Домашнее задание №1. Введение/Выпуклые множества/Выпуклые функции. Векторное дифференцирование.

ВВЕДЕНИЕ (3 балла)

1. Для успешного выступления на соревнованиях спортсмен должен потратить как минимум K килокалорий. Вечером перед соревнованиями он идёт в магазин, чтобы купить продукты себе на ужин. В магазине представлено m наименований товаров, цена единицы каждого наименования равна p_i, $i = 1, ..., m$. Также известно, что i-ая единица каждого товара придаёт студенту энергию равную k_i килокалорий. Поставьте задачу определения содержимого корзины минимальной стоимости для успешного выступления. Является ли поставленная задача выпуклой и почему?

   Задача является выпуклой, если целевая функция и допустимое множество - выпуклые. Если вы считаете, что $x \in \mathbb{N}$ или $x \in \mathbb{Z}$ - задача не будет выпуклой, так как допустимое множество не является выпуклым.

2. Что такое норма вектора? Что такое эквивалентность норм? Показать эквивалентность l_1, l_2, l_∞ норм.

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

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

$$\|x\|_\infty = \max_i |x_i|$$

3. Что такое норма матрицы (линейного конечномерного оператора)? Чему равны $\mathrm{L}_1, \mathrm{L}_2, \mathrm{L}_\infty$ нормы матрицы? Что такое Фробениусова норма матрицы и какой аналог среди векторных норм векторов она имеет?

   Определение 1 (Операторная норма в $\mathbb{R}^{m\times n}$)**.** Назовем норму матрицы операторной, если она подчинена некоторой норме векторов:

$$\|\mathrm{A}\| = \sup_{x\in\mathbb{R}^n,\ \|x\|=1} \|\mathrm{A}x\| = \sup_{x\neq 0} \frac{\|\mathrm{A}x\|}{\|x\|}.$$

   Также называется подчинённой или индуцированной нормой.

$$\|\mathrm{A}\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}|$$
$$\|\mathrm{A}\|_2 = \sqrt{\lambda_{\max}\left(\mathrm{A}^\top \mathrm{A}\right)}$$
$$\|\mathrm{A}\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

Определение 2 (Стандартная фробениусова норма в $\mathbb{R}^{m\times n}$)**.**

$$\|\mathrm{A}\|_{\mathrm{F}} = \sqrt{\langle \mathrm{A}, \mathrm{A}\rangle} = \sqrt{\mathrm{Tr}\left(\mathrm{A}^\top \mathrm{A}\right)} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}.$$

Стандартная фробениусова норма также называется нормой Гильберта-Шмидта.

ВЫПУКЛЫЕ/КОНИЧЕСКИЕ МНОЖЕСТВА (2 балла)

1. Доказать по определению, что гиперболическое множество

$$\mathrm{Q} = \left\{ x \in \mathbb{R}^n_+ \;\middle|\; \prod_{i=1}^{n} x_i \geq 1 \right\} \text{ – выпуклое.}$$

- Мы знаем, что $\ln(x)$ – вогнутая функция $\left(f''(x) = -\frac{1}{x^2} < 0\right)$.
  Тогда из определения вогнутой функции получаем следующее неравенство

$$\ln(\alpha x + (1-\alpha)y) \geq \alpha \ln x + (1-\alpha)\ln y.$$

  Так как логарифм непрерывная, монотонно возрастающая функция, то

$$\alpha x + (1-\alpha) y \geq x^\alpha y^{1-\alpha},$$

  полученным соотношением мы воспользуемся далее.
- Пусть $x, y \in \mathrm{Q}$, тогда

$$\prod_{i=1}^{n} x_i \geq 1, \; \prod_{i=1}^{n} y_i \geq 1,$$

  необходимо показать, что $z = \alpha x + (1-\alpha) y \in \mathrm{Q}, \;\alpha \in [0,1]$:

$$\prod_{i=1}^{n} z_i = \prod_{i=1}^{n} (\alpha x_i + (1-\alpha) y_i) \geq \prod_{i=1}^{n} x_i^\alpha y_i^{1-\alpha} = \left(\prod_{i=1}^{n} x_i\right)^\alpha \left(\prod_{i=1}^{n} y_i\right)^{1-\alpha} \geq 1.$$

2. Пусть $\mathrm{P} \in \mathbb{S}^n_{++}$, $c \in \mathbb{R}^n$. Покажите, что множество

$$\left\{x \in \mathbb{R}^n \mid \langle \mathrm{P}x, x\rangle \leq \langle c, x\rangle^2,\ \langle c, x\rangle \geq 0\right\} \text{ – выпуклое.}$$

$$\langle \mathrm{P}x, x\rangle = x^\top \mathrm{P}^\top x = x^\top \mathrm{P}x, \quad \langle c, x\rangle^2 = x^\top cc^\top x$$

Докажем по определению, для этого мы хотим показать, что

$$(\alpha x + (1-\alpha)y)^\top \mathrm{P}(\alpha x + (1-\alpha)y) \leq (\alpha x + (1-\alpha)y)^\top cc^\top(\alpha x + (1-\alpha)y).$$

После раскрытия скобок получаем следующее выражение

$$\alpha^2 x^\top \mathrm{P}x + (1-\alpha)^2 y^\top \mathrm{P}y + 2\alpha(1-\alpha)x^\top \mathrm{P}y \leq \alpha^2 x^\top cc^\top x + (1-\alpha)^2 y^\top cc^\top y + 2\alpha(1-\alpha)x^\top cc^\top y.$$

Первое и второе слагаемые левой части не превосходят соответственно первого и второго слагаемых правой части, т.к. это в точности условия для x, y, задающие множество. Нам достаточно доказать неравенство между третьими слагаемыми. Так как $2\alpha(1-\alpha) \geq 0$, то достаточно показать, что $x^\top \mathrm{P}y \leq x^\top cc^\top y$,.

В силу того, что $\mathrm{P} \in \mathbb{S}^n_{++}$, то у нее существует положительно определенный квадратный корень $\mathrm{P}^{\frac{1}{2}}$, то есть $\mathrm{P} = \mathrm{P}^{\frac{1}{2}}\mathrm{P}^{\frac{1}{2}}$.

$$\langle c, x\rangle^2 \geq x^\top \mathrm{P}x = x^\top \mathrm{P}^{\frac{1}{2}}\mathrm{P}^{\frac{1}{2}}x = \|\mathrm{P}^{\frac{1}{2}}x\|_2^2$$

$$x^\top cc^\top y = \langle c, x\rangle\langle c, y\rangle \geq \|\mathrm{P}^{\frac{1}{2}}x\|\|\mathrm{P}^{\frac{1}{2}}y\| \geq \left\langle \mathrm{P}^{\frac{1}{2}}x, \mathrm{P}^{\frac{1}{2}}y\right\rangle = x^\top \mathrm{P}^{\frac{1}{2}}\mathrm{P}^{\frac{1}{2}}y = x^\top \mathrm{P}y$$

То есть мы получили неравенство $x^\top cc^\top y \geq x^\top \mathrm{P}y$, а оно влечет требуемое неравенство, из которого следует выпуклость исходного множества.

ВЫПУКЛЫЕ ФУНКЦИИ (2 балла)

1. Докажите выпуклость следующих функций:

   (a) $f(x) = \left(\sum\limits_{i=1}^{n} e^{x_i}\right)$;

   Выпуклая как сумма выпуклых функций.

   (b) $f(x) = \frac{\|\mathrm{A}x - b\|^2}{1 - x^\top x}$, $\mathrm{X} = \{x \mid \|x\|^2 \leq 1\}$;

   Сводится к пункту (c).

   (c) $F(x) = \frac{f^2(x)}{g(x)}$ при условии, что $f(x)$ является выпуклой и принимает только неотрицательные значения, $g(x)$ вогнута и принимает только положительные значения;

   Достаточно проверить на положительную полуопределенность матрицу Гессе $\nabla^2 F(x)$ (такое решение засчитывалось за полный балл).

   (d) $f(x,t) = -\log(t^2 - x^T x)$, $\mathrm{E} = \{(x,t)\} \in \mathbb{R}^{n\times 1} \mid \|x\|_2 < t\}$;

   Перепишем функцию $f(x,t) = -\log t - \log(t - \frac{1}{t}x^\top x)$

Замечание 1. Пусть $h(x)$ - вогнутая функция, $P(y)$ - выпуклая и невозрастающая. Тогда $f(x) = P(h(x))$ - выпуклая функция.

- $-\log t$ – выпуклая и невозрастающая функция
- покажем, что $h(x,t) = t - \frac{1}{t}x^\top x$ - вогнутая функция, для этого посчитаем матрицу Гессе:

$$\nabla^2 h(x,t) = \begin{pmatrix} -\frac{2}{t}\mathrm{I_n} & \frac{2x}{t^2} \\ \frac{2x^\top}{t^2} & -\frac{2}{t^3}(x^\top x) \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)},$$

  где I_n — единичная матрица $n \times n$.
- матрица вторых производных $\nabla^2 h \preceq 0$ отрицательно полуопределена, а значит функция $h(x,t)$ - вогнутая, следовательно функция $-\log(t - \frac{1}{t}x^\top x)$ является выпуклой по замечанию 1
- сумма выпуклых функций $-\log t - \log(t - \frac{1}{t}x^\top x)$ – выпуклая функция

Так же можно было проверить на положительную полуопределенность матрицу вторых производных всей функции $f(x,t)$.

(e) $f(x) = \sum\limits_{i=1}^n w_i \ln\left(1 + \exp(a_i^\top x)\right) + \frac{\mu}{2}\|x\|_2^2, \quad \mu, w_1, \ldots, w_n > 0, \; a_i \in \mathbb{R}^n;$

Достаточно проверить на положительную полуопределенность матрицу Гессе $\nabla^2 f(x)$. (см. тест на векторное дифференцирование пункт 3)

(f) $f(x) = \ln \sum\limits_{i=1}^n \exp\left(\max\{0, x_i\}^2\right).$

Докажем выпуклость функции $g(x) = \ln\left(\sum\limits_{i=1}^n \exp x_i\right)$ по определению, пользуясь неравенством Гёльдера.

Замечание 2. Неравенство Гёльдера: пусть $p > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, тогда

$$\sum_{i=1}^n |x_i y_i| \le \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n |y_i|^q\right)^{\frac{1}{q}}.$$

$$\ln\left(\sum_{i=1}^n \exp(\alpha x_i + (1-\alpha)y_i)\right) = \ln\left(\sum_{i=1}^n \exp(\alpha x_i)\exp((1-\alpha)y_i)\right) \le$$

$$\le \text{Неравенство Гёльдера} \le$$

$$\le \ln\left(\left(\sum_{i=1}^n \exp(x_i)\right)^{\alpha} \left(\sum_{i=1}^n \exp(y_i)\right)^{(1-\alpha)}\right) = \alpha \ln\left(\sum_{i=1}^n \exp(x_i)\right) + (1-\alpha)\ln\left(\sum_{i=1}^n \exp(y_i)\right)$$

- с помощью операций сохраняющих выпуклость получаем выпуклость $h(x) = \max\{0, x_i\}^2$
- замечаем, что функция $g(x)$ – возрастающая по каждой переменной
- применяем свойство выпуклости композиции выпуклой и возрастающей функции и выпуклой

ТЕСТ/ВЫПУКЛЫЕ ФУНКЦИИ (2 балла)

1. Какое максимальное количество точек минимума может быть у выпуклой функции?
   (a) Одна
   (b) Две
   (c) Счётное множество
   (d) Несчётное множество – $f(x) = \text{const}$
2. Какие из следующих функций выпуклы?
   (a) $f(x) = \sum_{i=1}^{r} x_{[i]}$, где $x_{[1]} \geq x_{[1]} \geq \ldots \geq x_{[n]}$ — упорядоченные по убыванию элементы вектора x
   (b) $f(x) = x^5$
   (c) $f(x) = \sum_{i=1}^{n} x_i \log x_i$, где $x_i \geq 0$ и $f(0) = 0$ по непрерывности
   (d) $f(A) = x^\top Ax$, где $A \in \mathbb{R}^{n\times n}$, $x \in \mathbb{R}^n$ — фиксированный вектор
   (e) $\|x\|_{1/3} = (\sum_{i=1}^{n} |x_i|^{1/3})^3$ – не является выпуклой (многие писали, что это норма, но это не так: для данной функции не выполняется неравенство треугольника)
3. Чем является касательная к графику выпуклой функции в каждой точке области определения?
   (a) Локальной оценкой сверху
   (b) Глобальной оценкой снизу
   (c) Локальной оценкой снизу – тоже верное утверждение
   (d) Глобальной оценкой сверху
4. На каком множестве выпукла функция $f = x_2^3 + x_2(x_1^2 + x_3^2)$?
   (a) $\{(x_1, x_2, x_3) \mid x_1 \geq 0,\ x_2 \geq 0,\ x_3 \geq 0\}$
   (b) $\{(x_1, x_2, x_3) \mid x_2 \leq 0,\ 3x_2^2 - x_1^2 \geq 0\}$
   (c) $\{(x_1, x_2, x_3) \mid x_2 \geq 0,\ 3x_2^2 - x_1^2 \geq x_3^2\}$ – необходимо проверить на положительную полуопределенность матрицу вторых производных исходной функции
   (d) $\{(x_1, x_2, x_3) \mid x_2 \geq 0,\ 6x_2^2 - x_1^2 \geq 0\}$
5. Может ли область определения выпуклой функции быть невыпуклым множеством?
   (a) Да
   (b) Нет
6. Отметьте все композиции $f(x) = h(g(x))$, такие что $f(x)$ — выпуклая функция.
   (a) $h(x) = \log x$, $g(x)$ – выпуклая функция
   (b) $h(x) = e^x$, $g(x)$ – вогнутая функция

(c) $g(x) = Ax + b$ для некоторой матрицы A и вектора b, h – выпуклая функция

(d) $h(x) = |x - 4|$, $g(x) = |x|$

7. Чему равна константа сильной выпуклости у функции $f(x) = \frac{1}{2}\|x\|_2^2$?

(a) 1

(b) 2

(c) 3

(d) 10

ТЕСТ/ВЕКТОРНОЕ ДИФФЕРЕНЦИРОВАНИЕ (2 балла)

1. Посчитайте $df(x)$ и $\nabla f(x)$ для функции $f(x) = \log(x^\top \mathrm{A}x)$ и выберите правильный вариант ответа.

(a) $\nabla f(x) = \frac{2\left(\mathrm{A}+\mathrm{A}^\top\right)x}{x^\top \mathrm{A}x}$, $df(x) = \frac{x^\top 2\left(\mathrm{A}+\mathrm{A}^\top\right)dx}{x^\top \mathrm{A}x}$

(b) $\nabla f(x) = \frac{\left(\mathrm{A}+\mathrm{A}^\top\right)x}{x^\top \mathrm{A}x}$, $df(x) = \frac{x^\top\left(\mathrm{A}+\mathrm{A}^\top\right)dx}{x^\top \mathrm{A}x}$

(c) $\nabla f(x) = \frac{x^\top\left(\mathrm{A}+\mathrm{A}^\top\right)dx}{x^\top \mathrm{A}x}$, $df(x) = \frac{\left(\mathrm{A}+\mathrm{A}^\top\right)x}{x^\top \mathrm{A}x}$

(d) $\nabla f(x) = \frac{x^\top 2\left(\mathrm{A}+\mathrm{A}^\top\right)dx}{x^\top \mathrm{A}x}$, $df(x) = \frac{2\left(\mathrm{A}+\mathrm{A}^\top\right)x}{x^\top \mathrm{A}x}$

2. Посчитайте $df(x)$, $d^2f(x)$ и $\nabla f(x)$, $\nabla^2 f(x)$ для функции $f(x) = \frac{1}{p}\|x\|_2^p$, $p > 1$ и выберите все правильные результаты.

(a) $\nabla f(x) = \|x\|_2^{p-1}x$

(b) $df(x) = \|x\|_2^{p-2}x^\top dx$

(c) $\nabla^2 f(x) = (p-2)\|x\|_2^{p-4}x^\top x + \|x\|_2^{p-2}$

(d) $\nabla^2 f(x) = (p-1)\|x\|_2^{p-3}xx^\top + \|x\|_2^{p-1}I$

(e) $d^2 f(x) = dx^\top\left((p-2)\|x\|_2^{p-4}xx^\top + \|x\|_2^{p-2}I\right)dx$

(f) $d^2 f(x) = dx^\top\left((p-1)\|x\|_2^{p-3}xx^\top + \|x\|_2^{p-1}I\right)dx$

3. Верно ли, что у функции $f(x) = \frac{1}{n}\sum\limits_{i=1}^{n}\log\left(1 + \exp(a_i^\top x)\right) + \frac{\mu}{2}\|x\|_2^2$, $a_i \in \mathbb{R}^n$, $\mu > 0$, гессиан $\nabla^2 f(x) = \frac{1}{n}\sum\limits_{i=1}^{n}\left(-\frac{\exp(2a_i^\top x)}{(1+\exp(a_i^\top x))^2}a_i a_i^\top + \frac{\exp(a_i^\top x)}{1+\exp(a_i^\top x)}a_i a_i^\top\right) + \mu \mathrm{I}$?

(a) Да

(b) Нет – в исходном тексте задания не хватало двойки, которая выделена красным вверху