# Домашнее задание №1

Бондарь София
Группа Б05-021

## Выпуклые множества

## Задача №1

### Условие:

Докажите, что множество C выпукло $\Leftrightarrow (\alpha+\beta)C = \alpha C + \beta C \ \forall \ \alpha, \beta \geq 0$

### Решение:

$\Rightarrow$

Пусть $x \in (\alpha+\beta)C$. Это значит, что $\exists c \in C$ такой, что $x = \alpha c + \beta c$. При этом, очевидно, что $\alpha c \in \alpha C$ и $\beta c \in \beta C \Rightarrow (\alpha+\beta)C \subset \alpha C + \beta C$

Пусть $x \in \alpha C + \beta C$. Это значит, что $\exists c_1, c_2 \in C$ такие, что $x = \alpha c_1 + \beta c_2$. Положим $\alpha_1 = \frac{\alpha}{\alpha+\beta}, \beta_1 = \frac{\beta}{\alpha+\beta}$, тогда $\alpha_1 + \beta_1 = 1$. C - выпукло по условию $\Rightarrow \alpha_1 c_1 + \beta_1 c_2 \in C \Rightarrow x = \alpha c_1 + \beta c_2 = (\alpha+\beta) * (\alpha_1 c_1 + \beta_1 c_2) \in (\alpha+\beta) * C \Rightarrow (\alpha+\beta)C \supset \alpha C + \beta C$ .

$\Leftarrow$

Пусть $(\alpha+\beta)C = \alpha C + \beta C$ . Возьмем $x_1, x_2 \in C, \alpha \in (0;1)$. Тогда $\alpha x_1 + (1-\alpha)x_2 = (\alpha + 1 - \alpha)c = c, c \in C \Rightarrow C-$ выпуклое по определению. $\square$

# Задача №2

## Условие:

Проверьте выпуклость следующих множеств:

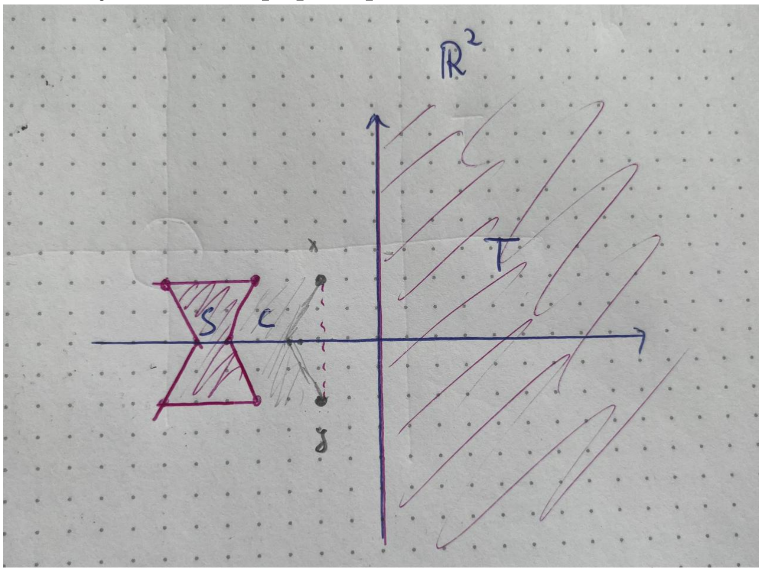
a) $C = \{x \in \mathbb{R}^n | dist(x, S) \leq dist(x, T)\}$ , где $S, T \subseteq \mathbb{R}^n$ - некоторые множества.

b) $C = \{x \in \mathbb{R}^n | x + S_1 \subseteq S_2\}$ где $S_1, S_2 \subseteq \mathbb{R}^n$ - некоторые множества, причем $S_2$ - выпукло.

c) $C = \{x \in \mathbb{R}^n | \|x - x_0\| \leq dist(x, T)\}$, где $T \subseteq \mathbb{R}^n$ - некоторое множество, $x_0 \in \mathbb{R}^n$

## Решение:

a) Не выпукло - контрпример в $\mathbb{R}^2$




b)

Пусть $x, y \in C, \alpha \in (0;1)$. Рассмотрим $\alpha x+(1-\alpha)y+S_1 = \alpha x+(1-\alpha)y+\alpha S_1-\alpha S_1+S_1 = \alpha(x+S_1)+(1-\alpha)(y+S_1)$. Так как $x, y \in C \Rightarrow (x+S_1) \subset S_2, (y+S_1) \subset S_2 \Rightarrow \alpha x+(1-\alpha)y+S_1 \subset S_2 \Rightarrow \alpha x + (1-\alpha)y \in C \Rightarrow C$ - выпуклое по определению

c)

Пусть $x, y \in C, \alpha \in (0;1)$. Рассмотрим $\alpha x + (1-\alpha)y$.

$\|\alpha x + (1-\alpha)y - x_0\| = \|\alpha x - \alpha x_0 + \alpha x_0 + (1-\alpha)y - x_0\| = \|\alpha(x-x_0) + (1-\alpha)(y-x_0)\| \leq \alpha\|x - x_0\| + (1-\alpha)\|y - x_0\| \leq \alpha dist(x, T) + (1-\alpha)dist(y, T) \leq dist(\alpha x + (1-\alpha)y, T) \Rightarrow \alpha x + (1-\alpha)y \in C \Rightarrow C$ - выпукло $\square$

# Задача №3

## Условие:

Матрица $D \in S^n$ - называется евклидовой, если найдется набор точек $x_1, ..., x_n \in \mathbb{R}^n$, такой что $D_{ij} = \|x_i - x_j\|_2^2 \ \forall i, j$. Известен следующий критерий: матрица $D$ является евклидовой тогда и только тогда, когда $D_{ii} = 0$ и $x^T D x \leq 0$ для всех x, таких что $\sum_{i=1}^n x_i = 1$. Покажите, что множество евклидовых матриц есть выпуклый конус.

## Решение:

Запишем определение выпуклого конуса. K - выпуклый конус, если

$\forall x_1, x_2 \in K \ \forall \theta_1, \theta_2 \geq 0 \Rightarrow \theta_1 x_1 + \theta_2 x_2 \in K$

Тогда обозначим за D - множество всех евклидовых матриц и возьмем $A, B \in D$. Возьмем также произвольные $\theta_1, \theta_2 \geq 0$.

Пусть $C = \theta_1 A + \theta_2 B$, хотим показать, что $C \in D$, тогда D - выпуклый конус по определению. Воспользуемся данным критерием и покажем, что оба условия выполняются.

- $C_{ii} = \theta_1 A_{ii} + \theta_2 B_{ii} = 0 + 0 = 0$
- Зафиксируем произвольный x такой что $\sum_{i=1}^n x_i = 1$. Тогда $x^T(\theta_1 A + \theta_2 B)x = x^T \theta_1 A x + x^T \theta_2 B x = \theta_1 x^T A x + \theta_2 x^T B x \leq 0$, ибо по условию $A, B \in D \Rightarrow x^T A x \leq 0, x^T B x \leq 0$, а также $\theta_1, \theta_2 \geq 0$.

Получается мы показали по критерию, что $C \in D \Rightarrow D$ - выпуклый конус $\square$

# Задача №4

## Условие:

Рассмотрим конус $K = \{0\} \cup \{x \in \mathbb{R}^n | x_1 = ... = x_k = 0, x_{k+1} > 0$ для некоторого $1 \leq k \leq n-1\}$, т.е. множество всех таких векторов x, что либо это нулевой вектор, либо первая ненулевая координата положительна. Найдите двойственный конус для этого конуса.

## Решение:

Запишем определение двойственного конуса. Если K - конус, то двойственный конус $K^* = \{y | y^T x \geq 0, \forall x \in K\}$.

Очевидно, что $0 \in K^*$.

Далее , рассмотрим все такие элементы z , $z = (0, ..0, t), t \geq 0$. Очевидно, что по определению все такие элементы лежат в K. Но тогда, чтоб все условия выполнялись, в двойственном конусе $K^*$ должны лежать элементы только с неотрицательной последней координатой. Теперь рассмотрим все такие элементы z, $z = (0, ..0, t, u), t \geq 0, u \leq 0$. Очевидно, что по определению все такие элементы лежат в K. И так как только что мы выяснили, что у всех элементов двойственного конуса последняя координата положительна, то чтоб $y^T z \geq 0$ надо, чтоб предпоследняя координата двойственного конуса была положительна, иначе $y_{n-1} z_{n-1} + y_n z_n < 0$. Но при этом, еcди мы устремим значение $z_n$ к -$\infty$, то для положительного произведения надо будет подгонять и элементы двойственного конуса, а значит все предыдущие не будут подходить, и так как $\forall y = (y_1, ..., y_n)$ можно таким способом подобрать элемент конуса z, что $y^T z < 0$. Однако, не стоит забывать про вырожденный случай когда мы доходим до первой координаты, когда $y = (t, 0, ..., 0), t \geq 0$, в этом случае, на любом элементе конуса произведение выйдет неотрицательным, так как первая координата никогда не может быть отрицательной, в отличие от всех остальных, но это значит, что двойственный конус просто-напросто не содержит элементов, отличных от 0 и векторов вида $y = (t, 0, ..., 0), t \geq 0$.

## Ответ:

$K^* = \{0\} \cup \{y = (t, 0, ..., 0) | t \geq 0\}$

# Матричное дифференцирование

# Задача №5

## Условие:

Пусть дана некоторая последовательность $\{x_k\}_{k=1}^{N} \subset \mathbb{R}^n$и некоторый постоянный вектор $y_0 \in \mathbb{R}^m$. Рассмотрим следующую рекурсивную модель:

$y_j(W, V, b) = \sigma(Wx_j + Vy_{j-1} + b), j \in \{1...N\}$

в которой $W \in \mathbb{R}^{m\times n}, V \in \mathbb{R}^{m\times m}, b \in \mathbb{R}^m$ - параметры модели, $\sigma(x) = \frac{1}{1+e^{-x}}$ - сигмоида (к вектору данная функция применяется поэлементно). Рассмотрим функцию:

$f(W, V, b) = \|y_N - Y\|_2^2$, где $Y \in \mathbb{R}^m$ есть некоторый константный вектор. Найдите градиенты функции f по W, V, b, как функции, зависящую только от этих параметров, $x_n, y_n$ и градиента от $y_{n-1}$ по этим параметрам.

## Решение:

$y_N = \sigma(Wx_N + Vy_{N-1} + b$

$d(y_N)_W = \sigma^{'}(y_N)(x_N + Vd(y_{N-1})_w)$

- $d(f)_W = d(\langle y_N - Y, y_N - Y\rangle) = 2\langle y_N - Y, d(y_N - Y)\rangle = 2\langle\sigma^{'}(y_N)(x_N + Vd(y_{N-1})_w), y_N - Y\rangle = \sigma^{'}(y_N)(x_N + Vd(y_{N-1})_w) \circ 2(y_N - Y)$
- $d(f)_V = 2\langle\sigma^{'}(y_N)(y_{N-1} + d(y_{N-1})_v), y_N - Y\rangle = \sigma^{'}(y_N)(y_{N-1} + d(y_{N-1})_v) \circ 2(y_N - Y)$
- $d(f)_b = 2\langle\sigma^{'}(y_N)(1 + Vd(y_{N-1})_b), y_N - Y\rangle = \sigma^{'}(y_N)(1 + Vd(y_{N-1})_b) \circ 2(y_N - Y)$

  $\square$

## Задача №6

### Условие:

Пусть $Y \in \mathbb{R}^{m\times n}, U \in \mathbb{R}^{m\times k}, V \in \mathbb{R}^{k\times n}$ - Найдите градиенты по U и V следующей функции $J(U,V) = \|UV - Y\|_F^2 + \frac{\lambda}{2}(\|U\|_F^2 + \|V\|_F^2)$, где $\|\bullet\|_F$- есть норма Фробениуса для матриц.

### Решение:

Норма Фробениуса определена следующим образом $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{ij}}$

Тогда если дифф-ть по V, получим, что $d(\|V\|_F^2) = d(\langle V, V\rangle) = 2\langle V, dV\rangle$.

Аналогично $d(\|U\|_F^2) = d(\langle U, U\rangle) = 2\langle U, dU\rangle$.

При этом фиксируя противоположную переменную:

$d(\|UV - Y\|_F^2)_v = d(\langle UV - Y, UV - Y\rangle)_v = 2\langle UV - Y, d(UV)\rangle = 2\langle U^T(UV - Y), dV\rangle$

$d(\|UV - Y\|_F^2)_u = d(\langle UV - Y, UV - Y\rangle)_u = 2\langle UV - Y, d(UV)\rangle = 2\langle (UV - Y)V^T, dU\rangle$

Тогда суммируя соответвующие слагаемые получим, что:

$\frac{\partial J(U,V)}{\partial V} = 2U^T(UV - Y) + \lambda V$

$\frac{\partial J(U,V)}{\partial U} = 2(UV - Y)V^T + \lambda U$

# Задача №7

## Условие:

Найдите градиент и гессиан функции $f(\mathbf{w}_1 \dots \mathbf{w}_K) = \sum\limits_{i=1}^{K} \log \frac{\exp(\langle \mathbf{w}_i, \mathbf{x} \rangle)}{\sum\limits_{k=1}^{K} \exp(\langle \mathbf{w}_k, \mathbf{x} \rangle)}$ по параметру $\mathbf{w}_j \in \mathbb{R}^n, j = \overline{1, K}$. Вектор $\mathbf{x}$ есть некоторый постоянный вектор из $\mathbb{R}^n$.

## Решение

Перепишем функцию f:

$f(\mathbf{w}_1 \dots \mathbf{w}_K) = \sum\limits_{i=1}^{K} (\langle \mathbf{w}_i, \mathbf{x} \rangle - ln \sum\limits_{k=1}^{K} \exp(\langle \mathbf{w}_k, \mathbf{x} \rangle))$

Тогда запишем компоненты градиента: Так как при фиксированном j все производные от второго слагаемого будет равны, то вторую сумму заменим на K одинаковых

$\frac{\partial f}{\partial w_j} = \mathbf{x} - K \frac{\mathbf{x} \exp(\langle \mathbf{w_j}, \mathbf{x} \rangle))}{\sum\limits_{k=1}^{K} \exp(\langle \mathbf{w}_k, \mathbf{x} \rangle)}$

Посчитаем вторую производную, тем самым вычислив $ij$ элемент гессиана

$\frac{\partial^2 f}{\partial w_i \partial w_j} = \frac{\partial}{\partial w_i} (\mathbf{x} - K \frac{\mathbf{x} \exp(\langle \mathbf{w_j}, \mathbf{x} \rangle))}{\sum\limits_{k=1}^{K} \exp(\langle \mathbf{w}_k, \mathbf{x} \rangle)}) = K \frac{\mathbf{x}\mathbf{x}^T \exp(\langle \mathbf{w}_i + \mathbf{w}_j, \mathbf{x} \rangle)}{(\sum\limits_{k=1}^{K} \exp(\langle \mathbf{w}_k, \mathbf{x} \rangle))^2}$

# Выпуклые функции

# Задача №8

## Условие:

Функция $f(x,t) = -\ln\left(t^2 - \mathbf{x}^\top \mathbf{x}\right)$, определённая на множестве $E = \left\{(\mathbf{x},t) \in \mathbb{R}^{n+1} \middle| \|\mathbf{x}\|_2 < t\right\}$ называется логарифмическим барьером для конуса второго порядка. Она используется при переходе от условной оптимизации с ограничением на конус второго порядка к безусловной с помощью введения штрафа за выход из этого конуса. Поэтому такие функции называются барьерными. Докажите, что функция f выпукла.

## Решение

Распишем доказательсво этой задачи "снизу - вверх чтоб было удобнее и не потеряться в переходах.

Для начала вспомним критерий сильвестра. Надо найти гессиан функции - матрицу вторых производных, определить знак с помощью критерия сильвестра. Далее с помощью критерия Сильвестра определяем, выпуклость функции: если определители угловых миноров гессиана всех порядков положительны, то функция выпукла вниз(вогнута); если знаки определителя угловых миноров гессиана всех порядков чередуются начиная с отрицательного, то функция выпукла вверх(выпукла); иначе неопределенность.

- Рассмотрим функцию $g(x,t) = \frac{x^2}{t}, t > 0$ и запишем ее гессиан:

  $Hess(\frac{x^2}{t}) = \begin{bmatrix} \frac{2}{t} & -\frac{2x}{t^2} \\ -\frac{2x}{t^2} & \frac{2x^2}{t^3} \end{bmatrix}$

  Первый главный минор $> 0$, второй $= 0 \Rightarrow$ функция $g(x,t) = \frac{x^2}{t}$ - выпуклая

- Теперь рассмотрим $\frac{x^T x}{t} = \sum_{i=1}^{n} \frac{x_i^2}{t}$ - которая является комбинацией выпуклых функций $\Rightarrow$ сама является выпуклой $\Rightarrow -\frac{x^T x}{t}$ - вогнута

- Функция $t - \frac{x^T x}{t}$ - также вогнутая, как линейная комбинация вогнутой и линейной функции $\Rightarrow$ функция $ln(t - \frac{x^T x}{t})$ - также является вогнутой как композиция вогнутой от вогнутой функции $\Rightarrow -ln(t - \frac{x^T x}{t})$ - выпукла

- Функция $-lnt$ - также является выпуклой, а значит, рассматривая функцию $f(x,t) = -lnt - ln(t - \frac{x^T x}{t}) = -ln(t(t - \frac{x^T x}{t})) = -ln(t^2 - \mathbf{x}^\top \mathbf{x})$ - она также является выпуклой, как сумма двух выпуклых

$\square$

# Задача №9

## Условие:

Докажите, что функция $f(\mathbf{x}) = \prod\limits_{k=1}^{n} x_k^{\alpha_k}$ является вогнутой на $\mathbb{R}^n_{++}$ для любых $\alpha_k$, таких что $\alpha_k \geq 0$ и $\sum\limits_{k=1}^{n} \alpha_k \leq 1$.

## Решение

Если доказать, что противоположная по знаку функция $g(x) = -f(x)$ - выпукла, то из этого будет следовать, что $f(x)$ - вогнута

Запишес теперь первые и вторые производгные функции g:

- $\frac{\partial g}{\partial x_i} = -\alpha_i x_i^{\alpha_i - 1} \prod\limits_{k=1, k\neq i}^{n} x_k^{\alpha_k}$

- $\frac{\partial^2 g}{\partial x_i^2} = -\alpha_i(\alpha_i - 1) x_i^{\alpha_i - 2} \prod\limits_{k=1, k\neq i}^{n} x_k^{\alpha_k}$

- $\frac{\partial^2 g}{\partial x_i \partial x_j} = -\alpha_i \alpha_j x_i^{\alpha_i - 1} x_j^{\alpha_j - 1} \prod\limits_{k=1, k\neq i,j}^{n} x_k^{\alpha_k}$

Запишем гессиан в матричном виде, так чтоб получались посчитанные элементы:

$H(x) = -\prod\limits_{k=1}^{n} x_k^{\alpha_k} * ((\mathbf{x}^{-1}\alpha^T)(\mathbf{x}^{-1}\alpha^T)^T - diag(\alpha_i x_i^{-2}))$ - если расписать все покоординатно, как раз получатся искомые производные.

Далее я хочу показать, что $\forall y, y^T H y \geq 0$

$y^T H y = -\prod\limits_{k=1}^{n} x_k^{\alpha_k} * (\langle \mathbf{x}^{-1}\alpha^T y, \mathbf{x}^{-1}\alpha^T y\rangle - \sum\limits_{i=1}^{n} (y_i x_i^{-1}\alpha_i)^2) = \prod\limits_{k=1}^{n} x_k^{\alpha_k} * (\sum\limits_{i=1}^{n} (y_i x_i^{-1}\alpha_i)^2 - \langle \mathbf{x}^{-1}\alpha^T y, \mathbf{x}^{-1}\alpha^T y\rangle)$

Оценим знак данного выражения:

- $\prod\limits_{k=1}^{n} x_k^{\alpha_k} > 0$

- Так как по условию $\sum\limits_{k=1}^{n} \alpha_k \leq 1$, то можно домножить и значение станет меньше. $\sum\limits_{i=1}^{n} (y_i x_i^{-1}\alpha_i)^2 - \langle \mathbf{x}^{-1}\alpha^T y, \mathbf{x}^{-1}\alpha^T y\rangle \geq (\sum\limits_{k=1}^{n} \alpha_k)(\sum\limits_{i=1}^{n} (y_i x_i^{-1}\alpha_i)^2) - \langle \mathbf{x}^{-1}\alpha^T y, \mathbf{x}^{-1}\alpha^T y\rangle \geq 0$

Последнее неравенство следует по нер-ву КБШ, из чего следует, что g - выпукла, а значит f - вогнута $\square$

# Задача №10

## Условие:

Докажите, что функция $f(X) = \mathrm{tr}(X^{-1})$ является выпуклой на множестве $\mathbb{S}^n_{++}$.

## Решение

Напомним, что $\mathbb{S}^n_{++}$ - симметрично положительно полуопределенные матрицы.

Введем вспомогательные матрицы $A, B$, такие что $A \in \mathbb{S}^n_{++}, B \in \mathbb{S}^n$, тогда для них можно записать спектральное разложение : $A^{-\frac{1}{2}}BA^{-\frac{1}{2}} = Q\Lambda Q^T$

Рассмтрим следующую функцию: $g(t) = f(A + tB) = \mathrm{tr}((A + tB)^{-1}) = \mathrm{tr}(A^{-1}(E + tA^{-\frac{1}{2}}BA^{-\frac{1}{2}})^{-1}) = \mathrm{tr}(A^{-1}Q(E + t\Lambda)^{-1}Q^T)) = \mathrm{tr}(Q^TA^{-1}Q(E + t\Lambda)^{-1}) = \sum\limits_{i=1}^{n}(Q^TA^{-1}Q)_{ii}(1 + t\lambda_i)^{-1}$

Функция $h(t) = \frac{1}{1+t\lambda_i}$ - выпукла, значит мы представили функцию g(t) в виде суммы выпуклых $\Rightarrow$ она сама выпуклая $\Rightarrow$ и f(X) - выпуклая

$\square$