Контрольная работа. Демо вариант. Решения

1. **(Матрично-векторное дифференцирование)**

   Найдите градиент:

$$f(x) = \left\| x^\top B \cdot \left\| b^\top x \right\|_2^2 \cdot Ax \right\|_F^2,$$

   где $b, x \in \mathbb{R}^d$ и $A, B \in \mathbb{R}^{d \times d}$.

   Решение. □

   Для начала поймем, что $b^\top x \in \mathbb{R}, x^\top BAx \in \mathbb{R}$, то есть, под нормами стоят скаляры. Тогда функцию можно переписать:

$$f(x) = (b^\top x)^4 (x^\top BAx)^2.$$

   Теперь, когда функция выглядит проще, можно перейти к вычислению дифференциала:

$$\begin{aligned} df(x) &= d\left((b^\top x)^4 (x^\top BAx)^2\right) \\ &= 4(b^\top x)^3 b^\top dx (x^\top BAx)^2 + (b^\top x)^4 \cdot 2(x^\top BAx)\left(dx^\top BAx + x^\top BAdx\right) \\ &= 4(b^\top x)^3 b^\top dx (x^\top BAx)^2 + (b^\top x)^4 \cdot 2(x^\top BAx) x^\top (A^\top B^\top + BA) dx \\ &= \langle 4(b^\top x)^3 (x^\top BAx)^2 b + 2(b^\top x)^4 (x^\top BAx)(A^\top B^\top + BA)x, dx \rangle. \end{aligned}$$

   Наконец, получаем:

$$\nabla f(x) = 4(b^\top x)^3 (x^\top BAx)^2 b + 2(b^\top x)^4 (x^\top BAx)(A^\top B^\top + BA)x.$$

   ■

   Ответ:

$$\nabla f(x) = 4(b^\top x)^3 (x^\top BAx)^2 b + 2(b^\top x)^4 (x^\top BAx)(A^\top B^\top + BA)x.$$

2. **(Выпуклость множеств)**

   Проверьте множество S на выпуклость:

$$S = \{X \in \mathbb{S}^{d \times d} \mid X \succ 0,\ \mathrm{tr}(X^{-1}) \leq 1\}.$$

   Решение. □

   Докажем два утверждения. Первое, что если $f(x)$ выпуклая функция, то множество

$$S^* = \{x \mid f(x) \leq 1\}$$

   выпукло. Нам нужно показать, что $\forall x_1, x_2 \in S^*$, то $\lambda \in [0, 1]$ выполняется:

$$\lambda x_1 + (1 - \lambda) x_2 \in S^*.$$

   Иными словами, нам нужно показать, что для любого $x_1, x_2 \in S^*$ выполняется:

$$f(\lambda x_1 + (1 - \lambda) x_2) \leq 1.$$

   Поскольку $x_1, x_2 \in S^*$, то по определению множества S^* имеем:

$$f(x_1) \leq 1 \text{ и } f(x_2) \leq 1.$$

   Так как $f(x)$ выпуклая, то по определению выпуклости функции для любых x_1 и x_2, а также для $\lambda \in [0, 1]$ выполняется:

$$f(\lambda x_1 + (1 - \lambda) x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2).$$

Подставляя $f(x_1) \leq 1$ и $f(x_2) \leq 1$, получаем:

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda \cdot 1 + (1-\lambda) \cdot 1 = 1.$$

Таким образом $\lambda x_1 + (1-\lambda)x_2 \in S^*$.

Теперь докажем, что для положительно определенной симметричной матрицы X: $\mathrm{tr} X^{-1}$ – выпуклая функция.

Пусть $C(t) = A + tB$, где $A, B \in S$. Достаточно показать, что

$$\frac{d^2}{dt^2} \mathrm{tr}\left(C(t)^{-1}\right)\Big|_{t=0} \geq 0.$$

При этом

$$C(t)^{-1} = (A(I + tA^{-1}B))^{-1} = A^{-1} - tA^{-1}BA^{-1} + t^2A^{-1}BA^{-1}BA^{-1} + \dots$$

Тогда

$$\frac{d^2}{dt^2} \mathrm{tr}\left(C(t)^{-1}\right)\Big|_{t=0} = 2\mathrm{tr}\left(A^{-1}BA^{-1}BA^{-1}\right).$$

Но $A^{-1}BA^{-1}BA^{-1} = A^{-1}BA^{-1}\left(A^{-1}B\right)^{\top}$ и $A^{-1}B$ – положительно определенная, поэтому $A^{-1}BA^{-1}BA^{-1}$ неотрицательно определена, т.е. $\mathrm{tr}\left(A^{-1}BA^{-1}BA^{-1}\right) \geq 0$.

■

Ответ: Да, выпукло.

3. **(Выпуклость функций)**

Докажите, что функция $f(x)$ строго выпуклая:

$$f(x) = -\sum_{i=1}^{d} \log(x_i),$$

где $x \in \mathbb{R}^d_{++}$.

Решение. □

Найдем первый и второй дифференциал функции $f(x)$.

$$df(x) = -\sum_{i=1}^{d} \frac{1}{x_i} dx_i,$$

$$d^2f(x) = \sum_{i=1}^{d} \frac{1}{x_i^2} dx_i^2 = \left\langle \mathrm{diag}\left(\frac{1}{x_i^2}\right) dx, dx \right\rangle.$$

Гессиан строго положительно определен:

$$\nabla^2 f(x) = \mathrm{diag}\left(\frac{1}{x_i^2}\right) \succ 0.$$

Тогда по дифференциальному критерию строгой выпуклости второго порядка функция f строго выпукла.

■

4. **(Субградиент)**

Найдите субградиент функции f, заданной в целых числах по формуле:

$$f(k) = \begin{cases} -\frac{1}{|k|}, & k \not\vdots\, 2, \\ 0, & k \vdots\, 2 \end{cases}$$

и продолженной в остальных точках до кусочно-линейной.

Решение. □

Зафиксируем точку $x_0 \in \mathbb{R}$, в которой вычисляем субградиент. Запишем определение:

$$f(x) \geq f(x_0) + g \cdot (x - x_0) \; \forall x \in \mathbb{R}.$$

Покажем, что $g \in \{0\}$. Пусть это не так. Тогда $\exists k \in \mathbb{Z} : f(k) = 0, |k - x_0| > \frac{-f(x_0)}{|g|}, \text{ sign}(k - x_0) = \text{sign}(g)$.

На этом k нарушается неравенство:

$$0 = f(k) \geq f(x_0) + g(k - x_0) = f(x_0) + |g(k - x_0)| > f(x_0) + |g|\frac{-f(x_0)}{|g|} = 0.$$

Получили противоречие, значит, $g \in \{0\}$.

Теперь поймем, при каких x_0 будет $g = 0$, а при каких $g \in \varnothing$.

Пусть $g = 0$. Согласно определению субградиента

$$f(x) \geq f(x_0) + 0 \cdot (x - x_0) = f(x_0) \; \forall x \in \mathbb{R}.$$

То есть, x_0 – точка глобального минимума. Таких точек две: ± 1. Для остальных нарушается определение субградиента при подстановке $x = 1$, а, учитывая, что мы до этого сузили множество, в котором может лежать g до $\{0\}$, для остальных точек субградиент не определен.

■

Ответ: $g = 0$ в точках ± 1, $g \in \varnothing$ в точках $\mathbb{R} \setminus \{\pm 1\}$.

5. **(Сопряженные множества)**

Пусть $A = \{x \in \mathbb{R}^d \mid \|x\|_2 \leq 3\}$, а $B = \{x \in \mathbb{R}^d \mid \|x - \mathbf{1}\|_1 \leq 2\}$, где $\mathbf{1}$ – вектор единиц размера d. Найдите сопряженные к A и B, а так же найдите их пересечение и объединение.

Решение. □

Оба множества являются шарами: A – шар по 2-норме радиуса 3 с центром в нуле, B – шар по 1-норме радиуса 2 с центром в $\mathbf{1}$.

Вспомним результат из пособия:

$$(B_{\|\cdot\|}(\mathbf{0}, r))^* = B_{\|\cdot\|_*}(\mathbf{0}, 1/r).$$

Применим его к множеству A:

$$A^* = (B_{\|\cdot\|_2}(\mathbf{0}, 3))^* = B_{\|\cdot\|_2}\left(\mathbf{0}, \frac{1}{3}\right).$$

Здесь был использован следующий факт: $(\|\cdot\|_2)_* = \|\cdot\|_2$.

Перейдем к множеству B. По определению сопряженное к нему задается как

$$B^* = \{y \in \mathbb{R}^d \mid \langle y, x \rangle \geq -1 \; \forall x \in B\}.$$

Обратимся к неравенству Гельдера при $p = 1,\ q = \infty$:

$$\langle y, x\rangle \geq -\|x\|_1 \|y\|_\infty.$$

Воспользуемся им для векторов y и $x - 1$:

$$\langle y, x - \mathbf{1}\rangle \geq -\|x - \mathbf{1}\|_1 \|y\|_\infty \geq -2\|y\|_\infty.$$

В последнем переходе использовали $\|x - \mathbf{1}\|_1 \leq 2$ для элементов B.

Преобразуем неравенство и получим:

$$\langle y, x\rangle \geq \langle y, \mathbf{1}\rangle - 2\|y\|_\infty.$$

Если теперь потребовать

$$\langle y, \mathbf{1}\rangle - 2\|y\|_\infty \geq -1,$$

то получим желаемое ограничение на $y \in B^*$, так как для неравенств, которые мы использовали, можно подобрать $x \in B$, чтобы они превращались в равенство:

$$B^* = \{y \in \mathbb{R}^d \mid \langle y, \mathbf{1}\rangle - 2\|y\|_\infty \geq -1\}.$$

Теперь найдем их пересечение и объединение.

$$A^* \cap B^* = B_{\|\cdot\|_2}\left(\mathbf{0}, \frac{1}{3}\right) \cap \{y \in \mathbb{R}^d \mid \langle y, \mathbf{1}\rangle - 2\|y||_\infty \geq -1\}.$$

Для объединения ничего интересного придумать не получится:

$$A^* \cup B^* = B_{\|\cdot\|_2}\left(\mathbf{0}, \frac{1}{3}\right) \cup \{y \in \mathbb{R}^d \mid \langle y, \mathbf{1}\rangle - 2\|y||_\infty \geq -1\}.$$

■

Ответ: $A^* = B_{\|\cdot\|_2}\left(\mathbf{0}, \frac{1}{3}\right)$, $B^* = \{y \in \mathbb{R}^d \mid \langle y, 1\rangle - 2\|y\|_\infty \geq -1\}$, $A^* \cap B^* = B_{\|\cdot\|_2}\left(\mathbf{0}, \frac{1}{3}\right) \cap \{y \in \mathbb{R}^d \mid \langle y, 1\rangle - 2\|y\|_\infty \geq -1\}$, $A^* \cup B^* = B_{\|\cdot\|_2}\left(\mathbf{0}, \frac{1}{3}\right) \cup \{y \in \mathbb{R}^d \mid \langle y, \mathbf{1}\rangle - 2\|y\|_\infty \geq -1\}$.

6. **(Сопряженные функции)**

Найдите сопряженную функцию для функции (в зависимости от p):

$$f(x) = |x|^p,\ p \leq 1,$$

где $x \in \mathbb{R}$.

Решение. □

Определение сопряженной к f функции в одномерном случае:

$$f^*(y) = \sup_{x \in \mathbb{R}} \{xy - f(x)\}.$$

Подставим нашу функцию.

$$f^*(y) = \sup_{x \in \mathbb{R}} \{xy - |x|^p\}$$

Выделим несколько случаев в зависимости от значения p:

- $p = 1$
  Под супремумом получаем функцию
  $$xy - |x| = xy - \max\{x, -x\} = \min\{x(y-1), x(y+1)\}.$$
  Рассматривая ее как функцию от x при фиксированном y, понимаем, что также выделяются несколько случаев.
  При $y \in [-1, 1]$ функция выглядит как галка с ветвями направленными вниз, либо одна из ветвей параллельна оси абсцисс. В любом случае, супремум достигается в вершине этой галки, то есть при условии
  $$x(y-1) = x(y+1) \iff x = 0.$$
  Подставляем $x = 0$ в определение сопряженной функции
  $$f^*(y) = \min\{x(y-1), x(y+1)\}|_{x=0} = 0,\ y \in [-1, 1],\ p = 1.$$
  Теперь пусть $y \in \mathbb{R} \setminus [-1, 1]$. Коэффициенты при x одного знака, функция под супремумом монотонная ломаная с 1 изломом. Либо на $+\infty$, либо на $-\infty$ функция будет стремиться к $+\infty$. Тогда
  $$f^*(y) = \sup_{x \in \mathbb{R}} \min\{x(y-1), x(y+1)\} = +\infty,\ y \in \mathbb{R} \setminus [-1, 1],\ p = 1.$$
- $p = 0$
  $$f^*(y) = \sup_{x \in \mathbb{R}}\{xy - |x|^0\} = \sup_{x \in \mathbb{R}}\{xy - 1\} = \begin{cases} -1,\ y = 0, \\ +\infty,\ y \neq 0 \end{cases}$$
- $p \in (-\infty, 0) \cup (0, 1)$
  При $y \neq 0$ можно устремить x либо к $+\infty$, если $y > 0$, либо к $-\infty$, если $y < 0$, и получить $+\infty$ в качестве супремума. Это так, потому что на бесконечности $xy - |x|^p \sim xy$ при $p < 1$.
  При $y = 0$ имеем
  $$\sup_{x \in \mathbb{R}}\{xy - |x|^p\} = \sup_{x \in \mathbb{R}}\{-|x|^p\} = 0,$$
  устремляя x либо к 0, либо к бесконечности.

■

Ответ:

$$f^*(y) = \begin{cases} 0,\ y \in [-1, 1], \\ +\infty,\ y \in \mathbb{R} \setminus [-1, 1]. \end{cases} \quad \text{при } p = 1,$$
$$f^*(y) = \begin{cases} -1,\ y = 0, \\ +\infty,\ y \neq 0. \end{cases} \quad \text{при } p = 0,$$
$$f^*(y) = \begin{cases} 0,\ y = 0, \\ +\infty,\ y \neq 0. \end{cases} \quad \text{при } p \neq 0, 1.$$

7. **(Двойственность)**

Составьте двойственную задачу:
$$\begin{aligned} \min_{X \in \mathbb{S}^d_{++}} \ & \log \det X^{-1} \\ \text{s.t. } & a_i^\top X a_i \leq 1,\ i \in \overline{1, m}, \end{aligned}$$

где $a_i \in \mathbb{R}^d$.

Решение. $\square$

Лагранжиан:

$$\mathcal{L}(X, \lambda) = \log \det X^{-1} + \sum_{i=1}^{m} \lambda_i (a_i^\top X a_i - 1).$$

Двойственная функция:

$$g(\lambda) = \inf_{X \in \mathbb{S}_{++}^d} \mathcal{L}(X, \lambda) = \inf_{X \in \mathbb{S}_{++}^d} \left(\log \det X^{-1} + \sum_{i=1}^{m} \lambda_i (a_i^\top X a_i - 1) \right).$$

Необходимое условие экстремума по X:

$$\nabla_X L(X, \lambda) = 0.$$

Найдем этот градиент, расписав дифференциал.

$$\begin{aligned}
d\mathcal{L}(X, \lambda) &= d \left(\log \det X^{-1} + \sum_{i=1}^{m} \lambda_i (a_i^\top X a_i - 1) \right) \\
&= \frac{1}{\det X^{-1}} \det X^{-1} \left\langle (X^{-1})^{-\top}, -X^{-1} dX X^{-1} \right\rangle + \sum_{i=1}^{m} d\lambda_i (a_i^\top X a_i - 1) + \sum_{i=1}^{m} \lambda_i a_i^\top dX a_i \\
&= -\operatorname{tr}\left((X^{-1})^{-1} X^{-1} dX X^{-1} \right) + \sum_{i=1}^{m} d\lambda_i (a_i^\top X a_i - 1) + \sum_{i=1}^{m} \lambda_i \operatorname{tr}(a_i^\top dX a_i) \\
&= -\operatorname{tr}\left(X^{-1} dX \right) + \sum_{i=1}^{m} d\lambda_i (a_i^\top X a_i - 1) + \operatorname{tr}\left(\sum_{i=1}^{m} \lambda_i a_i a_i^\top dX \right) \\
&= \left\langle -X^{-\top} + \sum_{i=1}^{m} \lambda_i a_i a_i^\top, dX \right\rangle + \sum_{i=1}^{m} d\lambda_i (a_i^\top X a_i - 1).
\end{aligned}$$

Здесь было использовано свойство, что след сохраняется при циклической перестановке матриц в произведении, то есть, $\operatorname{tr}(ABC) = \operatorname{tr}(CAB) = \operatorname{tr}(BCA)$.

Возвращаемся к условию экстремума:

$$\nabla_X (X, \lambda) = -X^{-\top} + \sum_{i=1}^{m} \lambda_i a_i a_i^\top = 0.$$

Находим доставляющий оптимум X^*:

$$X^* = \left(\sum_{i=1}^{m} \lambda_i a_i a_i^\top \right)^{-1}.$$

Подставляем его в определение двойственной функции:

$$
\begin{aligned}
g(\lambda) &= \inf_{X \in \mathbb{S}^d_{++}} \left(\log \det X^{-1} + \sum_{i=1}^m \lambda_i (a_i^\top X a_i - 1) \right) \\
&= \log \det \left(\left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right)^{-1} \right)^{-1} + \sum_{i=1}^m \lambda_i \left(a_i^\top \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right)^{-1} a_i - 1 \right) \\
&= \log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right) + \sum_{i=1}^m \lambda_i a_i^\top \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right)^{-1} a_i - \sum_{i=1}^m \lambda_i \\
&= \log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right) + \sum_{i=1}^m \operatorname{tr} \left(\lambda_i a_i^\top \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right)^{-1} a_i \right) - \sum_{i=1}^m \lambda_i \\
&= \log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right) + \sum_{i=1}^m \operatorname{tr} \left(\lambda_i a_i a_i^\top \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right)^{-1} \right) - \sum_{i=1}^m \lambda_i \\
&= \log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right) + \operatorname{tr} \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right)^{-1} \right) - \sum_{i=1}^m \lambda_i \\
&= \log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right) + n - \sum_{i=1}^m \lambda_i .
\end{aligned}
$$

Наконец, перейдем к двойственной задаче:

$$
\begin{aligned}
&\max_{\lambda} g(\lambda) \\
&s.t.\ \lambda \succeq 0.
\end{aligned}
$$

■

Ответ:

$$
\begin{aligned}
&\max_{\lambda} \left(\log \det \left(\sum_{i=1}^m \lambda_i a_i a_i^\top \right) + n - \sum_{i=1}^m \lambda_i \right) \\
&s.t.\ \lambda \succeq 0.
\end{aligned}
$$

8. **(KKT)**

Примените условия ККТ для поиска всех решений следующей задачи:

$$
\begin{aligned}
\min_{x \in \mathbb{R}^2} & e^{x_1 - x_2} \\
\text{s.t. } & e^{x_1} + e^{x_2} \leq 20, \\
& x_1 \geq 0.
\end{aligned}
$$

Можно ли вы показать, что эти точки являются локальными решениями? Глобальными решениями?

Решение. □

Для начала запишем Лагранжиан:

$$
\mathcal{L}(x_1, x_2, \lambda_1, \lambda_2) = e^{x_1 - x_2} + \lambda_1 (e^{x_1} + e^{x_2} - 20) - \lambda_2 x_1 .
$$

Запишем также частные производные Лагранжиана:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_1} = e^{x_1-x_2} + \lambda_1 e^{x_1} - \lambda_2, \\ \frac{\partial \mathcal{L}}{\partial x_2} = -e^{x_1-x_2} + \lambda_1 e^{x_2}. \end{cases}$$

Запишем условия стационарности:

$$\begin{cases} e^{x_1-x_2} + \lambda_1 e^{x_1} - \lambda_2 = 0 \\ -e^{x_1-x_2} + \lambda_1 e^{x_2} = 0 \end{cases}$$

Откуда:

$$\begin{cases} \lambda_1 = e^{x_1-2x_2}, \\ \lambda_2 = e^{2x_1-2x_2} + e^{x_1-x_2} \end{cases}$$

Запишем условия дополняющей нежесткости:

$$\begin{cases} \lambda_1(e^{x_1} + e^{x_2} - 20) = 0, \\ \lambda_2 x_1 = 0. \end{cases}$$

Тогда $\lambda_1, \lambda_2 \neq 0$ – значит, $x_1 = 0$ и $x_2 = \ln 19$. Тогда подставляя выше полученные выражения $\lambda_1 = \frac{1}{361}$ и $\lambda_2 = \frac{20}{361}$ – можно заметить, что выполняются условия неотрицательности.

Так как все функции в условии являются выпуклыми, можно утверждать, что полученное решение – глобальный минимум.

■

Ответ: $(x_1, x_2) = (0, \ln 19)$ – глобальный минимум

9. **(Нахождение констант гладкости и/или сильной выпуклости)**

Пусть $g : \mathbb{R} \to \mathbb{R}$ – дифференцируемая функция, производная которой является липшицевой с параметром $L > 0$. Пусть $a, x \in \mathbb{R}^d$, $b \in \mathbb{R}$, и пусть $f : \mathbb{R}^d \to \mathbb{R}$ – функция $f(x) := g(\langle a, x\rangle + b)$. Покажите, что градиент функции f является липшицевым с параметром $L\|a\|^2$.

Решение. □

Вычислим градиент функции $f(x) = g\left(\langle a, x\rangle + b\right)$. Пусть $u(x) = \langle a, x\rangle + b$, откуда $\nabla u(x) = a$:

$$\nabla f(x) = g'(u(x))\nabla u(x) = a g'(u(x)).$$

По условию $g'(x)$ L-липшицева, то есть

$$|g'(u(x_1)) - g'(u(x_2))| \leq L\left|u(x_1) - u(x_2)\right| = L\left|\langle a, x_1\rangle + b - \langle a, x_2\rangle - b\right| = L\left|\langle a, x_1 - x_2\rangle\right|.$$

Скалярное произведение можно оценить через КБШ:

$$|\langle a, x_1 - x_2\rangle| \leq \|a\| \cdot \|x_1 - x_2\|.$$

Тогда:

$$|g'(u(x_1)) - g'(u(x_2))| \leq L\|a\| \cdot \|x_1 - x_2\|.$$

Теперь оценим норму разности градиентов:

$$\|\nabla f(x_1) - \nabla f(x_2)\| = \left\|a\left(g'(u(x_1)) - g'(u(x_2))\right)\right\| = \|a\| \cdot \left|g'(u(x_1)) - g'(u(x_2))\right|.$$

Подставляя липшицевость $g'(x)$:

$$\|\nabla f(x_1) - \nabla f(x_2)\| \leq L\|a\|^2\|x_1 - x_2\|.$$

■

10. **(Градиентный спуск)**

Напишите как будет выглядеть метод градиентного спуска для задачи безусловной минимизации функции:

$$f(x, y) = 6x^2 - 4\sqrt{13}xy + 26y^2.$$

Решение. □

В общем виде шаг градиентного спуска выглядит так:

$$x^{k+1} = x^k - \gamma_k \nabla f(x^k).$$

Посчитаем градиент:

$$\nabla f(x, y) = \left(12x - 4\sqrt{13}y,\ -4\sqrt{13}x + 52y\right)^\top.$$

Остается подставить.

■

Ответ:

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \gamma_k \begin{pmatrix} 12x^k - 4\sqrt{13}y^k \\ -4\sqrt{13}x^k + 52y^k \end{pmatrix}$$

11. **(Метод Ньютона)**

Найдите итерацию метода Ньютона функции Розенброка (функции узкой долины), найдите её минимум и значение в нём:

$$\min_{x, y \in \mathbb{R}} (1 - x)^2 + 100(y - x^2)^2,$$

Решение. □

В общем виде итерация метода Ньютона имеет вид:

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k)\right)^{-1} \nabla f(x^k).$$

Найдем градиент:

$$\nabla f(x, y) = \left(2(x - 1) - 400x(y - x^2),\ 200(y - x^2)\right)^\top$$

Найдем гессиан:

$$\nabla^2 f(x, y) = \begin{pmatrix} 2 - 400(y - 3x^2) & -400x \\ -400x & 200 \end{pmatrix}$$

И матрицу обратную к нему:

$$(\nabla^2 f(x, y))^{-1} = \frac{1}{200x^2 - 200y + 1} \begin{pmatrix} \frac{1}{2} & x \\ x & 3x^2 - y + \frac{1}{200} \end{pmatrix}$$

Подставим:

$$(\nabla^2 f(x, y))^{-1} \nabla f(x, y) = \frac{1}{200x^2 - 200y + 1} \left(x - 1,\ x^2 + y - 2x - 200(x^2 - y)^2\right)^\top.$$

Остается только подставить это в исходное выражение для шага метода Ньютона.

Для поиска минимума функции Розенброка заметим, что оба слагаемых всегда неотрицательны, поэтому минимум функции $\geqslant 0$. Приравняв градиент $\nabla f(x, y)$ к $\mathbf{0}$, получаем $(x, y) = (1, 1)$ – подставив в исходную функцию, получим, что $f(1, 1) = 0$ – что по замечанию выше и будет решением задачи. ■

Ответ:

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \frac{1}{200x^2 - 200y + 1} \begin{pmatrix} x^k - 1 \\ (x^k)^2 + y^k - 2x^k - 200((x^k)^2 - y^k)^2 \end{pmatrix}.$$

Минимум $f(x, y) = 0$ в $(x, y) = (1, 1)$.

12. **(Метод Франк-Вульфа)**

Выпишите для этой задачи k-ую итерацию метода Франк-Вульфа в явном виде:

$$\begin{aligned} &\min_{x \in \mathbb{R}^d} \|Ax + b\|_2^2 \\ &\text{s.t. } \sum_{i=1}^{d} x_i = R, \\ &\quad x_i \geq 0, \ i \in \overline{1, d}, \end{aligned}$$

где $A \in \mathbb{R}^{n \times d}$ и $b \in \mathbb{R}^n$.

Подсказка: $\gamma_k = \frac{2}{k+2}$.

Решение. □

Бюджетное множество – симплекс $S = \left\{ x \in \mathbb{R}^d_+ \;\middle|\; \sum\limits_{i=1}^{d} x_i = R \right\}$.

Целевая функция и ее градиент:

$$\begin{aligned} f(x) &= \|Ax + b\|_2^2, \\ \nabla f(x) &= 2A^\top (Ax + b). \end{aligned}$$

Вспомним, как выглядит итерация метода Франк-Вульфа:

$$x^{k+1} = (1 - \gamma_k)x^k + \gamma_k s^k, \ s^k = \arg\min_{s \in S} \langle \nabla f(x^k), s \rangle.$$

Для того, чтобы выписать ее явно надо научиться решать $\arg\min$ на симплексе. Формула была получена в пособии:

$$s^k = R\mathbf{e_i}, \text{ где } i = \arg\min_j [\nabla f(x^k)]_j.$$

Объединяя все, получаем ответ.

■

Ответ:

$$x^{k+1} = \frac{k}{k+2} x^k + \frac{2}{k+2} R\mathbf{e_i}, \text{ где } i = \arg\min_j \left[2A^\top (Ax^k + b) \right]_j.$$