Continuous Optimization
Homework 2

Lev Khoroshansky

Задача 1. Рассмотрим каждый пункт отдельно.

- $f(\mathbf{x}) = \log(1 + \mathbf{x}^\mathbf{T}\mathbf{A}\mathbf{x})$, где $\mathbf{A} \in \mathbb{S}^n_+$. Начнём с дифференциала:

$$\begin{aligned}\mathbf{d}f(\mathbf{x}) &= \mathbf{d}\log(1 + \mathbf{x}^\mathbf{T}\mathbf{A}\mathbf{x}) = \mathbf{d}\log(1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle) \\ &= \frac{\mathbf{d}(1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle)}{1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle} = \frac{2\langle \mathbf{A}\mathbf{x}, \mathbf{dx}\rangle}{1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle}\end{aligned}$$

Используя равенство $\mathbf{d}f(x) = \langle \nabla f(\mathbf{x}), \mathbf{dx}\rangle$, получаем $\nabla f(\mathbf{x}) = \dfrac{2\mathbf{A}\mathbf{x}}{1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle}$.
Теперь найдём гессиан при помощи второго дифференциала:

$$\begin{aligned}\mathbf{d}^2 f(\mathbf{x}) = \mathbf{d}[\mathbf{d}f(\mathbf{x})] &= \mathbf{d}\left[\frac{2\langle \mathbf{A}\mathbf{x}, \mathbf{dx_1}\rangle}{1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle}\right] \\ &= \frac{\mathbf{d}[2\langle \mathbf{A}\mathbf{x}, \mathbf{dx_1}\rangle](1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle) - 2\langle \mathbf{A}\mathbf{x}, \mathbf{dx_1}\rangle\mathbf{d}[1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle]}{(1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle)^2} \\ &= \frac{2\langle \mathbf{A}\mathbf{dx_1}, \mathbf{dx_2}\rangle}{1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle} - \frac{4\langle \mathbf{A}\mathbf{x}, \mathbf{dx_1}\rangle\langle \mathbf{A}\mathbf{x}, \mathbf{dx_2}\rangle}{(1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle)^2} \\ &= \left\langle \frac{2\mathbf{A}}{1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle} - \frac{4\,\mathbf{A}\mathbf{x}\mathbf{x}^\mathbf{T}\mathbf{A}}{(1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle)^2}\,\mathbf{dx_1}, \mathbf{dx_2}\right\rangle\end{aligned}$$

Откуда $\nabla^2 f(\mathbf{x}) = \dfrac{2\mathbf{A}}{1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle} - \dfrac{4\,\mathbf{A}\mathbf{x}\mathbf{x}^\mathbf{T}\mathbf{A}}{(1 + \langle \mathbf{A}\mathbf{x}, \mathbf{x}\rangle)^2}$.

- $f(\mathbf{x}) = \langle \mathbf{x}, \mathbf{x}\rangle^{\langle \mathbf{x}, \mathbf{x}\rangle}$. Точно также, сначала найдём дифференциал:

$$\begin{aligned}\mathbf{d}f(\mathbf{x}) &= \mathbf{d}\left(\langle \mathbf{x}, \mathbf{x}\rangle^{\langle \mathbf{x}, \mathbf{x}\rangle}\right) = \mathbf{d}\exp\left(\langle \mathbf{x}, \mathbf{x}\rangle \ln\langle \mathbf{x}, \mathbf{x}\rangle\right) \\ &= \langle \mathbf{x}, \mathbf{x}\rangle^{\langle \mathbf{x}, \mathbf{x}\rangle} \cdot [\mathbf{d}\langle \mathbf{x}, \mathbf{x}\rangle\ \ln\langle \mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{x}, \mathbf{x}\rangle\,\mathbf{d}\ln\langle \mathbf{x}, \mathbf{x}\rangle] \\ &= \langle \mathbf{x}, \mathbf{x}\rangle^{\langle \mathbf{x}, \mathbf{x}\rangle} \cdot \left[2\langle \mathbf{x}, \mathbf{dx}\rangle \cdot \ln\langle \mathbf{x}, \mathbf{x}\rangle + \langle \mathbf{x}, \mathbf{x}\rangle \cdot \frac{\mathbf{d}\langle \mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle}\right] \\ &= \left\langle 2\,\underbrace{\langle \mathbf{x}, \mathbf{x}\rangle^{\langle \mathbf{x}, \mathbf{x}\rangle}}_{f(\mathbf{x})}(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)\,\mathbf{x}, \mathbf{dx}\right\rangle\end{aligned}$$

Выразим градиент: $\nabla f(\mathbf{x}) = 2f(\mathbf{x})(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)\,\mathbf{x}$. Второй дифференциал:

$$\begin{aligned}\mathbf{d}^2 f(\mathbf{x}) &= \mathbf{d}[2f(\mathbf{x})\langle \mathbf{x}, \mathbf{dx_1}\rangle(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)] \\ &= 2\,\underbrace{\mathbf{d}f(\mathbf{x}) \cdot [\langle \mathbf{x}, \mathbf{dx_1}\rangle(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)]}_{\Leo} + 2\,\underbrace{f(\mathbf{x}) \cdot \mathbf{d}[\langle \mathbf{x}, \mathbf{dx_1}\rangle(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)]}_{\Cancer}\end{aligned}$$

Раскроем каждое слагаемое по отдельности:

$$\leo \quad = 2f(\mathbf{x})\langle \mathbf{x}, \mathbf{dx_1}\rangle\langle \mathbf{x}, \mathbf{dx_2}\rangle(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)^2 = \langle 2f(\mathbf{x})(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)^2\, \mathbf{x}\mathbf{x}^\mathbf{T}\, \mathbf{dx_1}, \mathbf{dx_2}\rangle$$

$$\begin{aligned}
\cancer \quad &= f(\mathbf{x}) \left[\langle \mathbf{dx_1}, \mathbf{dx_2}\rangle(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1) + \langle \mathbf{x}, \mathbf{dx_1}\rangle\, \frac{\mathbf{d}\langle \mathbf{x}, \mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle}\right] \\
&= f(\mathbf{x}) \left[\langle \mathbf{dx_1}, \mathbf{dx_2}\rangle(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1) + \langle \mathbf{x}, \mathbf{dx_1}\rangle\, \frac{2\langle \mathbf{x}, \mathbf{dx_2}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle}\right] \\
&= \left\langle f(\mathbf{x}) \left[(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)\, \mathbf{I_n} + \frac{2\mathbf{x}\mathbf{x}^\mathbf{T}}{\langle \mathbf{x}, \mathbf{x}\rangle}\right] \mathbf{dx_1}, \mathbf{dx_2}\right\rangle
\end{aligned}$$

Итоговый результат:

$$\mathbf{d}^2 f(\mathbf{x}) = \left\langle \underbrace{2f(\mathbf{x}) \left[(\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)\, \mathbf{I_n} + 2\left((\ln\langle \mathbf{x}, \mathbf{x}\rangle + 1)^2 + \frac{1}{\langle \mathbf{x}, \mathbf{x}\rangle}\right) \mathbf{x}\mathbf{x}^\mathbf{T}\right]}_{\nabla^2 f(\mathbf{x})} \mathbf{dx_1}, \mathbf{dx_2}\right\rangle$$

- $f(\mathbf{x}) = \ln\left(\sum\limits_{i=1}^{m} \exp\langle \mathbf{a_i}, \mathbf{x}\rangle\right)$. Обозначим $g_i = \exp\langle \mathbf{a_i}, \mathbf{x}\rangle$ для удобства. Первый дифференциал:

$$\mathbf{d}f(\mathbf{x}) = \frac{\mathbf{d}\sum\limits_{i=1}^{m} g_i}{\sum\limits_{i=1}^{m} g_i} = \frac{\sum\limits_{i=1}^{m} g_i\, \langle \mathbf{a_i}, \mathbf{dx}\rangle}{\sum\limits_{i=1}^{m} g_i}$$

Откуда имеем $\nabla f(\mathbf{x}) = \dfrac{\sum\limits_{i=1}^{m} \exp\langle \mathbf{a_i}, \mathbf{x}\rangle\, \mathbf{a_i}}{\sum\limits_{i=1}^{m} \exp\langle \mathbf{a_i}, \mathbf{x}\rangle}$. Второй дифференциал:

$$\mathbf{d}^2 f(\mathbf{x}) = \left(\sum_{i=1}^{m} g_i\right)^{-2} \left[\underbrace{\mathbf{d}\left(\sum_{i=1}^{m} g_i\, \langle \mathbf{a_i}, \mathbf{dx_1}\rangle\right)}_{\leo} \left(\sum_{i=1}^{m} g_i\right) - \left(\sum_{i=1}^{m} g_i\, \langle \mathbf{a_i}, \mathbf{dx_1}\rangle\right) \underbrace{\mathbf{d}\left(\sum_{i=1}^{m} g_i\right)}_{\cancer}\right]$$

Рассмотрим обозначенные дифференциалы отдельно.

$$\leo = \sum_{i=1}^{m} \mathbf{d}g_i \cdot \langle \mathbf{a_i}, \mathbf{dx_1}\rangle + \cancel{g_i \cdot \mathbf{d}\langle \mathbf{a_i}, \mathbf{dx_1}\rangle} = \sum_{i=1}^{m} g_i\, \langle \mathbf{a_i}, \mathbf{dx_2}\rangle\, \langle \mathbf{a_i}, \mathbf{dx_1}\rangle = \sum_{i=1}^{m} g_i\, \langle \mathbf{a_i}\mathbf{a_i^T}\, \mathbf{dx_1}, \mathbf{dx_2}\rangle$$

$$\cancer = \sum_{i=1}^{m} g_i \langle \mathbf{a_i}, \mathbf{dx_2}\rangle$$

Всё вместе:

$$\mathbf{d}^2 f(\mathbf{x}) = \left\langle \underbrace{\left(\sum_{i=1}^{m} g_i\right)^{-2} \sum_{i=1}^{m}\sum_{j=1}^{m} g_i\, g_j\, \left[\mathbf{a_i}\mathbf{a_i^T} - \mathbf{a_j}\mathbf{a_i^T}\right]}_{\nabla^2 f(\mathbf{x})} \mathbf{dx_1}, \mathbf{dx_2}\right\rangle$$

- $f(\mathbf{x}) = \dfrac{\langle \mathbf{Ax}, \mathbf{x}\rangle}{\|\mathbf{x}\|_2^2}$, где $\mathbf{A} \in \mathbb{S}^n$. Первый дифференциал:

$$
\begin{aligned}
\mathbf{d}f(\mathbf{x}) = \mathbf{d}\frac{\langle \mathbf{Ax}, \mathbf{x}\rangle}{\langle \mathbf{x}, \mathbf{x}\rangle} &= \langle \mathbf{x}, \mathbf{x}\rangle^{-2}\left[\mathbf{d}\langle \mathbf{Ax}, \mathbf{x}\rangle \cdot \langle \mathbf{x}, \mathbf{x}\rangle - \langle \mathbf{Ax}, \mathbf{x}\rangle \cdot \mathbf{d}\langle \mathbf{x}, \mathbf{x}\rangle\right] \\
&= \langle \mathbf{x}, \mathbf{x}\rangle^{-2}\left[2\langle \mathbf{Ax}, \mathbf{dx}\rangle \cdot \langle \mathbf{x}, \mathbf{x}\rangle - 2\langle \mathbf{Ax}, \mathbf{x}\rangle \cdot \langle \mathbf{x}, \mathbf{dx}\rangle\right] \\
&= 2\langle \mathbf{x}, \mathbf{x}\rangle^{-2}\left\langle \langle \mathbf{x}, \mathbf{x}\rangle\, \mathbf{Ax} - \langle \mathbf{Ax}, \mathbf{x}\rangle\, \mathbf{x}, \mathbf{dx}\right\rangle
\end{aligned}
$$

Градиент: $\nabla f(\mathbf{x}) = 2\,\dfrac{\langle \mathbf{x}, \mathbf{x}\rangle\, \mathbf{Ax} - \langle \mathbf{Ax}, \mathbf{x}\rangle\, \mathbf{x}}{\langle \mathbf{x}, \mathbf{x}\rangle^2}$. Второй дифференциал:

$$
\mathbf{d}^2 f(\mathbf{x}) = 2\left[\underbrace{\mathbf{d}\left(\langle \mathbf{x}, \mathbf{x}\rangle^{-1}\langle \mathbf{Ax}, \mathbf{dx_1}\rangle\right)}_{\leo} - \underbrace{\mathbf{d}\left(\langle \mathbf{x}, \mathbf{x}\rangle^{-2}\langle \mathbf{Ax}, \mathbf{x}\rangle\langle \mathbf{x}, \mathbf{dx_1}\rangle\right)}_{\cancer}\right]
$$

Обозначенные дифференциалы:

$$
\begin{aligned}
\leo &= \mathbf{d}\langle \mathbf{x}, \mathbf{x}\rangle^{-1} \cdot \langle \mathbf{Ax}, \mathbf{dx_1}\rangle + \langle \mathbf{x}, \mathbf{x}\rangle^{-1} \cdot \mathbf{d}\langle \mathbf{Ax}, \mathbf{dx_1}\rangle \\
&= -2\langle \mathbf{x}, \mathbf{x}\rangle^{-2}\langle \mathbf{x}, \mathbf{dx_2}\rangle\langle \mathbf{Ax}, \mathbf{dx_1}\rangle + \langle \mathbf{x}, \mathbf{x}\rangle^{-1}\langle \mathbf{Adx_2}, \mathbf{dx_1}\rangle \\
&= \left\langle \frac{\mathbf{x^T x\, A} - 2\mathbf{xx^T A}}{\langle \mathbf{x}, \mathbf{x}\rangle^2}\, \mathbf{dx_1}, \mathbf{dx_2}\right\rangle \\
\cancer &= \mathbf{d}\langle \mathbf{x}, \mathbf{x}\rangle^{-2} \cdot \left[\langle \mathbf{Ax}, \mathbf{x}\rangle\langle \mathbf{x}, \mathbf{dx_1}\rangle\right] + \langle \mathbf{x}, \mathbf{x}\rangle^{-2} \cdot \mathbf{d}\left[\langle \mathbf{Ax}, \mathbf{x}\rangle\langle \mathbf{x}, \mathbf{dx_1}\rangle\right] \\
&= -4\langle \mathbf{x}, \mathbf{x}\rangle^{-3}\langle \mathbf{x}, \mathbf{dx_2}\rangle\langle \mathbf{Ax}, \mathbf{x}\rangle\langle \mathbf{x}, \mathbf{dx_1}\rangle + \langle \mathbf{x}, \mathbf{x}\rangle^{-2}[2\langle \mathbf{Ax}, \mathbf{dx_2}\rangle\langle \mathbf{x}, \mathbf{dx_1}\rangle + \langle \mathbf{Ax}, \mathbf{x}\rangle\langle \mathbf{dx_1}, \mathbf{dx_2}\rangle] \\
&= \left\langle \frac{-4\,\mathbf{xx^T Axx^T}}{\langle \mathbf{x}, \mathbf{x}\rangle^3}\, \mathbf{dx_1}, \mathbf{dx_2}\right\rangle + \left\langle \frac{2\,\mathbf{Axx^T} + \mathbf{x^T Ax\, I_n}}{\langle \mathbf{x}, \mathbf{x}\rangle^2}\, \mathbf{dx_1}, \mathbf{dx_2}\right\rangle
\end{aligned}
$$

Итого:

$$
\mathbf{d}^2 f(\mathbf{x}) = \left\langle \underbrace{\left[\frac{2\mathbf{x^T x\, A} - 4\mathbf{Axx^T} - 4\,\mathbf{xx^T A} - 2\,\mathbf{x^T Ax\, I_n}}{\langle \mathbf{x}, \mathbf{x}\rangle^2} + \frac{8\,\mathbf{xx^T Axx^T}}{\langle \mathbf{x}, \mathbf{x}\rangle^3}\right]}_{\nabla^2 f(\mathbf{x})} \mathbf{dx_1}, \mathbf{dx_2}\right\rangle
$$

Задача 2. Рассмотрим каждый пункт отдельно.

- $f(\mathbf{x}) = \|\mathbf{x}\|_2$. Первый дифференциал:

$$
\mathbf{d}f(\mathbf{x}) = \mathbf{d}\langle \mathbf{x}, \mathbf{x}\rangle^{1/2} = \frac{1}{2}\,\langle \mathbf{x}, \mathbf{x}\rangle^{-1/2}\, \mathbf{d}\langle \mathbf{x}, \mathbf{x}\rangle = \|\mathbf{x}\|_2^{-1}\ \langle \mathbf{x}, \mathbf{dx}\rangle
$$

Градиент: $\nabla f(\mathbf{x}) = \|\mathbf{x}\|_2^{-1}\ \mathbf{x}$. Второй дифференциал:

$$
\begin{aligned}
\mathbf{d}^2 f(\mathbf{x}) &= \mathbf{d}\,\|\mathbf{x}\|_2^{-1} \cdot \langle \mathbf{x}, \mathbf{dx_1}\rangle + \|\mathbf{x}\|_2^{-1} \cdot \mathbf{d}\langle \mathbf{x}, \mathbf{dx_1}\rangle \\
&= -\,\|\mathbf{x}\|_2^{-2}\left(\|\mathbf{x}\|_2^{-1}\,\langle \mathbf{x}, \mathbf{dx_2}\rangle\right)\langle \mathbf{x}, \mathbf{dx_1}\rangle + \|\mathbf{x}\|_2^{-1} \cdot \langle \mathbf{dx_2}, \mathbf{dx_1}\rangle \\
&= \|\mathbf{x}\|_2^{-1} \cdot \langle \mathbf{dx_1}, \mathbf{dx_2}\rangle - \|\mathbf{x}\|_2^{-3}\,\langle \mathbf{xx^T dx_1}, \mathbf{dx_2}\rangle \\
&= \left\langle \underbrace{\|\mathbf{x}\|_2^{-1}\left(\mathbf{I_n} - \|\mathbf{x}\|_2^{-2}\ \mathbf{xx^T}\right)}_{\nabla^2 f(\mathbf{x})} \mathbf{dx_1}, \mathbf{dx_2}\right\rangle
\end{aligned}
$$

Заметим, что гессиан является положительно полуопределённым.
Пусть $\mathbf{y} \in \mathbb{R}^n$ – случайный вектор, тогда

$$
\begin{aligned}
&0 \leqslant \left\langle \nabla^2 f(\mathbf{x})\,\mathbf{y}, \mathbf{y} \right\rangle = \|\mathbf{x}\|_2^{-1}\,\mathbf{y}^\mathbf{T}\left(\mathbf{I_n} - \|\mathbf{x}\|_2^{-2}\,\mathbf{x}\mathbf{x}^\mathbf{T}\right)\mathbf{y} \\
&0 \leqslant \langle \mathbf{y}, \mathbf{y} \rangle - \|\mathbf{x}\|_2^{-2}\,\langle \mathbf{y}, \mathbf{x} \rangle^2 && \text{(норма не влияет на знак)} \\
&0 \leqslant \langle \mathbf{y}, \mathbf{y} \rangle - \|\mathbf{x}\|_2^{-2}\,\|\mathbf{x}\|_2^2\,\|\mathbf{y}\|_2^2 \leqslant \langle \mathbf{y}, \mathbf{y} \rangle - \|\mathbf{x}\|_2^{-2}\,\langle \mathbf{y}, \mathbf{x} \rangle^2 && \text{(неравенство Коши-Буняковского)} \\
&0 \leqslant \langle \mathbf{y}, \mathbf{y} \rangle - \langle \mathbf{y}, \mathbf{y} \rangle
\end{aligned}
$$

- $f(\mathbf{x}) = -(1 + \langle \mathbf{x}, \mathbf{x} \rangle)^{-1}$. Первый дифференциал:

$$
\mathbf{d}f(\mathbf{x}) = (1 + \langle \mathbf{x}, \mathbf{x} \rangle)^{-2}\mathbf{d}\langle \mathbf{x}, \mathbf{x} \rangle = 2\,\underbrace{(1 + \langle \mathbf{x}, \mathbf{x} \rangle)^{-2}}_{f(\mathbf{x})^2}\langle \mathbf{x}, \mathbf{dx} \rangle
$$

Градиент: $\nabla f(\mathbf{x}) = 2f(\mathbf{x})^2\,\mathbf{x}$. Второй дифференциал:

$$
\begin{aligned}
\mathbf{d}^2 f(\mathbf{x}) &= 2[\mathbf{d}f(\mathbf{x})^2 \cdot \langle \mathbf{x}, \mathbf{dx_1} \rangle + f(\mathbf{x})^2 \cdot \mathbf{d}\langle \mathbf{x}, \mathbf{dx_1} \rangle] \\
&= 2\left[4f(\mathbf{x})^3\,\langle \mathbf{x}, \mathbf{dx_2} \rangle \langle \mathbf{x}, \mathbf{dx_1} \rangle + f(\mathbf{x})^2 \langle \mathbf{dx_1}, \mathbf{dx_2} \rangle\right] \\
&= \left\langle \underbrace{\left[8f(\mathbf{x})^3\,\mathbf{x}\mathbf{x}^\mathbf{T} + 2f(\mathbf{x})^2\,\mathbf{I_n}\right]}_{\nabla^2 f(\mathbf{x})}\,\mathbf{dx_1}, \mathbf{dx_2} \right\rangle
\end{aligned}
$$

Проверим гессиан на (полу)определённость. Пусть $\mathbf{y} \in \mathbb{R}^n$, тогда:

$$
\begin{aligned}
\left\langle \nabla^2 f(\mathbf{x})\,\mathbf{y}, \mathbf{y} \right\rangle &= 8f(\mathbf{x})^3 \langle \mathbf{x}, \mathbf{y} \rangle^2 + 2f(\mathbf{x})^2 \langle \mathbf{y}, \mathbf{y} \rangle = 2f(\mathbf{x})^2\left[\langle \mathbf{y}, \mathbf{y} \rangle + 4f(\mathbf{x})\langle \mathbf{x}, \mathbf{y} \rangle^2\right] \\
&\geqslant 2f(\mathbf{x})^2\left[\langle \mathbf{y}, \mathbf{y} \rangle - \frac{4\langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle}{1 + \langle \mathbf{x}, \mathbf{x} \rangle}\right] = \underbrace{2f(\mathbf{x})^2 \langle \mathbf{y}, \mathbf{y} \rangle}_{\geqslant 0}\,\frac{1 - 3\langle \mathbf{x}, \mathbf{x} \rangle}{1 + \langle \mathbf{x}, \mathbf{x} \rangle}
\end{aligned}
$$

Выше мы воспользовались неравенством Коши-Буняковского, где равенство достигается при пропорциональных векторах. Рассмотрим несколько случаев:

$\langle \mathbf{x}, \mathbf{x} \rangle < \dfrac{1}{3}$: гессиан положительно определён,

$\langle \mathbf{x}, \mathbf{x} \rangle = \dfrac{1}{3}$: гессиан положительно полуопределён,

$\langle \mathbf{x}, \mathbf{x} \rangle > \dfrac{1}{3}$: если $n = 1$, то гессиан отрицательно определён (равенство всегда выполняется в силу пропорциональности), а иначе – неопределён.

- В этом и следующем пунктах было сказано посчитать только первые дифференциалы.
$f(\mathbf{X}) = \mathbf{X}^{-1}$. Первый дифференциал:

$$
\mathbf{d}f(\mathbf{X}) = -\mathbf{X}^{-1}[\mathbf{dX}]\,\mathbf{X}^{-1}
$$

- $f(\mathbf{X}) = \langle \mathbf{X}^{-1}\,\mathbf{v}, \mathbf{v} \rangle$.
Мы знаем, что если скалярное произведение для матриц определено при помощи произведения Фробениуса, то для матрицы $\mathbf{A}$ и вектора $\mathbf{u}$ выражение $\langle \mathbf{Au}, \mathbf{u} \rangle$ эквивалентно $\langle \mathbf{u}\mathbf{u}^\mathbf{T}, \mathbf{A} \rangle$.
Воспользуемся этим при поиске первого дифференциала:

$$
\begin{aligned}
\mathbf{d}\langle \mathbf{X^{-1}}\,\mathbf{v}, \mathbf{v} \rangle &= \mathbf{d}\langle \mathbf{v}\mathbf{v}^\mathbf{T}, \mathbf{X^{-1}} \rangle = \langle \mathbf{v}\mathbf{v}^\mathbf{T}, \mathbf{dX^{-1}} \rangle = \langle \mathbf{v}\mathbf{v}^\mathbf{T}, -\mathbf{X^{-1}}[\mathbf{dX}]\,\mathbf{X^{-1}} \rangle \\
&= \langle \underbrace{-[\mathbf{X^{-1}}]^\mathbf{T}\mathbf{v}\mathbf{v}^\mathbf{T}[\mathbf{X^{-1}}]^\mathbf{T}}_{\nabla f(\mathbf{X})}, \mathbf{dX} \rangle
\end{aligned}
$$

Задача 3. Рассмотрим каждый пункт отдельно.

- Выпишем исходную задачу:

$$\begin{cases} f_0(\mathbf{x}) = \sum\limits_{k=1}^{n} c_k\,\mathbf{x}_k^{-1} \\ f_k(\mathbf{x}) = -\mathbf{x}_k \leqslant 0 \quad \forall k = 1,\ldots,n \\ f_{n+1}(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x}\rangle - b \leqslant 0 \end{cases}$$

Тогда лагранжиан выглядит следующим образом:

$$\mathcal{L}(\mathbf{x},\lambda) = f_0(\mathbf{x}) + \sum_{k=1}^{n} \lambda_k f_k(\mathbf{x}) + \lambda_{n+1} f_{n+1}(\mathbf{x})$$

Условия Каруша-Куна-Такера:

$$\begin{cases} \nabla_{\mathbf{x}_k}\mathcal{L}(\mathbf{x},\lambda) = -c_k\mathbf{x}_k^{-2} - \lambda_k + \lambda_{n+1}\mathbf{a}_k = 0 & \forall k = 1,\ldots,n \\ \lambda_k\,\mathbf{x}_k = 0 & \forall k = 1,\ldots,n \\ \lambda_{n+1}(\langle \mathbf{a},\mathbf{x}\rangle - b) = 0 & \\ \lambda_k \geqslant 0,\;\; -\mathbf{x}_k \leqslant 0 & \forall k = 1,\ldots,n \\ \lambda_{\mathbf{n+1}} \geqslant 0,\;\; \langle \mathbf{a},\mathbf{x}\rangle - b \leqslant 0 & \end{cases}$$

Заметим, что по условию $\mathbf{x}_k > 0$, откуда следует, что $\lambda_k = 0 \;\; \forall k = 1,\ldots,n$. Упростим:

$$\begin{cases} \nabla_{\mathbf{x}_k}\mathcal{L}(\mathbf{x},\lambda) = -c_k\mathbf{x}_k^{-2} + \lambda_{n+1}\mathbf{a}_k = 0 \quad \forall k = 1,\ldots,n \\ \lambda_{n+1}(\langle \mathbf{a},\mathbf{x}\rangle - b) = 0 \\ \lambda_{n+1} \geqslant 0,\;\; \langle \mathbf{a},\mathbf{x}\rangle - b \leqslant 0 \end{cases}$$

Также видно, что при $\lambda_{n+1} = 0$ градиент Лагранжиана станет строго отрицательным, что противоречит условиям. Остаётся случай, где $\lambda_{n+1} > 0$ и $\langle \mathbf{a},\mathbf{x}\rangle = b$:

$$\nabla_{\mathbf{x}_k}\mathcal{L}(\mathbf{x},\lambda) = -c_k\mathbf{x}_k^{-2} + \lambda_{n+1}\mathbf{a}_k = 0 \quad \longrightarrow \quad \mathbf{x}_k = \sqrt{\frac{c_k}{\lambda_{n+1}\,\mathbf{a}_k}}$$

$$\langle \mathbf{a},\mathbf{x}\rangle = \sum_{k=1}^{n}\sqrt{\frac{c_k\,\mathbf{a}_k}{\lambda_{n+1}}} = b \quad \longrightarrow \quad \lambda_{n+1} = \left[b^{-1}\sum_{k=1}^{n}\sqrt{\mathbf{a}_k\,c_k}\right]^2$$

Для простоты восприятия, не будем явно подставлять выражение λ_{n+1} через $\mathbf{a}_k, b, c_k$ в формулу для $\mathbf{x}_k$, достаточно лишь понимать, что $\mathbf{x}_k$ также можно выразить через $\mathbf{a}_k, b, c_k$.

Исходная задача является выпуклой (как композиция выпуклых функций), а для выпуклых задач условия ККТ являются достаточными, следовательно, мы нашли ответ.

- Выпишем исходную задачу:

$$\begin{cases} f_0(\mathbf{x}) = \mathbf{x}_1\mathbf{x}_3 - 2\mathbf{x}_2 \\ f_1(\mathbf{x}) = 2\mathbf{x}_1 - \mathbf{x}_2 - 3\mathbf{x}_3 - 10 \leqslant 0 \\ f_2(\mathbf{x}) = -\mathbf{x}_2 \leqslant 0 \\ h(\mathbf{x}) = 3\mathbf{x}_1 + 2\mathbf{x}_2 + \mathbf{x}_3 - 6 = 0 \end{cases}$$

Лагранжиан:

$$\mathcal{L}(\mathbf{x},\lambda,\nu) = f_0(\mathbf{x}) + \sum_{k=1}^{2}\lambda_k f_k(\mathbf{x}) + \nu h(\mathbf{x})$$

Условия Каруша-Куна-Такера:

$$\begin{cases} \nabla_{\mathbf{x}_1}\mathcal{L}(\mathbf{x},\lambda,\nu) = \mathbf{x}_3 + 2\lambda_1 + 3\nu = 0 \\ \nabla_{\mathbf{x}_2}\mathcal{L}(\mathbf{x},\lambda,\nu) = -2 - \lambda_1 - \lambda_2 + 2\nu = 0 \\ \nabla_{\mathbf{x}_3}\mathcal{L}(\mathbf{x},\lambda,\nu) = \mathbf{x}_1 - 3\lambda_1 + \nu = 0 \\ \lambda_1[2\mathbf{x}_1 - \mathbf{x}_2 - 3\mathbf{x}_3 - 10] = 0 \\ \lambda_2[-\mathbf{x}_2] = 0 \\ \lambda_1 \geqslant 0, 2\mathbf{x}_1 - \mathbf{x}_2 - 3\mathbf{x}_3 - 10 \leqslant 0 \\ \lambda_2 \geqslant 0, -\mathbf{x}_2 \leqslant 0 \\ 3\mathbf{x}_1 + 2\mathbf{x}_2 + \mathbf{x}_3 - 6 = 0 \end{cases}$$

1. При $2\mathbf{x}_1 - \mathbf{x}_2 - 3\mathbf{x}_3 - 10 = 0, \mathbf{x}_2 = 0$ выходит, что $\lambda_2 < 0$, что противоречит условию. Такая же ситуация и при $\lambda_1 = 0, \mathbf{x}_2 = 0$ – в решении присутствует $\lambda_2 < 0$.

2. Пусть $2\mathbf{x}_1 - \mathbf{x}_2 - 3\mathbf{x}_3 - 10 = 0, \lambda_2 = 0$, тогда $\mathbf{x}_1 = \dfrac{2}{7}, \mathbf{x}_2 = \dfrac{174}{35}, \mathbf{x}_3 = -\dfrac{24}{5}, \lambda_1 = \dfrac{18}{35}, \nu = \dfrac{44}{35}$. При этом решении количество активных ограничений меньше, чем необходимо для выполнения достаточного условия первого порядка, исследуем условие второго порядка. Ограничения для вектора $\mathbf{z} \in \mathbb{R}^3$ выглядят следующим образом:

   $$\begin{cases} 2\mathbf{z}_1 - \mathbf{z}_2 - 3\mathbf{z}_3 = 0 \\ 3\mathbf{z}_1 + 2\mathbf{z}_2 + \mathbf{z}_3 = 0 \\ \langle \nabla^2\mathcal{L}(\mathbf{x}^*, \lambda^*, \nu^*)\,\mathbf{z}, \mathbf{z}\rangle = 2\mathbf{z}_1\mathbf{z}_3 \geqslant 0 \end{cases} \longrightarrow \begin{cases} 7\mathbf{z}_1 = 5\mathbf{z}_3 \\ 2\mathbf{z}_1\mathbf{z}_3 = \frac{10}{7}\,\mathbf{z}_3^2 > 0 \quad (\mathbf{z}_3 > 0, \text{ иначе } \mathbf{z} = 0) \end{cases}$$

   Видно, что в данном случае найденная точка действительно является локальным минимумом, причём целевая функция достигает значения $-\dfrac{396}{35}$.

3. Пусть $\lambda_1 = 0, \lambda_2 = 0$, тогда $\mathbf{x}_1 = -1, \mathbf{x}_2 = 6, \mathbf{x}_3 = -3, \nu = 1$. В данном случае было истрачено множество попыток доказать, является ли эта точка экстремумом или нет, но никаких результатов получено не было. С большой вероятностью это не экстремум.

4. При рассмотрении задачи максимизации (достаточно наложить условия $\lambda \leqslant 0$ на исходную задачу) имеет место одно решение, при котором $\mathbf{x}_1 = 1, \mathbf{x}_2 = 0, \mathbf{x}_3 = 3, \lambda_1 = 0, \lambda_2 = -4, \nu = 1$, причём аналогично ранее разобранному случаю, выполняется условие второго порядка со значением целовой функции 3.

Задача 4. Решим прямую задачу:

$$\begin{cases} 3x_1 + 2x_2 \to \max \\ x_1 + 3x_2 \leqslant 3 \\ x_1 + 2x_2 \leqslant 2 \\ 6x_1 - x_2 = 4 \\ x_1, x_2 \geqslant 0 \end{cases} \longrightarrow \begin{cases} x_1 \to \max \\ x_1 \leqslant {}^{15}\!/_{19} \\ x_1 \leqslant {}^{10}\!/_{13} \\ x_2 = 6x_1 - 4 \geqslant 0 \\ x_1 \geqslant {}^{3}\!/_{2} \end{cases} \longrightarrow \begin{cases} \max(3x_1 + 2x_2) = {}^{46}\!/_{13} \\ x_1 = {}^{10}\!/_{13} \\ x_2 = {}^{8}\!/_{13} \end{cases}$$

Теперь составим двойственную задачу:

$$\begin{cases} 3y_1 + 2y_2 + 4y_3 \to \min \\ y_1 + y_2 + 6y_3 \geqslant 3 \\ 3y_1 + 2y_2 - y_3 \geqslant 2 \\ y_1, y_2 \geqslant 0, y_3 \in \mathbb{R} \end{cases}$$

Выразим целевую функцию как переменную и решим задачу:

$$\begin{cases} z = 3y_1 + 2y_2 + 4y_3 \to \min \\ y_2 = \frac{1}{2}\left[z - 3y_1 - 4y_3\right] \geqslant 0 \\ -y_1 + z + 8y_3 \geqslant 6 \\ z - 5y_3 \geqslant 2 \\ y_1 \geqslant 0, y_3 \in \mathbb{R} \end{cases} \longrightarrow \begin{cases} z \to \min \\ y_2 = \frac{1}{2}\left[z - 3y_1 - 4y_3\right] \geqslant 0 \\ z - 5y_3 \geqslant 2 \\ z + 8y_3 \geqslant 6 \\ z - 4y_3 \geqslant 0 \\ y_3 \in \mathbb{R} \end{cases}$$

$$\begin{cases} z \to \min \\ y_2 = \frac{1}{2}\left[z - 3y_1 - 4y_3\right] \geqslant 0 \\ 2z \geqslant 6 - z \\ 8z - 16 \geqslant 30 - 5z \end{cases} \longrightarrow \begin{cases} \min(3y_1 + 2y_2 + 4y_3) = {}^{46}\!/_{13} \\ y_1 = 0 \\ y_2 = {}^{15}\!/_{13} \\ y_3 = {}^{4}\!/_{13} \end{cases}$$

Заметим, что целевая функция и ограничения являются линейными, а из курса дискретной математики мы знаем, что в таких случаях при достижении оптимальных решений зазор между двойственной и прямой задачей равен 0.

Задача 5. Сначала составим двойственную задачу:

$$\begin{cases} 16y_1 + 4y_2 \to \min \\ 4y_1 + 6y_2 \geqslant 5 \\ -4y_2 \geqslant 1 \\ y_1 - y_2 \geqslant 1 \\ 3y_1 + y_2 \geqslant 1 \\ y_1, y_2 \in \mathbb{R} \end{cases}$$

Обозначим $z = 4y_1 + y_2$, откуда $y_2 = z - 4y_1$, тогда:

$$\begin{cases} z \to \min \\ 6z - 20y_1 \geqslant 5 \\ 16y_1 - 4z \geqslant 1 \\ 5y_1 - z \geqslant 1 \\ z - y_1 \geqslant 1 \\ y_1 \in \mathbb{R} \end{cases} \longrightarrow \begin{cases} z \to \min \\ 24z - 20 \geqslant 20z + 5 \\ 6z - 5 \geqslant 4z + 4 \\ 16z - 16 \geqslant 4z + 1 \\ 5z - 5 \geqslant z + 1 \end{cases} \longrightarrow \begin{cases} \min(16y_1 + 4y_2) = 25 \\ y_1 = {}^{13}\!/_8 \\ y_2 = -{}^{1}\!/_4 \end{cases}$$

Вернёмся к решению исходной задачи:

$$\begin{cases} \max(5x_1 + x_2 + x_3 + x_4) = 25 \\ 4x_1 + x_3 + 3x_4 = 16 \\ 6x_1 - 4x_2 - x_3 + x_4 = 4 \\ x_1, x_2, x_3, x_4 \geqslant 0 \end{cases} \longrightarrow \begin{cases} x_1 = \frac{1}{4}[-x_3 - 3x_4 + 16] \\ x_2 = \frac{1}{16}[-10x_3 - 14x_4 + 80] \\ 7x_3 + 29x_4 = 0 \\ x_1, x_2, x_3, x_4 \geqslant 0 \end{cases} \longrightarrow \begin{cases} x_1 = 4 \\ x_2 = 5 \\ x_3 = 0 \\ x_4 = 0 \end{cases}$$