Домашняя работа №2

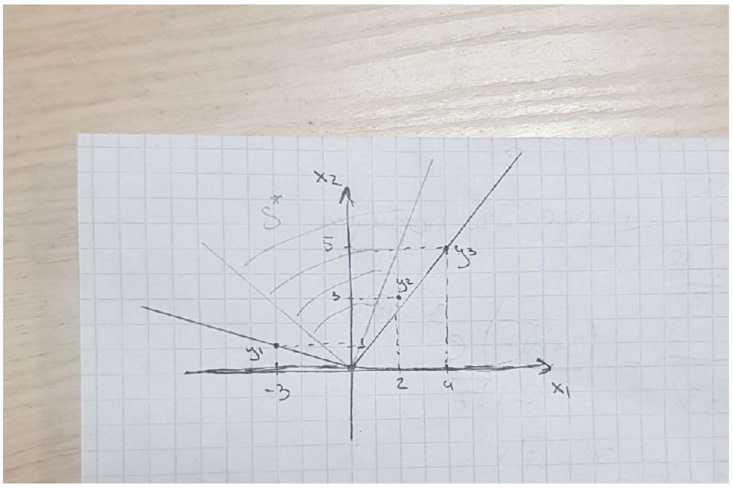
Бредихин Александр

Conjugate sets

Задача 1

Find and sketch on the plane a conjugate set to a multi-faceted cone: $S = \textbf{cone}\{(-3,1),(2,3),(4,5)\}$

Изобразим наше множество (видим, что конус строится по точкам y_1 и y_3):



Теперь использую теорему, доказанную на семинаре, что сопряжённым к множеству

$$S = \mathrm{conv}\,(x_1, \ldots, x_k) + \mathrm{cone}\,(x_{k+1}, \ldots, x_m)$$

является многогранник

$$S^* = \left\{p \in \mathbb{R}^n \mid \langle p, x_i\rangle \geq -1, i = \overline{1,k}; \langle p, x_i\rangle \geq 0, i = \overline{k+1,m}\right\}$$

В нашем случае: $S = \textbf{cone}\{y_1, y_3\}$, следовательно,

$$S^* = \{p = \begin{pmatrix} p^1 \\ p^2 \end{pmatrix} \in R^n | \quad p^T y_1 \geq 0; \quad p^T y_3 \geq 0\}$$

То есть S^* задаётся как:

$$\begin{array}{l} p^1 \cdot (-3) + p^2 \cdot 1 \geqslant 0 \\ p^1 \cdot (4) + p^2 \cdot 5 \geqslant 0 \end{array}$$

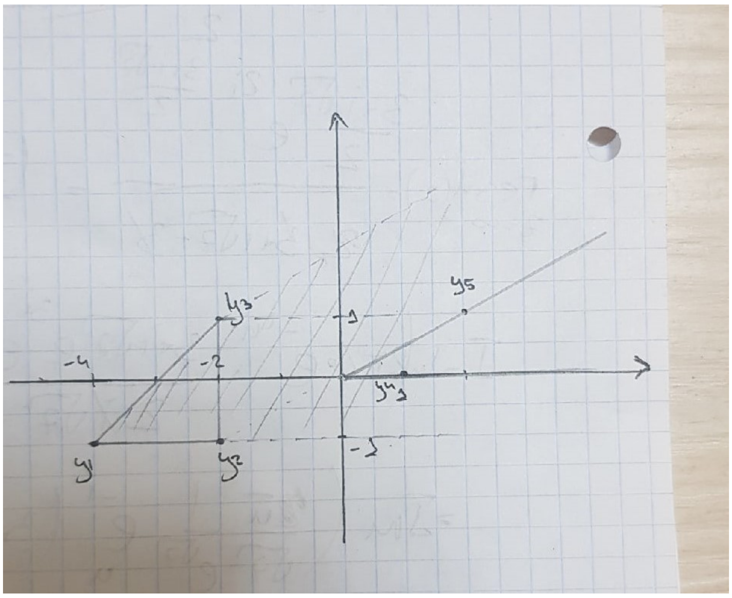
Изображено на том же рисунке.

Задача 5

Find and sketch on the plane a conjugate set to a multifaced cone:

$$S = \mathbf{conv}\,\{(-4,-1),(-2,-1),(-2,1)\} + \mathbf{cone}\,\{(1,0),(2,1)\}$$

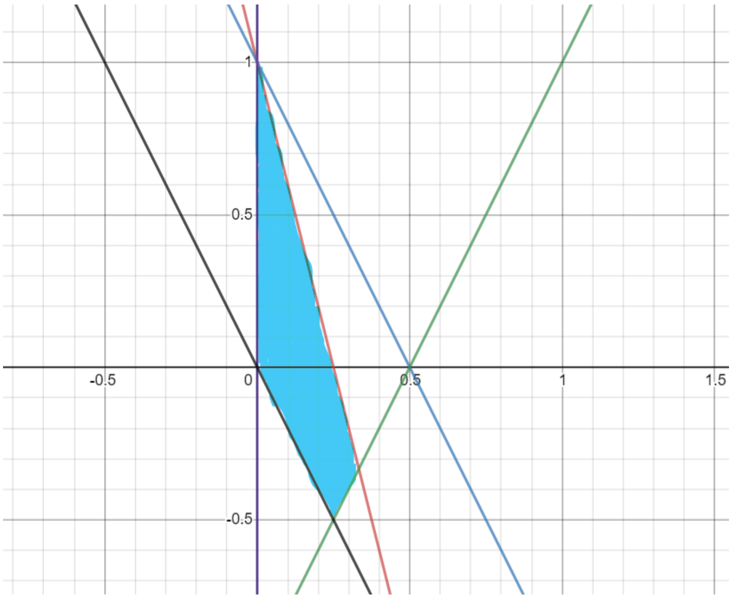
Изобразим наше множество: передвигаем конус (коническая оболочка y_4, y_5) по треугольнику, который получается как выпуклая оболочка точек y_1, y_2, y_3. В итоге он заметает область, закрашенную на рисунке



Используем ту же теорему, что и в предыдущей задаче, получаем неравенства:

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

Построим полученное множество:

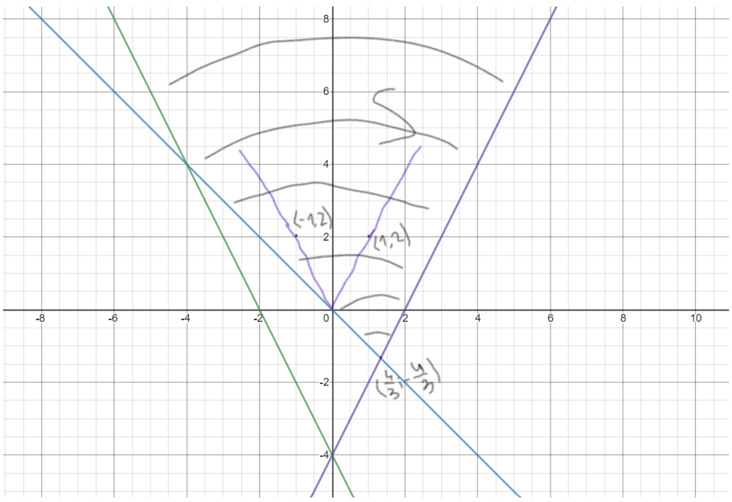


Задача 2

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

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

Изобразим наше множество S:



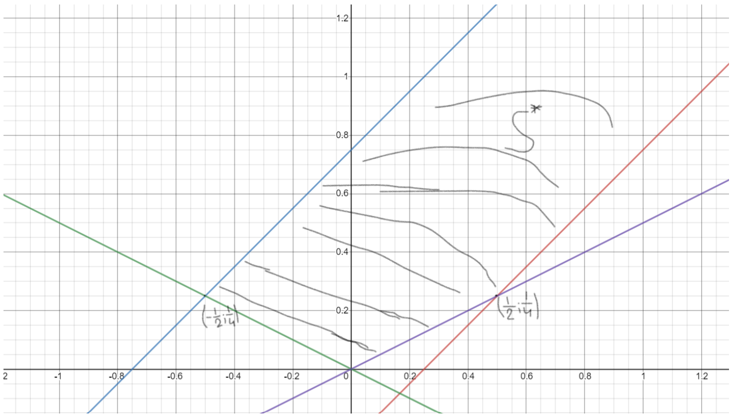
Видим, что его можно задать, как

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

Используем теорему, 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^*:



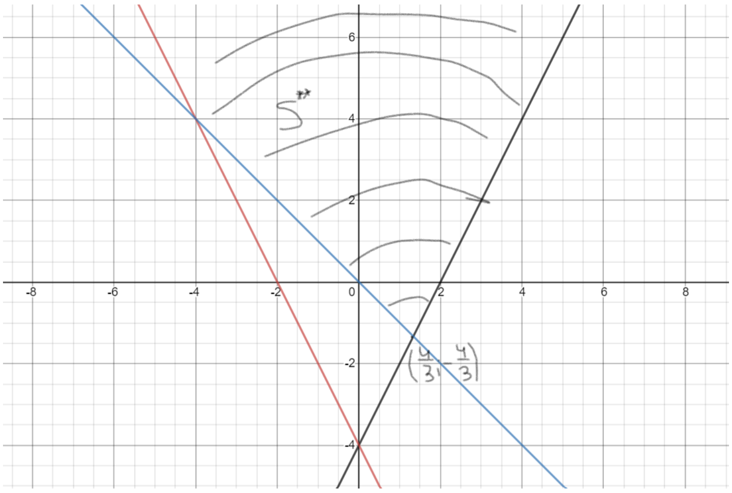
Видим, что множество S^* можно задать как:

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

Прямой $p^2 = p^1$ бегаем по треугольнику (который получается как выпуклое множество 3х точек) и заметам как раз нужную нам область. Снова можем приметить теорему и получить такие ограничения на множество S^{**}:

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

Построим его:



Видим, что $S = S^{**}$! (все ограничения на множества совпадают). Также на семинаре было свойство, что если S - замкнуто (в нашем случае это так, так как все неравенства нестрогие), выпукло и включает 0, то $S^{**} = S$, что и получили.
Следовательно, $S^{***} = S^*$, которое уже нашли выше.

Задача 3

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. Сначала, что $(\mathbb{A}_n)^* \subset \mathbb{S}_n$.
пусть матрица $B \in (\mathbb{A}_n)^*$. $\forall A$ - антисимметричной (то есть $A^T = -A$)

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

Пользуясь тем, что $A^T = -A$, получаем:

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

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

Складываем эти 2 неравенства, получаем:

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

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

Мы хотим показать, что матрица $B \in \mathbb{S}_n$, то есть, что $B = B^T$. Докажем это от противного, пусть $B \neq B^T$ тогда у матрицы $B - B^T$ на диагонале стоят 0 (так как эти элементы не меняются при транспонировании)

и есть хотя бы один ненулевой элемент. Матрица A - произвольная антисимметричная (значит на диагонали у неё тоже стоят нули). Так как матрица A произвольная, то выберем её элементы таким образом, чтобы на диагонали у матрицы $A(B - B^T)$ был хотя бы 1 элемент.
Мы так можем сделать всегда, так как: пусть без ограничения общности ненулевой элемент у матрицы $(B - B^T)$ будет в первом столбце: b_{i1} (но не на диагонале). Тогда в первой строке матрицы A выбираем $a_{1i} \neq 0$ а остальные равные нулю. Следовательно при произведении получаем, на диагонале ненулевой элемент: $e = b_{i1} \cdot a_{1i}$.
След матрицы равен сумме диагональных элементов. Б.о.о. пусть у $(B - B^T)$ только 1 ненулевой элемент, значит $tr\left(A \cdot (B - B^T)\right) = e$. Так как матрица A произвольная, то можем взять её с элементом $a_{1i} \cdot \frac{10}{e}$. Тогда $tr\left(A \cdot (B - B^T)\right) = 10 > 2$ получаем противоречие, что $B \in (\mathbb{A}_n)^*$.
Это случай, когда в матрице $(B - B^T)$ 1 ненулевой элемент, но это вообще не обязательно так, может быть 2 и в итоге мы не покажем этими рассуждениями, что след равен 0. Рассмотрим более стого:
заметим, что матрица $(B - B^T)$ - антисимметричная (простой факт из линейной алгебры), поэтому нам нужно показать, что для произвольной антисимметричной матрицы $B^* = (B - B^T)$ можно подобрать такую антисимметричную A, что $tr(AB^*) \neq 0$, это так, так как если мы просто будем брать такую же, то будем получать отрицательно определённую (на диагонале будут стоять числа одного знака и след не занулится), ну а дальше аналогично рассуждениям выше можем взять A такую, что след произведения был больше двух и мы получим противоречие

Значит, $B = B^T$, то есть симметричная: $B \in \mathbb{S}_n \to (\mathbb{A}_n)^* \subset \mathbb{S}_n$

Покажем 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 $B \in (\mathbb{A}_n)^* \to (\mathbb{A}_n)^* \supseteq \mathbb{S}_n$

Показав вложения друг в друга получаем, что $(\mathbb{A}_n)^* = \mathbb{S}_n$

Задача 4

Find the conjugate cone for the exponential cone:

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

Рассуждаем по определению: нам нужно найти такие вектора из R^3 с коэффициентами (a, b, c) такие что:

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

Сделаем замену: обозначим за $u = \frac{x}{y}$ и $v = \frac{z}{y}$, тогда получим условие:

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

Так как K^* - конус, то можно рассмотреть только 3 случая: $c = -1$, $c = 0$, $c = 1$ все остальные случаи получаются домножением на положительную константу.

Рассмотрим $c = 1$. Нам нужно подобрать такие a, b, что в области $v \geq -au - b$ лежали все значения $v \geq e^u$. Решаем это графически: строим множество точек $v \geq e^u$ и смотрим, как может распологаться прямая $v = -au - b$. Попередвигая прямую в desmos получааем:

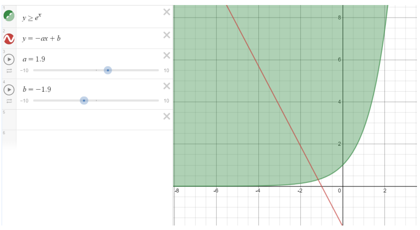


Рис. 1: $a > 0$

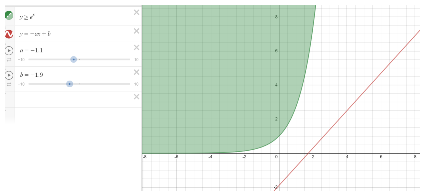


Рис. 2: $a < 0$

Делаем такие выводы: при $a > 0$ - невозможно (зелёная зона должна быть полностью выше прямой). При $a = 0$ это возможно только при $b \geq 0$. При $a < 0$ ситуация сложнее. Для понимания подходящих условий найдём точку касания кривых: $v \geq e^u$ и $v = -au - b$ для этого приравниваем значения функции и её производной, получаем связь коэффициентов b и a:

$$\begin{cases} e^u = -au - b \\ e^u = -a \end{cases} \implies b = a(1 - \ln(-a))$$

В итоге для случая $c = 1$ получаем:

$$\begin{aligned} a < 0 : &\quad b \geq a(1 - \ln(-a)) \\ a = 0 : &\quad b \geq 0 \end{aligned}$$

$$a > 0 : \text{ нет подходящих } b$$

Теперь рассмотрим $c = 0$ то есть в области $au + b \geq 0$ лежат все точки $v \geq e^u$, но таких a и b мы не найдём, так как $au + b = 0$ - вертикальная прямая поэтому понятно, что не при каких a и b не получим нужное нам условие.

Для случая $c < 0$ множество точек $v \geq e^u$ должно лежать выше прямой $au + b \geq v$, что не выполняется не при каких a и b (можно также понять графически). Если строго, то множество $v \geq e^u$ не ограничено сверху, а прямая делает это ограничение, следовательно, не при каких a и b наши требования не выполнятся.

Разобрав все случаи можем записать ответ: помним, что так как мы ищем сопряжённый конус, то все полученные тройки можно домнажать на произвольную положительную константу - α, и заметим, что если все три коэффициента будут равняться 0, то такой вектор тоже принадлежит нашему сопряжённому конусу (по определению), поэтому:

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

Задача 6

Prove that if we define the conjugate set to S as follows:

$$S^* = \{y \ \in \mathbb{R}^n \mid \langle y, x \rangle \leq 1 \ \ \forall x \in S\},$$

then unit ball with the zero point as the center is the only self conjugate set in $\mathbb{R}^n$.

Для такого определения хотим найти множество S, такое что $S = S^*$ (то есть самосопряжённое). Зафиксируем какой-то элемент $p \in S^*$: по определению: $\forall x \in S \rightarrow \langle y, x \rangle \leq 1$, то есть $\max\limits_{x \in S} \ x^T p \leq 1$.
Заметим, что если $|p| > 1$, то мы можем взять $x = p$ (так как $S = S^*$) но $p^2 > 1$, следовательно $p \notin S^*$. Получили противоречие, значит $|p| \leq 1$ (показали, что если во множестве есть хотя бы один элемент норма которого больше 1, то такое множество уже не может быть самосопряжённым по введённому определению).

Покажем, что единичный шар с центром в нуле будет самосопряжён: действительно геометрически (и из свойств скалярного произведения) понятно, что чтобы максимизировать $\max\limits_{x \in S} \ x^T p$ для фиксированного p из

единичного шара нужно брать $x = \frac{p}{\|p\|}\|x\|$. Значит, $p \le \frac{1}{\|x\|} \to p \le 1$ (так как x - произвольный из S).
Обратное включение: $\forall p \in S^* = B_{r=1}$ и $\forall x \in S = B_{r=1} \quad x^T \cdot p \le 1$ (следует из неравенства Коши-Буняковского и того, что оба элемента из единичного шара), следовательно, верно обратное включение и значит $S = S^* = B_{r=1}$

Покажем, что шары с меньшим радиусом, не будут самосопряжёнными. Рассмотрим шар радиусом $r < 1$, тогда сопряжённым к нему будет шар радиусом $\frac{1}{r}$.
Так как для фиксированного $p \in S^* \in B_{\frac{1}{r}}$ $\max\limits_{x \in S} \; x^T p$ достигается при $x \in S$ сонаправленном с p, то из определения сопряжённого множества: $x^T \cdot p \le 1 \to p \le \frac{1}{r}$, тем самым показали включение $B_{\frac{1}{r}} \subseteq S^*$
Обратное включение: $x^T p \le \frac{1}{r} \cdot r = 1 \le 1$, следовательно, $S^* \subseteq B_{\frac{1}{r}}$.
Но для $r < 1$, $B_r \ne B_{\frac{1}{r}}$, значит, не является самосопряжённым.

Покажем, что множество произвольной формы, которое лежит в шаре радиусом 1 не будет являться самосопряжённым (можно было не доказывать абзац выше, но не дошла такая идея сразу):
рассмотрим произвольное $S \subset B_{r=1}$, по определению сопряжённого множества для $p \in S^*$, $\forall x \in S \to x^T p \le 1$.
Рассматриваем $S = S^* \ne B_{r=1}$, следовательно, $\exists z \notin S$ и $z \in B_{r=1}$.
Так как z лежит в единичном шаре, то $\|z\| \le 1$, следовательно:
$\forall x \in S \subset B_{r=1} \to z^T x \le 1$, значит по определению $z \in S = S^*$.
Получаем противоречие, что z принадлежит и не принадлежит S, значит множества произвольной формы лежащие внутри единичного шара не могут быть самосопряжёнными.

Показали, что множества произвольной формы, где есть элемент с модулем больше 1, не могуть быть самомсопряжёнными. Что единичный шар - самосопряжённое множество и что все множества, которые вложены в него не будут самосопряжёнными.

Значит, для такого определения только единичный шар с центром в точке 0 является самомсопряжённым.

Задача 7

Find the conjugate set to the ellipsoid:

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

Перепишем определение конуса через матрицу, чтобы было удобнее работать:

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

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

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

Поэтому задать эллипс мы можем как:

$$S = \{x \in \mathbb{R}^n \mid \|Ax\|_2 \leq 1\}$$

покажем, что такое задание эллипса эквивалентно следующему (откуда придумал такой вид? - в Бойде в теме convex set, нашёл, что можно задавать так и это сильно помогло упростить решение):

$$E = \left\{A^{-1}u \mid \|u\|_2 \leq 1\right\}$$

Покажем вложение множеств друг в друга, чтобы показать их эквивалентность.
Из 2го в 1ое: пусть $x = A^{-1}u$, где $\|u\|_2 \leq 1$, подставим в 1ое определение, получим, что $\|AA^{-1}u\|_2 \leq 1$ - верно, следовательно, $E \subseteq S$
Из 1го во 2ое: возьмём x такой что $\|Ax\|_2 \leq 1$, обозначим $y = Ax$, получается, что $\|y\|_2 \leq 1$. Из введённого обозначения выразим $x: \ x = A^{-1}y$, где $\|y\|_2 \leq 1$, следовательно $x \in E$ и $S \subseteq E$
Получаем, что $E = S$ - один и тот же эллипсойд.

Работаем с определением конуса $E = \{A^{-1}u \mid \|u\|_2 \leq 1\}$, тогда по определению хотим найти все такие векторы $p \in E^*$, что $\forall x \in E \rightarrow \langle p, x\rangle \geq -1$, так как любой вектор $x \in E$ представим как $x = A^{-1}u$, где $\|u\|_2 \leq 1$, то нам подходят p такие что выполнено:

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

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

Применяем неравенство Коши-Буняковского, получаем:

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

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

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

Такой вид неравенства уже видели выше, оно задаёт эллипс, но теперь с другой матрицей. Заметим, что A - диагональная, поэтому при транспонировании ничего не меняется, а при взятии обратной все элементы на диагонали становятся обратными (по определению обратной матрицы, чтобы при перемножении с исходной получить единичную), то есть $A^{-T} = \mathrm{diag}(\frac{\varepsilon}{a_i})$, если переходить к привычному заданию в виде суммы, то получим:

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

что эквивалентно

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

Чтобы проверить себя можно рассмотреть частный случай эллипсоида - шар (то есть все $a_i = 1$), тогда получим сопряжённое множество - шар с радиусом $\frac{1}{R}$, что верно (рассматривали на семинаре)

Conjugate function

Задача 1

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

Запишем определение сопряжённой функции:

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

Посмотрим, на область определения $f^*(y)$, то есть при каких y, $\sup\limits_x \left(xy + \frac{1}{x}\right)$ - ограничен. Рассмотрим случаи:

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: $f^*(y) = +\infty$ (или можно сказать, что функция не определена)

Задача 2

Find $f^*(y)$, if $f(x) = -0,5 - \log x, \;\; x > 0$

Запишем определение сопряжённой функции:

$$f^*(y) = \sup_x (xy - f(x)) = \sup_x (xy + 0,5 + \log x) = \sup_x f(x, y)$$

Ищем максимум $f(x, y)$ для этого

$$\nabla_x f(x, y) = y + \frac{1}{x} = 0 \Rightarrow x^* = -\frac{1}{y}$$

Видим, что область определения $f^*(y)$: $y < 0$ (чтобы $x \in \mathbb{R}_{++}$), по определению:

$$f^*(y) = f(x^*, y) = y \cdot \left(-\frac{1}{y}\right) + \log\left(-\frac{1}{y}\right) + 0,5 = -0,5 - \log(-y)$$

Ответ: $f^*(y) = -0,5 - \log(-y)$, область определения $y < 0$ то есть:

$$f^*(y) = \begin{cases} -0,5 - \log(-y) & , \text{при } y < 0 \\ +\infty & , \text{при } y \geq 0 \end{cases}$$

Задача 3

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

По определению сопряжённой функции:

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

Взяв градиент по x от функции $\langle y, x\rangle - f(x)$ и приравняв его к нулю (для нахождения максимума (супремума)), получаем такие равенства:

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

Если подставить в определение для двойственной функции, получим:

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

Получили, что $f^*(y_i) = y_i \log y_i$ - знакомая нам кросс-энтропия!, ведь это верно, тогда и только тогда когда $y \succ 0$ and $\mathbf{1}^T y = 1$ (что соответствует, что y - вектор вероятности: все $y_i > 0$ и $\sum\limits_{i=1}^{n} y_i = 1$).

Покажем, что область определения $f^*(y)$ это как раз вектора вероятности:

- пусть существует компонента вектора y, такая что $y_i < 0$, тогда возьмём вектор x таким: $x_k = -t$, и $x_i = 0, i \neq k$ и устремим t к бесконечности. Получаем, что $y^T x - f(x)$ стремится к бесконечности, как $t - \log t$, следовательно $f^*(y)$ - неограничена.
- Если $y \succeq 0$ но $\mathbf{1}^T y \neq 1$, выберем $x = t\mathbf{1}$, получим

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

  Если $\mathbf{1}^T y > 1$, то $f^*(y)$ - неограничена при $t \to \infty$
  если $\mathbf{1}^T y < 1$, то $f^*(y)$ - неограничена при $t \to -\infty$

В итоге:

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

то есть сопряжённая функция - кросс-энтропия для y - вектора вероятности.

Задача 4

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

По определению сопряжённой функции:

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

Получили то, что нужно: $f^*(y) = \alpha g^*(y/\alpha)$

Задача 5

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

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

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

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

Видим, что неограничено при $t \to \infty$.
Если $Y \prec 0$, то найдём максимум, взяв градиент по матрице X (из первого задания знаю, что $\nabla_X \log\det X = X^{-1}$ и $\nabla_X \operatorname{tr}(Y^T X) = Y$), получаем:

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

что даёт: $X = -Y^{-1}$ (видим, что X - положительно определена, значит подходит. Ещё так как X - симметричная, то Y тоже должна быть симметричной), подставляя в определение, получаем
Ответ:

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

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

Задача 6

Prove, that if $f(x) = g(Ax)$, then $f^*(y) = g^*(A^{-\top}y)$
Доказываем по определению функции $f^*(y)$, затем сводим к определению функции g^*

$$f^*(y) = \sup_x \left(x^\top y - f(x)\right) = \sup_x p\left(x^\top y - g(Ax)\right) =$$

$$= \left[z = Ax, \quad x = A^{-1}z\right] = \sup_z \left(\left(A^{-1}z\right)^\top y - g(z)\right) =$$

$$= \sup_z \left(z^\top A^{-T} y - g(z)\right) = g^*\left(A^{-T}y\right)$$

Можем переходить от супремума по x к супремуму по z, так как $z = Ax$ - линейная замена.
Получили, что $f^*(y) = g^*(A^{-\top}y)$

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 $\forall x \in S$

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

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

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

Но это и есть определение минимума выпуклой функции! (так как $f(x)$ - выпуклая, то $\forall x \in S$, а не для окрестности x_0). Получаем, что утверждения эквивалентны.

Задача 2

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

По теореме Дубовитского-Милютина, так как 0 и x - выпуклые функции на выпуклом множестве $S \subseteq \mathbb{R}^n$

$$\partial_S f\left(x_0\right) = \operatorname{conv}\left\{\bigcup_{i \in I(x_0)} \partial_S f_i\left(x_0\right)\right\}$$
$$\text{где } I(x) = \{i \in [1:m] : f_i(x) = f(x)\}$$

Применяя эту теорему (при $x > 0$ - активная функция $y = x$, при $x < 0$ - активная функция 0, в точке 0 обе функции активны, поэтому берём выпуклую оболочку), получим:

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

Задача 3

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

Сначала напишем для $f(x) = \|x\|_1$, по определению это:

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

где $s_i = \{-1, 1\}$

Рассмотрим эту функцию, как поточечный максимум линейных по x функций, по теореме Дубовицкого-Милютина в каждой точке

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

где субдифференциал g в силу её линейности определяется как ($\partial g_i(x) = \nabla g_i(x) = s_i$):

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

и «активные» индексы в каждой точке мы выбираем как (в одномерном случае):

- Если j-ая координата точки отрицательна, $s_i^j = -1$
- Если j-ая координата точки положительна, $s_i^j = 1$
- Если j-ая координата точки равна нулю, то подходят оба варианта коэффициентов и соответствующих им функций, а значит, необходимо включать субградиенты этих функций в объединение в теореме Дубовицкого - Милютина

Чтобы красиво записать ответ, рассмотрим двумерный случай:

Тогда выпуклая оболочка аналогично правилам выше - квадрат, но если обобщать результат, то квадрат - это «сфера» в бесконечной норме, поэтому можем записать ответ как:

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

Случай $p = 2$, тогда по определению нормы

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

Это норма дифференцируемая, следовательно, субдифференциал равен просто градиенту, то есть:

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

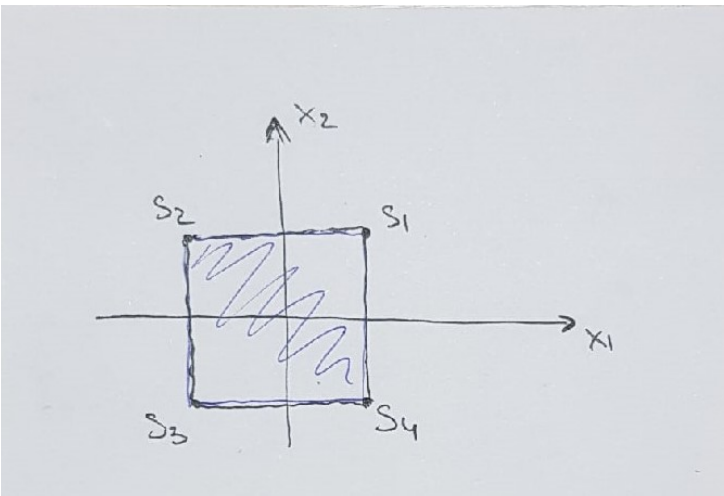


рисунок к $p = 1$

Случай $p = \infty$: по определению бесконечной нормы

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

Снова применяем теорему Дубовитского-Милютина о поточечном максимуме (так как $|x_i|$ - выпуклые функции), то

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

где $I(x) = \{i \in [1:m] : f(x) = |x_i|\}$
Субдифференциал для модуля x - уже много раз считали и знаем ответ наизусть:

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

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 (из 1го задания любая операторная норма - выпуклая функция), тогда $f(x) = \|Ax - b\|_1 = g(Ax - b)$. Так как $g(x)$ - выпуклая функция, то можем применить свойство субдифференциального исчисления:

$$\partial f(x) = \partial(g(Ax + b))(x) = A^T \partial g(Ax + b), g \text{ - выпуклая функция}$$

Из предыдущей задачи:

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

Получаем ответ:

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

Задача 5

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

Рассмотрим, как сложную функцию от $g(x) = \|x\|_p$, для которой субдифференциал мы нашли в задаче 3.
По chain rule для субдифференциала (так как возведение в экспоненту монотонно неубывающая функция и ещё выпуклая, этот факт был на семинаре в 1ом задании. Операторная форма - выпуклая функция, было в дз 1, значит его можно использовать), получаем (с учётом, что экспонента - дифференцируемая функция):

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

где $g(x) = \|x\|_p$ и $\partial g(x)$ смотреть в задаче 3

Задача 6

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

Введём $f_i = \{\langle a_i, x\rangle + b_i\}$ для $i = 1, \ldots, m$, тогда: $f(x) = \max\limits_{i=\overline{1,m}} \{f_i\}$ - выпуклые функции на множестве $S = \mathbf{R^n}$
Можем использовать теорему Дубовитского-Милютина:

$$\partial_S f(x) = \operatorname{conv}\left\{\bigcup\nolimits_{i \in I(x)} \partial_S f_i(x)\right\}$$
$$\text{где } I(x) = \{i \in [1 : m] : f_i(x) = f(x) - \text{активные индексы}\}$$

$f_i(x)$ - линейные функции и их субдифференциал равняется градиенту, то есть: $\partial_S f_i(x) = \nabla f_i(x) = a_i$, поэтому:

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

где $I(x) = \{i \in [1 : m] : f(x) = f_i(x) = a_i x + b_i\}$
Можно переписать в более красивой форме, записав что такое выпуклая оболочка точек, которые определяются активными индексами:

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

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