$Ex = \sum_i a_i p_i = a^T p$

$$\forall p_1, p_2 \in P \hookrightarrow a^T(\lambda p_1 + (1-\lambda)p_2) = \lambda a^T p_1 + (1-\lambda) a^T p_2$$

Поэтому функция и вогнута, и выпукла (нестрого в обоих случаях).

2.

$$P(x \geq a) = \sum_{i: a_i \geq a} p_i$$

Поскольку индексы суммирования одни и те же для любого элемента p вероятностного симплекса, то из линейности данной функции снова получаем и выпуклость, и вогнутость аналогично пункту 1.

3. $P(a \leq x \leq b$ Аналогично, выпукла и вогнута из-за линейности.

4.$\sum_{i=1}^n p_i \log p_i$ показана выпуклость в номере 2).

5.$V(x) = E(x^2) - (Ex)^2 = \sum_i p_i a_i^2 - (\sum_i p_i a_i)^2 = b^T - (pTa)^2 = b^T - pTaa^T p$
$-V(x) = f(p), A = aa^T, b_i = a_i^2, A \in S_+^n$
$df = -d < b, p > +d < p, Ap >= - < b, dp > + < (A + A^T)p, dp >$
$d^2 f = 2 < Adp_2, dp_1 >\Rightarrow f"(p) = 2A \succeq 0 \Rightarrow f-$ выпуклая, значит, $V(x)-$ вогнутая.

6.-

4)
1.
$$a(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \frac{\partial a}{\partial x_i} = \frac{1}{n} \Rightarrow \bigtriangledown a = \frac{1}{n}\bar{1} \Rightarrow \bigtriangledown^2 a = 0_{n\times n}$$
Значит, функция и выпукла, и вогнута.

2. $g(x) = \prod_{i=1}^{n} (x_i)^{\frac{1}{n}}$
Так как степень корня произвольнная, то для корректности задачи подразумевается, что числа $x_i$ положительные.
Воспользуемся неравенством средних $\prod_{i=1}^{n} (x_i)^{\frac{1}{n}} \le \dfrac{1}{n}\sum_{i=1}^{n} x_i$.
Любое положительное число можем представить как отношение двух ненулевых чисел $\forall i \hookrightarrow x_i = \dfrac{a_i}{b_i}$.
Имеем $\dfrac{\partial g}{\partial x_i} = \dfrac{1}{n} x_i^{\frac{1}{n}-1} \prod_{i=1;j\ne i}^{n} (x_i)^{\frac{1}{n}} = \dfrac{1}{nx_i} \prod_{i=1}^{n} (x_i)^{\frac{1}{n}}$
Тогда из неравенства имеем
$$\prod_{i=1}^{n}\left(\frac{a_i}{b_i}\right)^{\frac{1}{n}} \le \frac{1}{n}\left(\sum_{i=1}^{n}\frac{a_i}{b_i}\right)+1-1 \Rightarrow \prod_{i=1}^{n}(a_i)^{\frac{1}{n}} \le \prod_{i=1}^{n}(b_i)^{\frac{1}{n}} + \prod_{i=1}^{n}(b_i)^{\frac{1}{n}}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{a_i}{b_i}-1\right) =$$
$$= \prod_{i=1}^{n}(b_i)^{\frac{1}{n}} + \sum_{j=1}^{n}\frac{1}{nb_j}\prod_{i=1}^{n}(b_i)^{\frac{1}{n}}(a_j - b_i)$$
Заметим здесь,что
$$\prod_{i=1}^{n}(a_i)^{\frac{1}{n}} = g(a);\ \sum_{j=1}^{n}\frac{1}{nb_j}\prod_{i=1}^{n}(b_i)^{\frac{1}{n}}(a_j - b_i) = \bigtriangledown^T g(b)(a-b) \Rightarrow$$
$$\Rightarrow g(a) \le g(b) + \bigtriangledown^T g(b)(a-b)$$
Поэтому функция $g(x)-$ вогнутая.

5)

Рассмотрим двумерную функцию f (при n=2). При этом имеем $f(x_1, x_2) = \dfrac{1}{x_1 - \dfrac{1}{x_2}}$

Воспользуемся теоремой из книжки Бойда (обозначена 3.10 в разделе операции, сохраняющие выпуклость), утверждающую, что если $f(x) = h(g(x))$ и f выпуклая если h выпцклая невозрастающая и g вогнутая. В нашем случае $g(x_2) = x_1 + (-\dfrac{1}{x_2})-$ сумма двух вогнутых функций есть вогнутая функция. А функция $h(t) = \dfrac{1}{t}-$ выпуклая убывающая. В итоге, f - выпуклая для двумерного случая.

При повышении размерности получаем рекурентную формулу: $f_3(x_1, x_2, x_3) = \dfrac{1}{x_3 - f_2(x_1, x_2)}$, где повторим все те же рассуждения, зная, что $f_2(x_1, x_2)-$ выпуклая. Так дял любой $f_n(x_1, \cdots, x_n)$ покажем, что она выпукла.

6)

$f(x) = -x \ln x = (1 - x) \ln 1 - x = -g(x)$

Область определения функции $x \in [0, 1]$. Имеем $g"(x) = \dfrac{1}{x(1-x)} \Rightarrow g"(x) > 0 \Leftrightarrow x \in [0, 1]$. Значит, функция $g$ выпукла на своей области определения, а $f-$ вогнута.