1 Matrix calculus

Задача №1

Find $\bigtriangledown f(x), f(x) = \frac{1}{2}\|Ax - b\|_2^2, x \in \mathbb{R}^n$
Функцию можно переписать в виде:

$$f(x) = \frac{1}{2}(Ax - b)^T(Ax - b)$$

Воспользуемся свойством, что $d < x, x >= 2 < x, dx >$, тогда:

$$df(x) = \frac{1}{2}d(< Ax-b, Ax-b >) = \frac{1}{2}2 < Ax-b, d(Ax-b) >=< Ax-b, Adx >=< A^T(Ax-b), dx >$$

$=> \bigtriangledown f(x) = A^T(Ax - b)$

Задача №2

Find $\triangledown f(X)$, **if** $f(X) =< x, x >^{<x,x>}, x \in \mathbb{R}^n \setminus \{0\}$
Сделаем замену $g(x) =< x, x >$, тогда можно переписать функцию $f(x) = g(x)^{g(x)}$. Можем взять градиент от функции и умножить на градиент по самой функции, $d < x, x >= 2 < x, dx >$, тогда:

Задача №3

Calculate the Frobenious norm derivative: $\frac{\partial}{\partial X}||X||_F^2$
По свойству нормы Фробениуса верно, что $||X||_F^2 = \mathbf{tr}\left(X^T X\right) = \mathbf{tr}\left(X X^T\right) = \langle X, X\rangle$, тогда:

$$\frac{\partial}{\partial X}||X||_F^2 = \frac{\partial}{\partial X}\langle X, X\rangle$$

По определению производной скалярного произведения:

$$d||X||_F^2 = 2\langle X, dX\rangle => \frac{\partial}{\partial X}||X||_F^2 = 2X$$

Задача №4

Calculate the first and the second derivative of the following function $f : \mathbf{S} \to \mathbb{R}$

$$f(t) = det(A - tI_n),$$

where $A \in \mathbb{R}^{nxn},\ S := \{t \in \mathbb{R} : det(A - tI_n) \neq 0\}$.

Из свойств дифференцирования:

$$
\begin{aligned}
f(t) = det(A - tI_n) => df(t) = det(A - tI_n) < (A - tI_n)^{-T}, d(A - tI_n) >= \\
= det(A - tI_n) < (A - tI_n)^{-T}(-I_n)^T, dt) >=> \\
=> f'(t) = det(A - tI_n)(-I_n)^T(A - tI_n)^{-T}
\end{aligned}
$$

Задача №5

Смотри файл autograd.py

2 Convex sets

Задача №1

Prove that the set of square symmetric positive definite matrices is convex.

Докажем, что множество является выпуклым конусом. По определению: $\forall \theta_1, \theta_2 \geqslant 0$ и $\forall A, B \in \mathbb{S}^n_+$

$$\theta_1 A + \theta_2 B \in \mathbb{S}^n_+$$

Матрица, которая получена с помощью такой линейной комбинации остаётся квадратной и симметричной (сумма одинаковых чисел, домноженных на одинаковые константы).
Матрица положительно определена, по определению $\forall x \in \mathbb{R}^n$:

$$x^T \left(\theta_1 A + \theta_2 B\right) x = \theta_1 x^T A x + \theta_2 x^T B x \geq 0$$

так как $A \succeq 0 \to x^T A x \geqslant 0, B \succeq 0 \to x^T B x \geqslant 0$ и $\theta_1, \theta_2 \geq 0$
По определению - выпуклый конус => выпуклое.

Задача №2

Show, that conv$\{xx^T : x \in \mathbf{R}^n, \| x \| = 1\} = \{A \in \mathbf{S}^n_+ : \mathbf{tr}(A) = 1\}$.

Задача №3

Show that the hyperbolic set of $x \in \mathbb{R}^n_+ | \prod\limits_{i=1}^n x_i \geq 1$ **is convex. Hint: For** $0 \leq \theta \leq 1$ **it is valid, that** $a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b$ **with non-negative** a, b.

Возьмём такие a и b, чтобы они принадлежали множеству по определению: $a = \prod\limits_i x_i \geq 1$ и $b = \prod\limits_i y_i \geq 1$. Подставляем в неравенство:

$$\prod_i \left(\theta x_i + (1-\theta) y_i\right) = \theta \prod_i x_i + (1-\theta) \prod_i y_i \geq \prod_i x_i^\theta y_i^{1-\theta} = \left(\prod_i x_i\right)^\theta \left(\prod_i y_i\right)^{1-\theta} \geq 1$$

По определению выпуклого множества, $0 \leq \theta \leq 1$ и $\forall a, b \in S$, где S наше множество, выполнено, что $\theta a + (1-\theta) b \in S$ => показали выпуклость.

Задача №4

Prove, that the set $S \subseteq \mathbb{R}^n$ is convex if and only if $(\alpha+\beta)S = \alpha S+\beta S$ for all non-negative α, β

►⇒ Если $(\alpha + \beta)S = \alpha S + \beta S$ - множество S выпуклое.

1) По определению, $\forall x_1, x_2 \in S$ и $\forall \theta \in [0, 1] => \theta x_1 + (1 - \theta)x_2 \in S$ $\forall x \in \alpha S + \beta S$.
Представим в виде: $x = \alpha z_1 + \beta z_2$, где $z_1 \in S$и $z_2 \in S$

2) $x = (\alpha + \beta)z_3$, где $z_3 \in S$ (так как $(\alpha + \beta)S = \alpha S + \beta S$)
Выносим $\alpha + \beta$ из первого представления:
$$x = (\alpha + \beta)\left(\frac{\alpha}{\alpha + \beta}z_1 + \frac{\beta}{\alpha + \beta}z_2\right) \in S$$

3) Получили $z_3 = \frac{\alpha}{\alpha+\beta}z_1 + \frac{\beta}{\alpha+\beta}z_2 \in S$
Заметим, что $\frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta} = 1$. Пусть $\theta = \frac{\alpha}{\alpha+\beta} \Rightarrow \frac{\beta}{\alpha+\beta} = 1 - \theta$, следовательно:
$z_3 = \theta z_1 + (1 - \theta)z_2 \in S$ Так как $(\alpha + \beta)S = \alpha S + \beta S$ верно для произвольных неотрицательных α и β, то $\theta = \frac{\alpha}{\alpha+\beta} \in [0, 1]$. ◄

►⇐ S - выпуклое множество, доказываем, что $(\alpha + \beta)S = \alpha S + \beta S$. Доказываем два включения:

1) $(\alpha + \beta)S \subseteq \alpha S + \beta S$
Возьмём $z \in (\alpha + \beta)S => z = (\alpha + \beta)x = \alpha x + \beta x$, где $x \in S$. Тогда $\forall z \in S$ представим как $z = \alpha x + \beta x$. Это и есть наше включение: $(\alpha + \beta)S \subseteq \alpha S + \beta S$

2) $\alpha S + \beta S \subseteq (\alpha + \beta)S$
$z \in \alpha S + \beta S$, можно представить как: $z = \alpha z_1 + \beta z_2$, где $z_1, z_2 \in S$, можно записать, как $z = (\alpha + \beta)z_3$, где $z_3 \in S$.
Как и в доказательстве в другую сторону вынесем скобку $(\alpha + \beta)$. Получим:
$$z = (\alpha + \beta)\left(\frac{\alpha}{\alpha + \beta}z_1 + \frac{\beta}{\alpha + \beta}z_2\right)$$
Снова берём $\theta = \frac{\alpha}{\alpha+\beta} \Rightarrow \frac{\beta}{\alpha+\beta} = 1 - \theta => z = (\alpha + \beta)(\theta z_1 + (1 - \theta)z_2)$
S - выпуклое множество, значит $(\theta z_1 + (1 - \theta)z_2) = z_3 \in S, \forall \theta$
Тогда, $z = (\alpha + \beta) z_3 \in S$, то есть мы показали обратное включение.

Доказано, что S-выпуклая $\Rightarrow (\alpha + \beta)S = \alpha S + \beta S$ ◄

Задача №5

Let $x \in \mathbb{R}$ **is a random variable with a given probability distribution of** $\mathbb{P}(x = a_i) = p_i,$ **where** $i = 1, \dots, n, a_1 < \dots < a_n$. **It is said that the probability vector of outcomes of** $p \in \mathbb{R}^n$ **belongs to the probabilistic simplex, i.e.** $P = p \mid \mathbf{1}^T p = 1, p \succeq 0 = p \mid p_1 + \dots + p_n = 1, p_i \geq 0$. **Determine if the following sets of p are convex:**

1) $\mathbb{P}(x > \alpha) \leq \beta$
   В первых трёх пунктах можем сводить к ограничениям на p, как ограничения на полупространство из чего будет следовать, что множество выпуклое.
   По определению: $\mathbb{P}(x > \alpha) = \sum\limits_{i: a_i \geqslant \alpha} p_i$, следовательно

$$\mathbb{P}(x > \alpha) \leq \beta \Leftrightarrow \sum_{i: a_i \geqslant \alpha_i} p_i \leq \beta$$

   Геометрически это полупространство и оно выпуклое, значит, наше множество тоже выпуклое.

2) $\mathbb{E}|x^{201}| \leq \alpha \mathbb{E}|x|$ По определению матожидания:

$$\sum_{i=1}^{n} p_i \left(\left| a_i^{201} \right| - \alpha \left| a_i \right| \right) \leq 0$$

   Не важно, какой коэффициент, главное, чтобы относительно p было линейное выражение. Можно переписать как $\sum\limits_{i=1}^{n} c_i p_i$, где $c_i = (|a_i^{201}| - \alpha |a_i|)$ - некоторые числовые коэффициенты. Значит, геометрически оно задаёт полупростронство, а оно выпуклое.

3) $\mathbb{E}|x^2| \geq \alpha$
   По опредлению матожидания:

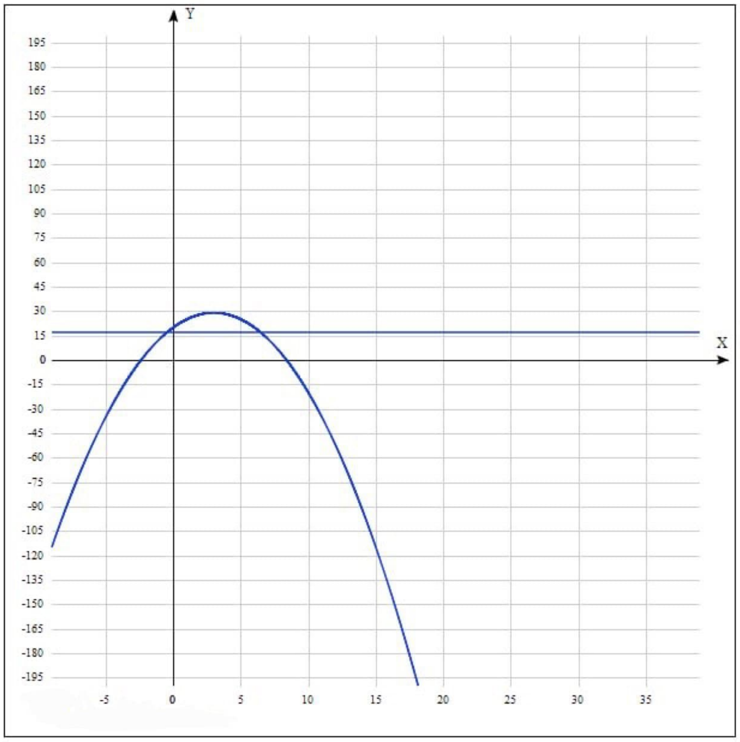
$$\sum_i p_i a_i^2 \geqslant \alpha$$

   Тогда из прошлых примеров можем сказать, что оно выпуклое.

4) $\mathbb{V}x \geq \alpha$
   По определению дисперсии:

$$\mathbb{V}x = \mathbb{E}x^2 - (\mathbb{E}x)^2 = \sum_{i=1}^{n} a_i^2 p_i - \left(\sum_{i=1}^{n} a_i p_i \right)^2 = b^T p - p^T A p \geqslant \alpha$$

   где $b_i = a_i^2$ и $A = aa^T$ - положительно определённая матрица. Заметим, что такое выражение геометрически задаёт параболу , ветви которой направлены вниз



и гиперплоскость α. Нашему множеству принадлежат точки, ограниченные этой гиперплоскостью. Можно утверждать о выпуклости через понятие надграфика: если надграфик функции является выпуклой фигурой, то функция - выпуклая, если выпуклым является подграфик, то функция вогнутая.
Перевернутая парабола - вогнутое => подграфик выпуклое множество.

3 Convex functions

Задача №1

Prove, that function $f(X) = \mathbf{tr}(X^{-1}), X \in S^n_{++}$ **is convex, while** $g(X) = (\det X)^{1/n}, X \in S^n_{++}$ **is concave.**

▶ Начнем с доказательства, что $g(x)$ вогнутая.
Функция возвращает скаляр => можем представить как:

$$g(t) = f(Z + tV), \text{ где } Z \succ 0 \text{ и } V \in \mathbb{S}^n$$

Получили функцию от скалярной переменной t, которая принимает значения, что $\{t \mid Z + tV \succ 0\}$.

$$\begin{aligned} g(t) &= (\det(Z + tV))^{1/n} \\ &= \left(\det Z^{1/2} \det\left(I + tZ^{-1/2}VZ^{-1/2}\right) \det Z^{1/2}\right)^{1/n} \\ &= (\det Z)^{1/n} \left(\prod_{i=1}^{n} (1 + t\lambda_i)\right)^{1/n} \end{aligned}$$

Где λ_i - собственные числа матрицы $Z^{-1/2}VZ^{-1/2}$.
Из последнего равенства наша функция $g(t)$ - вогнута на множестве $\{t \mid Z+tV \succ 0\}$, так как $\det Z > 0$, $Z \succ 0$ и геометрическое среднее $\prod\limits_{i=1}^{n} (x_i)^{1/n}$ - вогнуто, а $1 + t\lambda_i$ - вогнутая функция, то в силу линейности можно говорить, что $g(X)$ - вогнуто.

Для $f(X) = \mathbf{tr}(X^{-1}), X \in S^n_{++}$. Также представим в виде: $S(t) = Z+tV$, где $Z \in S^n_{++}$ и V - симметрична. Теперь рассматриваем функцию от скалярной переменной t. Достаточно показать (по дифференциальному критерию второго порядка), что для $t = 0$

$$\frac{d^2}{dt^2}\mathbf{tr}(S(t)^{-1})|_{t=0} \geq 0$$

Матричное разложение в ряд Тейлора в $t = 0$

$$S(t)^{-1} = (Z(I + tZ^{-1}V))^{-1} = Z^{-1} - tZ^{-1}VZ^{-1} + t^2Z^{-1}VZ^{-1}VZ^{-1} + \ldots$$

поэтому

$$\frac{d^2}{\partial t^2}\mathbf{tr}(S(t)^{-1})\bigg|_{t=0} = 2\mathbf{tr}(Z^{-1}VZ^{-1}VZ^{-1})$$

Но $Z^{-1}VZ^{-1}VZ^{-1} = UZ^{-1}U^T$, где $U = Z^{-1}V$ и Z^{-1} положительно определены (так как $Z \in S^n_{++}$ и V - симметрична), поэтому $UZ^{-1}U^T$ положительно полуопределена, а значит $\mathbf{tr}(UZ^{-1}U^T) \geq 0$.
Следовательно доказали, что $f(X) = \mathbf{tr}(X^{-1})$ - выпуклая функция. ◀

Задача №2

Kullback–Leibler divergence between $p, q \in \mathbb{R}^n_{++}$:

$$D(p,q) = \sum_{i=1}^{n}(p_i \log(p_i/q_i) - p_i + q_i)$$

Prove, that $D(p,q) \geq 0 \forall p, q \in \mathbb{R}^n_{++}$, $D(p,q) = 0 \leftrightarrow p = q$

Hint: $D(p,q) = f(p) - f(q) - \nabla f(q)^T(p-q), \quad f(p) = \sum\limits_{i=1}^{n} p_i \log p_i$

► По определению дивергенция Кульбака-Лейблера:

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

$$D(p,q) = f(p) - f(q) - \nabla f(q)^T(p-q), \quad f(p) = \sum_{i=1}^{n} p_i \log p_i$$

Тогда:

$$\frac{\partial f}{\partial p_i} = \log p_i + 1$$

$\nabla f(p) = \log p + 1$, где p - вектор. Значит:

$$D(p,q) = f(p) - f(q) - \nabla f(q)^T(p-q) = \sum_{i=1}^{n} p_i \log p_i - q_i \log q_i - (p_i - q_i)\log q_i - p_i + q_i$$

$$= \sum_{i=1}^{n}\left(p_i \log p_i/q_i - p_i + q_i\right)$$

Получается определение дивергенции Кульбака-Лейблера.По дифференциальному критерию первого порядка функция выпуклая тогда и только тогда, когда

$$f(p) \geqslant f(q) + \nabla f(q)^T(p-q) \Leftrightarrow D(p,q) = f(p) - f(q) - \nabla f(q)^T(p-q) \geqslant 0$$

Но функция $f(p) = \sum_{i=1}^{n} p_i \log p_i$ - выпуклая, так как это неотрицательная сумма функций $x \log x$ (задача №6) $=>$ $D(p,q) \geqslant 0$.
Если $p = q$, то подставив получим $D(p,q) = 0$ ◄

Задача №3

Let x be a real variable with the values $a_1 < a_2 < \ldots < a_n$ **with probabilities** $\mathbb{P}(x = a_i) = p_i$. **Derive the convexity or concavity of the following functions from p on the set of** $\left(p | \sum\limits_{i=1}^{n} p_i = 1, p_i \geqslant 0\right)$

1) $\mathbb{E}x$
Линейная функция вида $f(x)c^T x + b$ - выпуклая и вогнутая одновременно (доказывается по определению). По определению матожидания:

$$\mathbb{E}x = \sum_{i=1}^{n} a_i p_i.$$

Получили линейную функцию, следовательно, матожидание - выпуклая и вогнутая функция.

2) $\mathbb{P}\{x \geqslant \alpha\}$
Аналогично задаче на множества:

$$\mathbb{P}\{x \geqslant \alpha\} = \sum_{i:a_i \geq \alpha} p_i$$

Получили линейную функцию, следовательно - выпуклая и вогнутая функция.

3) $\mathbb{P}\{\alpha \leq x \leq \beta\}$
Рассматриваем прошлый пункт для другого диапазона:

$$\mathbb{P}\{\alpha \leq x \leq \beta\} = \sum_{i:\alpha \leq a_i \leq \beta} p_i$$

Получили линейную функцию, следовательно - выпуклая и вогнутая функция.

4) $\sum\limits_{i=1}^{n} p_i \log p_i$
В задаче 6 показали, что функция $x \ln x$ - выпуклая. Тогда наша функция - неотрицательная сумма выпуклых, сепарабельных функций. Тогда получили выпуклую функцию, ведь неотрицательная сумма сохраняет выпуклость.

5) $\mathbb{V}x = \mathbb{E}(x - \mathbb{E}x)^2 = \mathbb{E}x^2 - (\mathbb{E}x)^2$
Выпуклая функция:

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

Существует контрпример $n = 2, a_1 = 0, a_2 = 1, p_1 = (1, 0), p_2 = (0, 1), \theta = \frac{1}{2}$.

$$\mathbb{V}(p_1) = \mathbb{E}p_1^2 - (\mathbb{E}p_1)^2 = 0$$

$$\mathbb{V}(p_2) = \mathbb{E}p_2^2 - (\mathbb{E}p_2)^2 = 1 - 1 = 0$$

$$\mathbb{V}(p = (\frac{1}{2}, \frac{1}{2})) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$$

$\frac{1}{4} \nleq 0$ => дисперсия - вогнутая функция.

6) $\textbf{quartile}(x) = \inf\left(\beta | \mathbb{P}\{x \leqslant \beta\} \geqslant 0.25\right)$
   quartile(x) не является непрерывной функцией, так как x может принимать только дискретные значения $a_1 < a_2 < \ldots < a_n$ => она определена на дискритном множестве точек, которое не является выпуклым. Значит, она не является ни выпуклой, ни вогнутой.

Задача №4

Is the function returning the arithmetic mean of vector coordinates is a convex one: $a(x) = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$, **what about geometric mean:** $g(x) = \prod\limits_{i=1}^{n}(x_i)^{1/n}$?

$a(x) = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ - выпуклая функция. По дифференциальному критерию первого порядка:

$$a(x + \Delta x) \geq a(x) + \nabla a^T(x)\Delta x$$

Градиент берем по координатно, получаем: $\nabla a^T(x) = \left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$. Подставляем в критерий:

$$\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n}\Delta x_i \geq \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=s}^{n}\Delta x_i$$

Это неравенство верно, следовательно, среднее арифмитическое - выпуклая функция.

Геометрическое среднее, которое определяется как:

$$g(x) = \prod_{i=1}^{n}(x_i)^{1/n}$$

является вогнутой функцией. Воспользуемся дифференциальным критерием второго порядка, который звучит так: если мы покажем $\nabla^2 f(x) \preceq 0$, то $f(x)$ - вогнутая.
Если матрица гессиана отрицательно определена => для любого вектора v, верно

$$v^T\nabla^2 f(x) v \leqslant 0$$

Для начала подсчитаем гессиан. Подсчитаю сначала поэлементно, затем перепишу в матричном виде.
Певая производная:

$$\frac{\partial}{\partial x_k} f(x) = \frac{1}{n}\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}-1} \cdot \prod_{\substack{i=1\\ i\neq k}}^{n} x_i$$

Вторая производная, по этому же элементу:

$$\frac{\partial^2}{\partial x_k^2} f(x) = -\frac{n-1}{n^2}\left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}-2} \cdot \left(\prod_{\substack{i=1\\ i\neq k}}^{n} x_i\right)^2 = -\frac{n-1}{n^2} \times x_k^{\frac{1-2n}{n}}\left(\prod_{\substack{i=1\\ i\neq k}}^{n} x_i\right)^{\frac{1}{n}} =$$

$$= -\frac{n-1}{n^2 x_k^2} \left(\prod_{i=1}^{n} x_i \right)^{\frac{1}{n}}$$

Получаем, что

$$\frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1) \frac{\left(\prod_{i=1}^{n} x_i\right)^{1/n}}{n^2 x_k^2}$$

Вторая производная по x_l, где $k \neq l$

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{\left(\prod_{i=1}^{n} x_i\right)^{1/n}}{n^2 x_k x_l}$$

$$\nabla^2 f(x) = -\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2} \left(n \operatorname{diag}\left(1/x_1^2, \ldots, 1/x_n^2\right) - a a^T \right)$$

где $a_i = \frac{1}{x_i}$
Умножаем гессиан слева и справа на вектор v:

$$v^T \nabla^2 f(x) v = -\frac{\prod_{i=1}^{n} x_i^{1/n}}{n^2} \left(n \sum_{i=1}^{n} v_i^2 / x_i^2 - \left(\sum_{i=1}^{n} v_i / x_i \right)^2 \right) \leq 0$$

и он ≤ 0. Скобки в конце выражения - неравенство Коши-Буняковского, которое звучит как:

$$\left(\sum_{i=1}^{n} a_i \right)^2 \leq \left(\sum_{i=1}^{n} 1 \right) \sum_{i=1}^{n} a_i^2 = n \sum_{i=1}^{n} a_i^2$$

Так как $a_i = \frac{v_i}{x_i}$ => геометрическое среднее - вогнутая функция.

Задача №5

Is $f(x) = -x \ln x - (1-x) \ln(1-x)$ convex?

Рассмотрим отдельно $f(x) = x \ln x$, где x - скалярная переменная и покажем по дифференциальному критерию второго порядка, что она выпуклая.

$$f'(x) = \ln x + 1, \quad f''(x) = 1/x$$

Так как x стоит под логарифмом, то $x > 0$, следовательно, $f''(x) > 0$. Значит $f(x) = x \ln x$ строго выпуклая функция. Функция $g(x) = (1-x) \ln(1-x)$ тоже строго выпуклая при $1 - x > 0$.
Неотрицательная сумма выпуклых функций сохраняет выпуклость, поэтому $x \ln x + (1-x) \ln(1-x)$ - строго выпуклая функция. Но она стоит с минусом, значит, вогнутая по определению.

Задача №6

Let $f : \mathbb{R}^n \to \mathbb{R}$ be the following function:

$$f(x) = \sum_{i=1}^{k} x_{\lfloor i \rfloor},$$

where $1 \leqslant k \leqslant n$, while the symbol $x_{\lfloor i \rfloor}$ stands for the i-th component of sorted ($x_{\lfloor 1 \rfloor}$-maximum component of x and$x_{\lfloor n \rfloor}$-minimum component of x)
Show, that f is a convex function.

Можно рассмотреть как максимум различных сумм:

$$f(x) = \sum_{i=1}^{k} x_{\lfloor i \rfloor} = max\{x_{i_1} + ... + x_{i_r} \mid 1 \leq i_1 < i_2 < ... < i_r \leq n\},$$

Тогда можно говорить, что получили поточечный максимум для $\frac{n!}{r!(n-r)!}$ линейных функций. А так как функции линейны, можем сделать вывод, что функция f(x) будет выпуклой.

4 Conjugate sets

Задача №1

Let $\mathbb{A}_n$ be the set of all n dimensional antisymmetric matrices. Show that $(\mathbb{A}_n)^* = \mathbb{S}_n$.

Покажем вложения в обе стороны.
► $(\mathbb{A}_n)^* \subset \mathbb{S}_n$. Пусть матрица $B \in (\mathbb{A}_n)^*$,$\forall A$ - такой, что $A^T = -A$.

$$tr\left(A^T B\right) = tr\left(AB^T\right) \geq -1$$

Пользуясь тем, что $A^T = -A$:

$$tr\left(-AB\right) \geq -1$$
$$tr\left(AB^T\right) \geq -1$$

Складываем эти 2 неравенства:

$$tr\left(A \cdot (B^T - B)\right) \geq -2$$

$$tr\left(A \cdot (B - B^T)\right) \leq 2$$

Матрица $B \in \mathbb{S}_n$, то есть $B = B^T$. Доказательство от противного:
Пусть $B \neq B^T$ диагональ не меняется при транспонировании, значит, если вычесть эти матрицы - получим все 0 и хотя бы один ненулевой элемент. Матрица A была взята такой, чтобы у нее на диагонали были нули. Выбираем элементы так, чтобы на диагонали у матрицы $A(B - B^T)$ был хотя бы 1 элемент.
Можем это сделать всегда: пусть ненулевой элемент у матрицы $(B - B^T)$ будет в первом столбце: b_{i1} $(i \neq 1)$. Тогда в первой строке матрицы A выбираем $a_{1i} \neq 0$ а остальные равные нулю. При произведении получим ненулевой элемент на диагонали $c = b_{i1} \cdot a_{1i}$.
След матрицы равен сумме диагональных элементов. Пусть у $(B - B^T)$ только 1 ненулевой элемент, значит $tr\left(A \cdot (B - B^T)\right) = c$. Так как матрица A произвольная, то можем взять её с элементом $a_{1i} \cdot \frac{3}{c}$. Тогда $tr\left(A \cdot (B - B^T)\right) = 3 > 2$ получаем противоречие, что $B \in (\mathbb{A}_n)^*$.

Если в матрице $(B - B^T) > 1$ ненулевого элемента, нельзя доказать, что след равен 0: матрица $(B - B^T)$ - антисимметричная, для произвольной антисимметричной матрицы $B^* = (B - B^T)$ можно подобрать такую антисимметричную A, что $tr(AB^*) \neq 0$. Просто будем брать такую же и получаем отрицательно определённую (на диагонале будут стоять числа одного знака и след не занулится), применяем рассуждения для одного ненулевого элемента - и получаем противоречие. Значит, $B = B^T$, $B \in \mathbb{S}_n \to (\mathbb{A}_n)^* \subset \mathbb{S}_n$ ◄

▶ Вложение в другую сторону: пусть $B \in \mathbb{S}_n$, то есть $B = B^T$ а A - произвольная антисимметричная матрица, тогда:

$$tr(AB^T) = tr(AB)$$

$$tr(A^T B) = -tr(AB)$$

Получается, что $tr(AB) = -tr(AB) \to tr(AB) = 0 > -1$, значит $B \in (\mathbb{A}_n)^* \to (\mathbb{A}_n)^* \supseteq \mathbb{S}_n$
Получаем, что $(\mathbb{A}_n)^* = \mathbb{S}_n$ ◀

Задача №2

Find the conjugate set to the ellipsoid:

$$S = \left\{ x \in \mathbb{R}^n \mid \sum_{i=1}^{n} a_i^2 x_i^2 \leq \varepsilon^2 \right\}$$

Определение конуса через матрицу:

$$S = \left\{ x \in \mathbb{R}^n \mid \sum_{i=1}^{n} \left(\frac{a_i}{\varepsilon}\right)^2 x_i^2 \leq 1 \right\}$$

$A = \text{diag}(\frac{a_i}{\varepsilon})$ - диагональная матрица, тогда

$$\|Ax\|_2^2 = \sum_{i=1}^{n} \left(\frac{a_i}{\varepsilon}\right)^2 x_i^2$$

Задаем эллипс:

$$S = \{x \in \mathbb{R}^n \mid \|Ax\|_2 \leq 1\} = E = \left\{A^{-1}u \mid \|u\|_2 \leq 1\right\} \text{ (Бойд)}$$

Где $A^{-1} = diag\frac{\varepsilon}{a_i}$. Что эквивалентно, покажем через вложение множеств одно в другое:
$2 \to 1$: $x = A^{-1}u$, где $\|u\|_2 \leq 1$, подставим в S, тогда $\|AA^{-1}u\|_2 \leq 1$ - следовательно, $E \subseteq S$
$2 \to 1$: $\|Ax\|_2 \leq 1$, обозначим $y = Ax$, получается, что $\|y\|_2 \leq 1$. Тогда $x = A^{-1}y$, где $\|y\|_2 \leq 1$, следовательно $x \in E$ и $S \subseteq E$

По определению хотим найти все векторы $p \in E^*$, что $\forall x \in E \to \langle p, x\rangle \geq -1$. Любой вектор $x \in E$ представим как $x = A^{-1}u$, где $\|u\|_2 \leq 1$, => для p:

$$\langle p, A^{-1}u\rangle \geq -1, \text{ где } \|u\|_2 \leq 1$$

$$\langle A^{-T}p, u\rangle \geq -1, \text{ где } \|u\|_2 \leq 1$$

Из неравенства Коши-Буняковского:

$$|\langle A^{-T}p, u\rangle| \leq \|u\|_2 \cdot \|A^{-T}p\|_2 \leq [\|u\|_2 \leq 1] \leq \|A^{-T}p\|_2$$

Так как вектор u - произвольный, то равенство может достигаться всегда, поэтому нам подходят все p, для которых выполнено:

$$-1 \leq -\|A^{-T}p\|_2 \Leftrightarrow \|A^{-T}p\|_2 \leq 1$$

Такоенеравенство задает эллипс с другой матрицей $A^{-T} = \mathrm{diag}(\frac{\varepsilon}{a_i})$. Или в привычном виде:

$$E^* = \left\{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} \left(\frac{\varepsilon}{a_i}\right)^2 x_i^2 \leq 1\right\}$$

что эквивалентно

$$E^* = \left\{x \in \mathbb{R}^n \mid \sum_{i=1}^{n} \left(\frac{1}{a_i}\right)^2 x_i^2 \leq \left(\frac{1}{\varepsilon}\right)^2\right\}$$

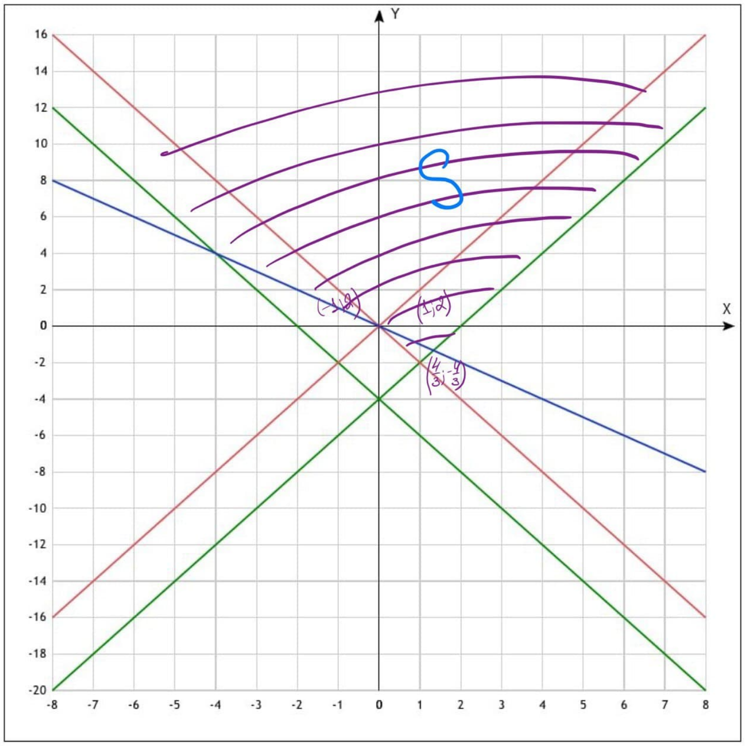
Задача №3

Find the sets S^*, S^{**}, S^{***},

$$S = \{x \in \mathbb{R}^2 \mid x_1 + x_2 \geq 0, \;\; 2x_1 + x_2 \geq -4, \;\; -2x_1 + x_2 \geq -4\}$$

S - замкнуто, выпукло и включает 0, то $S^{**} = S$, $S^{***} = S^*$. Смотрим на рисунках, что это верно:
Изображаем S:

$$S = \mathrm{conv}\left((-4, 4), (\frac{4}{3}, -\frac{4}{3})\right) + \mathrm{cone}\left((1, 2), (-1, 2)\right)$$

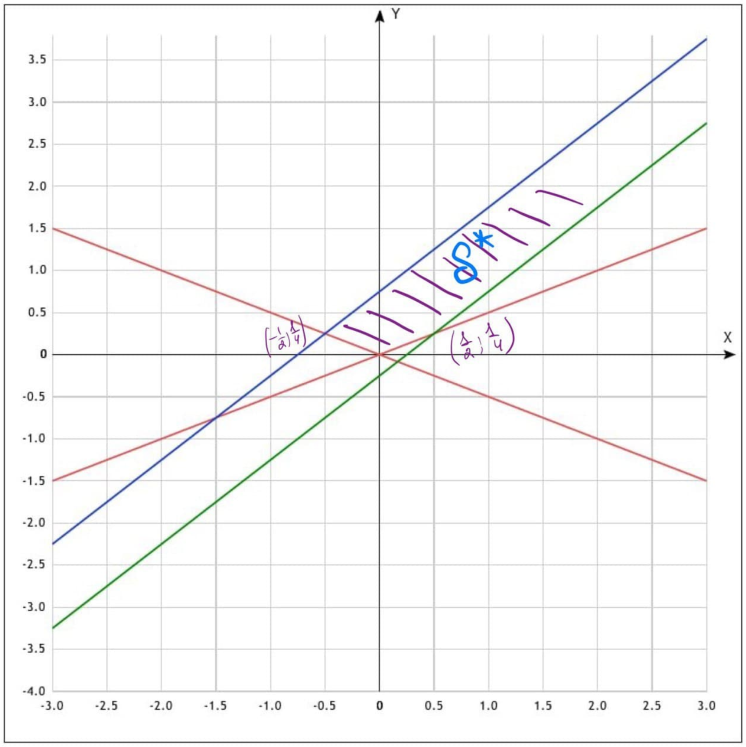


Получаем следующие неравенства:

$$\begin{cases} -4p^1 + 4p^2 \geq -1 \\ 4p^1 - 4p^2 \geq -3 \\ p^1 + 2p^2 \geq 0 \\ -p^1 + 2p^2 \geq 0 \end{cases}$$

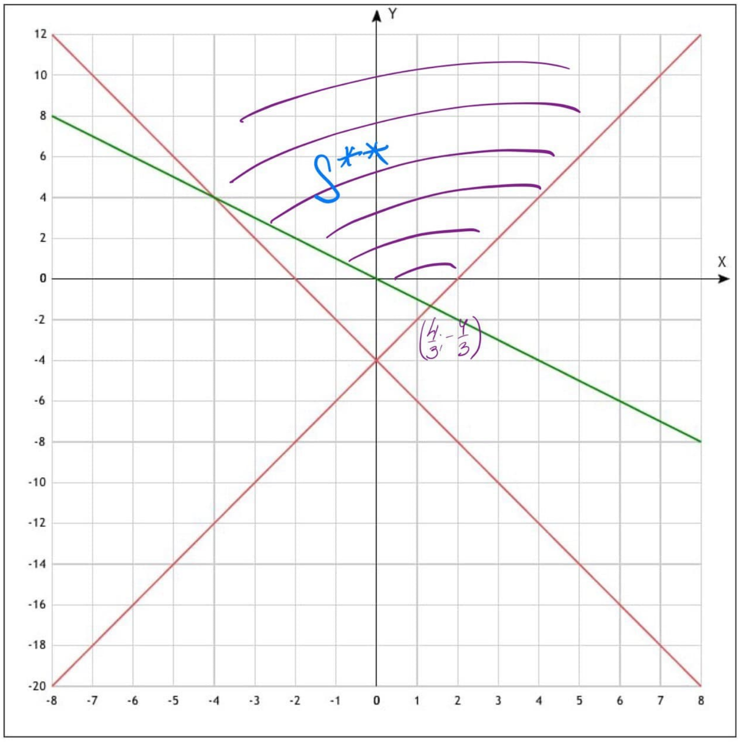
Изобразим S^*:

$$S^* = \operatorname{conv}\left(\left(-\frac{1}{2}, \frac{1}{4}\right), (0,0), \left(\frac{1}{2}, \frac{1}{4}\right)\right) + \operatorname{cone}((1,1))$$



$$\begin{cases} -2p^1 + p^2 \geq -4 \\ 2p^1 + p^2 \geq -4 \\ p^1 + p^2 \geq 0 \end{cases}$$

Изобразим S^{**}:



Задача №4

Find the conjugate cone for the exponential cone:

$$K = \{(x, y, z) \mid y > 0, ye^{x/y} \leq z\}$$

По определению, ищем вектора из $\mathbb{R}^3$ с коэффициентами (a, b, c) такие что:

$$(x, y, z)\begin{pmatrix} a \\ b \\ c \end{pmatrix} \geq 0, \text{ где}(x, y, z) : \begin{cases} ye^{x/y} \leq z \\ y > 0 \end{cases}$$

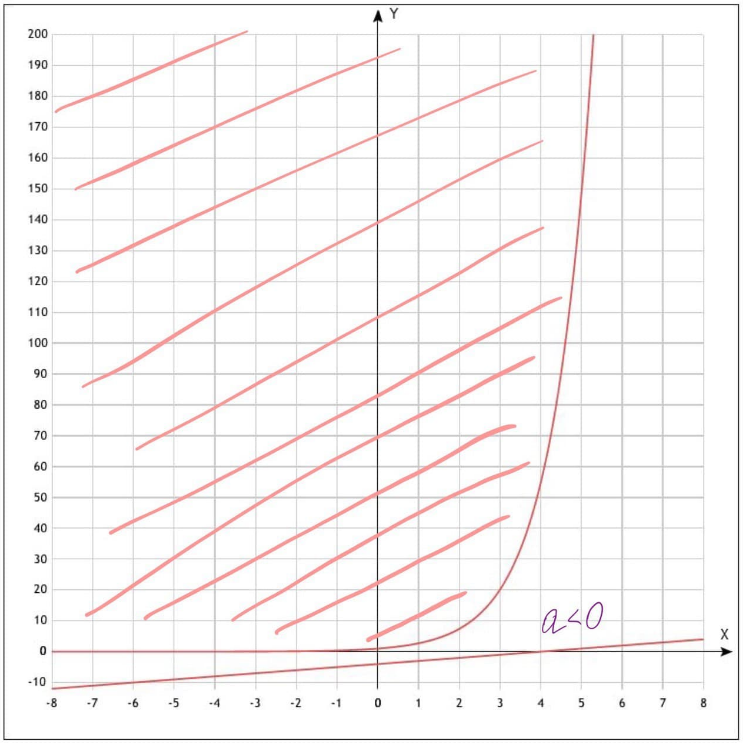
Замена: $u = \frac{x}{y}$ и $v = \frac{z}{y}$:

$$au + b + cv \geq 0, \quad \text{где } (u, v) : \begin{cases} v > 0 \\ v \geq e^u \end{cases}$$

K^* - конус, рассматриваем 3 случая: $c = -1$, $c = 0$, $c = 1$, так как любой другой можно получить домножив на положительную константу, что не изменит результатов.

$c = 1$: Выбираем a,b, чтобы в $v \geq -au - b$ лежали все значения $v \geq e^u$.

$$a < 0 : \quad b \geq a(1 - \ln(-a))$$
$$a = 0 : \quad b \geq 0$$
$$a > 0 : \text{ нет подходящих } b$$



c = 0: В области $au + b \geq 0$ лежат все точки $v \geq e^u$, но таких a,b нет ($au + b = 0$ - вертикальная прямая).

c = -1: Множество точек $v \geq e^u$ должно лежать выше прямой $au + b \geq v$, но таких a,b нет (видно из графика).

$$K^* = \left\{ \alpha \cdot (a, b, 1) \mid \alpha \geq 0, \quad \begin{matrix} \text{при } a = 0, \text{ и } b \geq 0 \\ \text{при } a < 0, \text{ и } b \geq a(1 - \ln(-a)) \end{matrix} \right\}$$

Задача №5

Prove, that B_p, B_{p_*} are inter-conjugate, i.e$(B_p)^* = B_{p_*}, (B_{p_*})^* = B_p,$ where B_p is the unit ball (w.r.t. p- norm) and p, p_* are conjugated, i.e. $p^{-1} + p_*^{-1} = 1.$
You can assume, that $p_* = \infty, p = 1$ and vice versa.

Единичный шар $B := \{x \in V :\| x \|< 1\}$, является выпуклым, симметричным по x, замкнутым, норму можно записать как:

$$\| x \|= (sup\{t \geq 0 \mid tx \in B\})^{-1}$$

Из свойств сопряженной нормы:

$$(\| x \|_p)_* = \| x \|_{p_*}, \text{ if } \frac{1}{p} + \frac{1}{p_*} = 1$$

У нас не бесконечномерное пространство => должно выполняться, $\| x \|_{**} = \| x \|, \; \forall x$

5 Conjugate function

Задача №1

Find $f^*(y)$, if $f(x) = -\dfrac{1}{x}, \;\; x \in \mathbb{R}_{++}$

По определению сопряжённой функции:

$$f^*(y) = \sup_x \left(xy - f(x)\right) = \sup_x \left(xy + \frac{1}{x}\right) = \sup_x f(x, y)$$

Область определения $f^*(y)$ (при каких y, $\sup\limits_x \left(xy + \frac{1}{x}\right)$ - ограничен):

1) если $y > 0$, то при $x \to +\infty$, $xy \to +\infty$, а $\frac{1}{x} \to 0$, следовательно $f(x, y) \to +\infty$

2) если $y < 0$, то при $x \to -\infty$, $xy \to +\infty$, а $\frac{1}{x} \to 0$, следовательно $f(x, y) \to +\infty$

3) если $y = 0$, то $\forall x \; xy = 0$. При $x \to 0$, то $\frac{1}{x} \to +\infty$, следовательно, $f(x, y) \to +\infty$

Ответ: $f^*(y) = +\infty$

Задача №2

Prove, that if $f(x_1, x_2) = g_1(x_1) + g_2(x_2),$ **then** $f^*(y_1, y_2) = g_1^*(y_1) + g_2^*(y_2)$

По определению сопряжённой функции:

$$\begin{aligned} f^*(y_1, y_2) = g_1^*(y_1) + g_2^*(y_2) &= sup_{x_1}\{x_1^T y_1 - g_1(x_1)\} + sup_{x_2}\{x_2^T y_2 - g_2(x_2)\} = \\ &= sup_{x_1, x_2}\{x_1^T y_1 - g_1(x_1) + x_2^T y_2 - g_2(x_2)\} = sup_{x_1, x_2}\{x_1^T y_1 + x_2^T y_2 - f(x_1, x_2)\} \end{aligned}$$

Задача №3

Find $f^*(y)$, if $f(x) = \log\left(\sum\limits_{i=1}^{n} e^{x_i}\right)$

По определению сопряжённой функции:

$$f^*(y) = \sup_{x \in \operatorname{dom} f} \left(\langle y, x\rangle - f(x)\right)$$

Берем градиент:

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}, \quad i = 1, \ldots, n$$

Будем использовать определение двойственной функции:

$$f^*(y) = \sum_{i=1}^{n} y_i x_i - \log\left(\sum_{i=1}^{n} e^{x_i}\right) = \sum_{i=1}^{n} \log e^{x_i y_i} - \log\left(\sum_{i=1}^{n} e^{x_i}\right) = \sum_{i=1}^{n} \log \frac{e^{x_i} e^{y_i}}{\sum\limits_{i=1}^{n} e^{x_i}} = \sum_{i=1}^{n} y_i \log y_i$$

$f^*(y_i) = y_i \log y_i$ - кросс-энтропия ($y \succ 0$ and $\mathbf{1}^T y = 1$, y - вектор вероятности: все $y_i > 0$ и $\sum\limits_{i=1}^{n} y_i = 1$).

Область определения:

1) Пусть существует компонента вектора y, такая что $y_i < 0$. Возьмём вектор $x_k = -t,$ и $x_i = 0, i \neq k, t \to \infty$. Тогда

$$y^T x - f(x) \to \infty$$

как $t - \log t \Rightarrow f^*(y)$ - неограничена.

2) Если $y \succeq 0$ но $\mathbf{1}^T y \neq 1,$ выберем $x = t\mathbf{1}$

$$y^T x - f(x) = t\mathbf{1}^T y - t - \log n$$

2.1) $\mathbf{1}^T y > 1 \Rightarrow f^*(y)$ - неограничена при $t \to \infty$

2.2) $\mathbf{1}^T y < 1 \Rightarrow f^*(y)$ - неограничена при $t \to -\infty$

$$f^*(y) = \begin{cases} \sum_{i=1}^{n} y_i \log y_i & \text{если } y \succeq 0 \text{ и } \mathbf{1}^T y = 1 \\ \infty & \text{иначе} \end{cases}$$

Задача №4

Prove, that if $f(x) = \alpha g(x)$, then $f^*(y) = \alpha g^*(y/\alpha)$

По определению сопряжённой функции:

$$\begin{aligned} f^*(y) &= \sup_x (xy - f(x)) = \sup_x (xy - \alpha g(x)) = \\ &= \sup_x \left(\alpha \left[x \cdot \frac{y}{\alpha} - g(x)\right]\right) = \alpha \sup_{x, \alpha > 0} \left(x \cdot \frac{y}{\alpha} - g(x)\right) = \\ &= \alpha \cdot g^* \left(\frac{y}{\alpha}\right) \end{aligned}$$

Задача №5

Find $f^*(Y)$, if $f(X) = -\ln\det X, X \in \mathbb{S}^n_{++}$
По определению сопряжённой функции:

$$f^*(Y) = \sup_{X\succ 0}(\operatorname{tr}(Y^TX) + \log\det X)$$

используем $\operatorname{tr}(Y^TX)$ как операцию скалярного умножения на матрицах. $f^*(Y)$ неограничена, если $Y \not\prec 0$ (Y имеет собственный вектор v с $\lambda \geq 0$ и без ограничения общности $\|v\|_2 = 1$), тогда возьмём $X = I + tvv^T$, получим:

$$\operatorname{tr}(Y^TX) + \log\det X = \operatorname{tr} Y^T + t\lambda + \log\det\left(I + tvv^T\right) = \operatorname{tr} Y^T + t\lambda + \log(1+t)$$

Неограничено при $t \to \infty$.
Если $Y \prec 0$, найдём максимум:

$$\nabla_X(\operatorname{tr}(Y^TX) + \log\det X) = Y + X^{-1} = 0$$

$X = -Y^{-1}$ (X - положительно определена, симметричная => Y симметричная). По определению:

$$f^*(Y) = \log\det(-Y)^{-1} - n$$

область определения: $\operatorname{dom} f^* = -\mathbb{S}^n_{++}$

Задача №6

Prove, that if $f(x) = inf_{u+v=x}(g(u) + h(v)),$ **then** $f^*(y) = g^*(y) + h^*(y)$
Запишем по определению:

$$f^*(y) = sup_x\{< x, y > -f(x)\}$$

Тогда можем подставить нашу f(x):

$$f^*(y) = sup_x\{< x, y > -inf(g(u) + h(v)) \mid u + v = x\}$$

Упрощаем:

$$\begin{aligned} f^*(y) &= sup_{x,u,v}\{x^Ty - g(u) - h(v) \mid u + v = x\} = sup_{u,v}\{(u+v)^Ty - g(u) - h(v)\} = \\ &= sup_u\{u^Ty - g(u)\} + sup_v\{v^Ty - h(v)\} = g^*(y) + h^*(y) \end{aligned}$$

6 Subgradient and subdifferential

Задача 1

Prove, that x_0 - is the minimum point of a convex function $f(x)$ if and only if $0 \in \partial f(x_0)$

По определению g - субградиент $f(x) : S \to R$ в точке x_0, если $\forall x \in S$

$$f(x) \geq f(x_0) + \langle g, x - x_0 \rangle$$

Множество всех субградиентов $f(x)$ в точке x_0 - субдифференциал f в x_0.

$$0 \in \partial f(x_0) \longrightarrow f(x) \geq f(x_0), \quad \forall x \in S$$

Получаем определение минимума выпуклой функции => утверждения эквивалентны.

Задача 2

Find $\partial f(x)$, if $f(x) = \text{ReLU}(x) = \max\{0, x\}$

По теореме Дубовитского-Милютина (0 и x - выпуклые функции на выпуклом множестве $S \subseteq \mathbb{R}^n$):

$$\partial_S f(x_0) = \text{conv}\left\{\bigcup_{i \in I(x_0)} \partial_S f_i(x_0)\right\}$$

где $I(x) = \{i \in [1:m] : f_i(x) = f(x)\}$

$x > 0$ - активная функция $y = x$

$x < 0$ - активная функция 0

$x = 0$ - обе функции активны, берём выпуклую оболочку

$$\partial f(x) = \begin{cases} 1, & x > 0 \\ \{0\}, & x < 0 \\ \{a \mid a = \lambda a', 0 \leq \lambda \leq 1, a' \in \partial x = 1\}, & x = 0 \end{cases}$$

Задача 3

Find $\partial f(x)$, if $f(x) = \|x\|_p$ **при** $p = 1, 2, \infty$

1. $f(x) = \|x\|_1$, по определению:

$$\|x\|_1 = |x_1| + |x_2| + \ldots + |x_n| = s_1 x_1 + s_2 x_2 + \ldots + s_n x_n$$

где $s_i = \{-1, 1\}$
Можно рассмотреть как поточечный максимум линейных функций, по теореме Дубовицкого-Милютина в каждой точке:

$$\partial f = \operatorname{conv}\left(\bigcup_{i \in I(x)} \partial g_i(x)\right)$$

где $\partial g_i(x) = \nabla g_i(x) = s_i$ (линейность):

$$\partial g(x) = \partial\left(\max\left\{s^\top x, -s^\top x\right\}\right) = \begin{cases} -s, & s^\top x < 0 \\ \operatorname{conv}(-s; s), & s^\top x = 0 \\ s, & s^\top x > 0 \end{cases}$$

1) Если j-ая координата точки отрицательна, $s_i^j = -1$
2) Если j-ая координата точки положительна, $s_i^j = 1$
3) Если j-ая координата точки равна нулю, субградиенты этих функций включены в объединение в теореме Дубовицкого - Милютина

$$\partial f(x) = \left\{g : \|g\|_\infty \leq 1, \quad g^\top x = \|x\|_1\right\}$$

2. По определению:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Функция дифференцируема во всех точках, кроме 0_n => если $x \neq 0_n$:

$$\partial f(x) = \nabla(\|x\|_2) = \frac{\mathbf{x}}{\|x\|_2}$$

Для 0_n:

$$f(0_n, p) = lim_{\alpha \to 0+} \frac{\| \alpha p \|_2}{\alpha} = \| p \|_2$$

Опорная функция единичного шара =>

$$\partial f(0_n) = \{x \mid \| x \|_2 \leqslant 1\}.$$

3. По определению:

$$f(x) = \|x\|_\infty = \max_i |x_i|$$

По теореме Дубовитского-Милютина о поточечном максимуме ($|x_i|$ - выпуклые функции):

$$\partial f(x) = \operatorname{conv}\left\{ \bigcup_{i \in I(x)} \partial |x_i| \right\}$$

где $I(x) = \{i \in [1:m] : f(x) = |x_i|\}$
Субдифференциал для модуля x:

$$\partial |x_i| = \begin{cases} \operatorname{sign}(x_i), x_i \neq 0 \\ [-1, 1], x_i = 0 \end{cases}$$

Получается,

$$\partial f(x) = \begin{cases} \operatorname{sign}(x_i), x_i \neq 0 \\ [-1, 1], x_i = 0 \end{cases}$$

где x_i - максимальный элемент в векторе

Задача 4

Find $\partial f(x)$, if $f(x) = \|Ax - b\|_1$

Пусть $g(x) = \|x\|_1$ - выпуклая функция, тогда $f(x) = \|Ax - b\|_1 = g(Ax - b)$.
Так как $g(x)$ - выпуклая функция, то можем применить свойство субдифференциального исчисления:

$$\partial f(x) = \partial(g(Ax + b))(x) = A^T \partial g(Ax + b)$$

$$\partial g(x) = \partial \|x\|_1 = \{\phi : \|\phi\|_\infty \leq 1, \phi^T x = \|x\|_1\}$$

$$\partial f(x) = A^T \cdot \{\phi : \|\phi\|_\infty \leq 1, \phi^T (Ax + b) = \|Ax + b\|_1\}$$

Задача 5

Find $\partial f(x)$, if $f(x) = e^{\|x\|}$

Рассматриваем, как сложную функцию от $g(x) = \|x\|_p$, для которой субдифференциал мы нашли в задаче 3.
Возведение в экспоненту монотонно неубывающая функция, дифференцируемая, выпуклая. Операторная форма - выпуклая функция =>

$$\partial f(x) = e^{\|x\|_p} \cdot \partial \|x\|_p$$

Задача 6

Find $\partial f(x)$, if $f(x) = \max\limits_i \{\langle a_i, x\rangle + b_i\}$, $a_i \in \mathbb{R}^n, b_i \in \mathbb{R}, i = 1, \dots, m$

Перепишем в виде: $f_i = \{\langle a_i, x\rangle + b_i\}$, $i = 1, \dots, m$. $f(x) = \max\limits_{i=\overline{1,m}} \{f_i\}$ - выпуклые функции на множестве $S = \mathbb{R}^n$
Можем использовать теорему Дубовитского-Милютина:

$$\partial_S f\,(x) = \operatorname{conv}\left\{\bigcup\nolimits_{i\in I(x)} \partial_S f_i\,(x)\right\}$$

$I(x) = \{i \in [1:m] : f_i(x) = f(x) -$ активные индексы$\}$ $f_i(x)$ - линейные функции и их субдифференциал равняется градиенту:

$$\partial_S f\,(x) = \operatorname{conv}\left\{\bigcup_{i\in I(x)} a_i\right\}$$

где $I(x) = \{i \in [1:m] : f(x) = f_i(x) = a_i x + b_i\}$

$$\partial_S f\,(x) = \left(\sum_{i\in I(x)} \lambda_i a_i : \sum_{i\in I(x)} \lambda_i = 1, \quad \lambda_i \geq 0, i \in I(x)\right)$$

где $I(x) = \{i \in [1:m] : f(x) = a_i x + b_i\}$