Методы оптимизации, HW5

Иванов Вячеслав, группа 699

29 октября 2018 г.

Оглавление

1 Задача 1

$P \in \mathbb{S}^n_+$, $\operatorname{rank}(A) = m < n$

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \frac{1}{2}\mathbf{x}^\top \mathbf{P}\mathbf{x}^\top + \mathbf{q}^\top\mathbf{x} + r \\ \text{s. t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

Условия ККТ в матричной записи:

$$\mathbf{K} := \begin{bmatrix} P & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \mathbf{x}_* \\ \lambda_* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \tag{1}$$

Доказать эквивалентность следующих условий невырожденности матрицы системы (1):

a) $\mathcal{N}(A) \cap \mathcal{N}(P) = 0$, где $\mathcal{N}(A)$ — ядро матрицы.

b) $Ax = 0,\ x \neq 0 \to x^T P x > 0$

c) $F^\top P F \succ 0$, где $F \in \mathbb{R}^{n\times(n-m)},\ \mathcal{R}(F) = \mathcal{N}(A),\ \mathcal{R}(F)$ — образ F (и пр-во решений $Ax = 0$).

d) $P + A^\top Q A \succ 0$ для некоторой $Q \succeq 0$

Доказательство. Докажем вначале попарную эквивалентность условий между собой.

- $(a) \iff (b)$

  1. $\neg(a) \implies \neg(b)$:

  $$\begin{aligned} &\exists x \neq 0 : x \in \mathcal{N}(A) \cap \mathcal{N}(P) \\ &\implies Ax = 0,\ Px = 0 \\ &\implies x^\top P x = 0 \end{aligned}$$

  2. $\neg(b) \implies \neg(a)$:

  $$\begin{aligned} &\exists x : Ax = 0,\ x \neq 0,\ x^\top P x = 0 \implies Px = 0 \\ &\implies x \neq 0 :\ x \in \mathcal{N}(A) \cap \mathcal{N}(P) \\ &\implies \mathcal{N}(A) \cap \mathcal{N}(P) \neq \{0\} \end{aligned}$$

- $(b) \iff (c)$

  1. $\neg(b) \implies \neg(c)$:

  $$\begin{aligned} &\exists x \in \mathcal{N}(A) :\ x \neq 0,\ x^\top P x = 0 \\ &\implies \exists y :\ x = Fy \ (\text{т.к. } \mathcal{N}(A) = \mathcal{P}(F)) \\ &\implies (Fy)^\top P (Fy) = 0 \\ &\implies y^\top (F^\top P F) y = 0 \\ &\implies F^\top P F \not\succ 0 \end{aligned}$$

  2. $\neg(c) \implies \neg(b)$:

  $$\begin{aligned} &F^\top P F \not\succ 0 \\ &\implies \exists y \neq 0 : y^\top (F^\top P F) y \leq 0 \\ &\implies (Fy)^\top P (Fy) \leq 0 \\ &\implies \exists x = Fy :\ x \in \mathcal{N}(A),\ x \neq 0,\ x^T P x \leqslant 0 \end{aligned}$$

- $(c) \iff (d)$:
  1. $(c) \implies (d)$:
$$\begin{aligned}
&(c) \iff (b) \\
&\quad \implies \forall x: \ x^\top (P + A^\top A)x = x^\top P x + x^\top A^\top A x = (Ax)^\top (Ax) + x^\top P x > 0 \\
&\quad \implies \text{при } Q = E \text{ верно } (d)
\end{aligned}$$
  2. $(d) \implies (c)$:
$$\begin{aligned}
P + A^T Q A \not\succ 0 &\iff \forall x : x^\top P x + (Ax)^T Q (Ax) > 0 \\
&\implies \forall x \in \mathcal{N}(A) : x^\top P x > 0 \\
&\implies F^T P F \succ 0
\end{aligned}$$
- $(a) - (d) \iff \det \mathbf{K} \neq 0$:
  1. $(\neg(a) \vee \ldots \vee \neg(d)) \implies \det \mathbf{K} = 0$:
$$\begin{aligned}
&(\neg(a) \vee \ldots \vee \neg(d)) \\
&\implies \neg(a) \\
&\implies \exists x \neq 0 : \ x \in \mathcal{N}(A) \cap \mathcal{N}(P) \\
&\implies \begin{pmatrix} P & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} Px \\ Ax \end{pmatrix} = 0 \\
&\implies \det \mathbf{K} = 0
\end{aligned}$$
  2. $\det \mathbf{K} = 0 \implies (\neg(a) \vee \ldots \vee \neg(d))$:
$$\begin{aligned}
&\det \mathbf{K} = 0 \\
&\iff \exists (x, \lambda) \neq \mathbf{0} : \begin{bmatrix} P & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
&\iff \begin{bmatrix} Px + A^\top \lambda \\ Ax + 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
&\implies \begin{cases} Ax = 0, \\ Px + A^\top \lambda = 0 \quad (*) \end{cases} \Bigg| \ x^\top \cdot \\
&\iff \begin{cases} Ax = 0, \\ x^\top P x + (Ax)^\top \lambda = 0 \end{cases} \\
&\iff x^\top P x = 0 \\
&\iff Px = 0 \ (\text{т.к. } P \in S^n_+) \\
&\implies A^T \lambda = 0 \ (\text{из } (*)) \\
&\implies \left[\begin{array}{l} x \neq 0 \implies \neg(b), \ \text{т.к. Ax = b} \\ x = 0 \implies \lambda \neq 0, \ \text{т.к. } (x, \lambda) \neq 0 \\ \qquad\quad \implies A^T \lambda = 0, \ \text{что противоречит полноте ранга } A \end{array} \right.
\end{aligned}$$

$\square$

2 Задача 2

Решить, построить двойственную задачу, проиллюстрировать двойственную функцию Лагранжа для нескольких значений параметра, проверить задачу на сильную двойственность.

$$\begin{aligned} &\min_{x \in \mathbb{R}} x^2 + 1 \\ &\text{s.t. } (x-2)(x-4) \le 0 \end{aligned}$$

Решение.

1. Выпишем Лагранжиан и запишем условия ККТ:

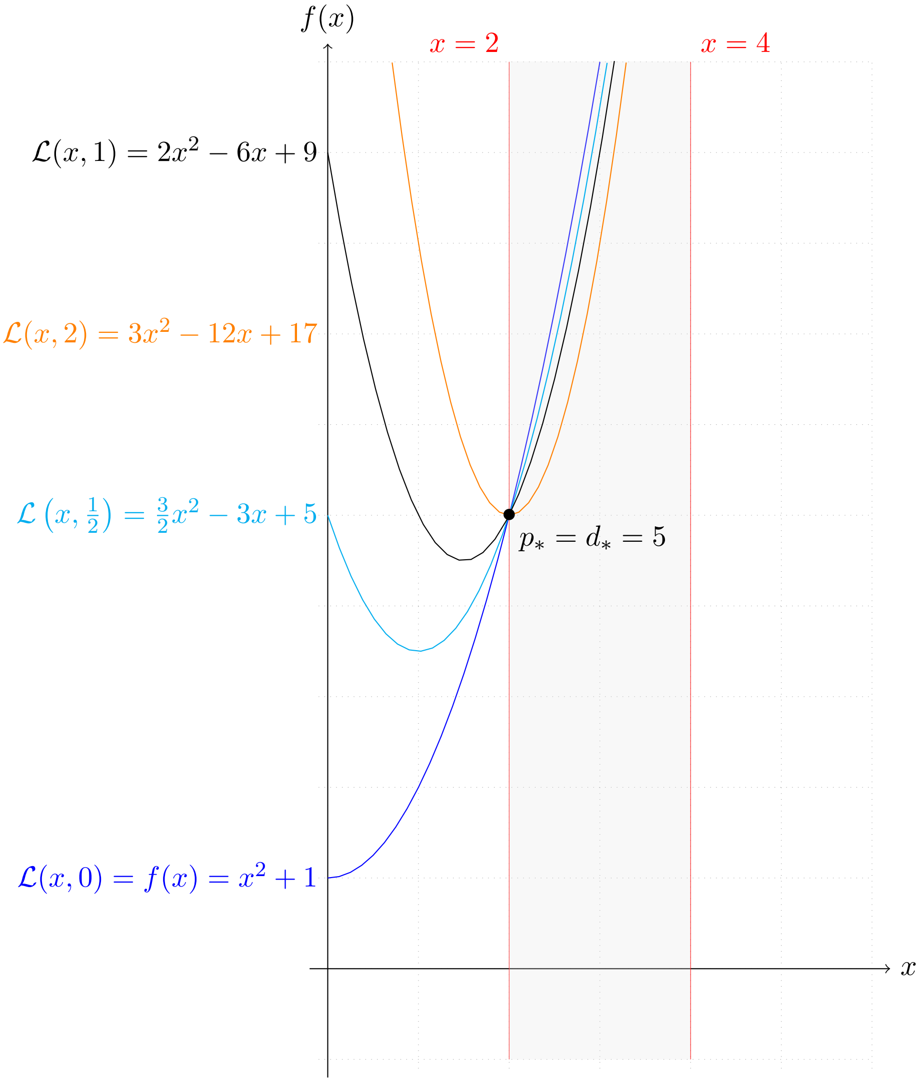
$$\begin{aligned} \mathcal{L}(x,\mu) &= x^2 + 1 + \mu(x-2)(x-4) = \\ &= x^2 + 1 + \mu(x^2 - 6x + 8) \\ \nabla_x \mathcal{L}(x,\mu) &= 2x + 2x\mu - 6\mu = \\ &= 2x(\mu+1) - 6\mu \end{aligned}$$

$$\begin{cases} (x-2)(x-4) \le 0 \\ \mu \ge 0 \\ \mu(x-2)(x-4) = 0 \\ x(\mu-1) - 3\mu = 0 \end{cases} \iff \left[\begin{array}{l} \begin{cases} \mu = 0, \\ x = 0, \\ (x-2)(x-4) \le 0 \end{cases} \quad - \perp \\ \begin{cases} \mu > 0, \\ (x-2)(x-4) = 0, \\ \mu = \frac{x}{3-x} \end{cases} \end{array}\right. \iff \left[\begin{array}{l} (1) \begin{cases} x = 2, \\ \mu = 2 \end{cases} \\ (2) \begin{cases} \mu > 0, \\ x = 4, \\ \mu = -4 \end{cases} \quad - \perp \end{array}\right.$$

Где $\perp$ означает "система несовместна".
Получаем, что условия ККТ выполняются при $x = \mu = 2$.

2. Найдём оптимальное значение и изобразим его на графике:

$$p^* = f(x^*) = f(2) = 5 \ge \inf_x L(x,\lambda)$$



3. Двойственная функция Лагранжа:

$$g(\mu) = \inf_x \left(x^2 + 1 + \mu(x-2)(x-4)\right) = inf_x(x^2 + 1 + \mu(x^2 - 6x + 8))$$

Двойственная задача:

$$\max_{\mu \in \mathbb{R}} g(\mu)$$
$$\text{s.t. } \mu \geq 0$$

Найдём стационарные точки:

$$\nabla_x \mathcal{L}(x,\mu) = 2x(\mu+1) - 6\mu = 0 \implies x = \frac{3\mu}{\mu+1}$$
$$\nabla_x^2 \mathcal{L}(x,\mu) = 2(\mu+1) > 0 \text{ при } \mu \geq 0$$
$$x^* = \frac{3\mu}{\mu+1} \text{ в силу выпуклости}$$

$$\begin{aligned} g(\mu) = \mathcal{L}(x_*,\mu) = \\ &= \frac{9\mu^2}{\mu+1} - \frac{18\mu^2}{\mu+1} + 8\mu + 1 = \\ &= -\frac{9\mu^2}{\mu+1} + 8\mu + 1 \end{aligned}$$

Явный вид двойственной задачи:

$$\begin{aligned}
&\max_{\mu\in\mathbb{R}} -\frac{9\mu^2}{\mu+1}+8\mu+1\\
&\text{s.t. } \mu \geq 0
\end{aligned}$$

Её тоже будем решать, используя условия ККТ. Найдём стационарные точки:

$$\begin{aligned}
\mathcal{L}(\mu,\nu) &= -\frac{9\mu^2}{\mu+1}+\mu+1-\mu\nu\\
\nabla_\mu\mathcal{L}(\mu,\nu) &= -9\frac{2\mu(\mu+1)-\mu^2}{(\mu+1)^2}+8-\nu=\\
&= \frac{-\mu^2-2\mu+8}{(\mu+1)^2}-\mu=\frac{9-(\mu+1)^2}{(\mu+1)^2}-\nu=\\
&= \frac{(2-\mu)(\mu+4)}{(\mu+1)^2}-\nu=0
\end{aligned}$$

Запишем их вместе с условиями дополняющей жёсткости:

$$\begin{aligned}
&\nu_*=\frac{(2-\mu)(\mu+4)}{(\mu+1)^2}\\
&\nu_*(-\mu_*)=0
\end{aligned}$$

$$\begin{cases}\nu_*=0 \implies \mu_*=2,\ g(\mu_*)=d^*=5=p^*\\ \mu_*=0 \implies g(\mu_*)=1<5\end{cases} \implies \mu_*=2$$

Т.к. $d^*=p^*$, то сильная двойственность выполняется.

$\square$

3 Задача 3

Найти двойственную задачу к задаче бинарного линейного программирования:

$$\begin{aligned}
&\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}\\
&\text{s.t. } x_i(1-x_i)=0,\ i=1,\dots,n\\
&\mathbf{Ax}\leqslant\mathbf{b}
\end{aligned}\tag{2}$$

Метод поиска приближённых решений в задачах дискретной оптимизации, основанный на построении двойственной задачи уже в непрерывном пространстве называется релаксацией Лагранжа. Доказать, что нижняя оценка, которую даёт релаксация Лагранжа совпадает с оценкой, которую даёт решение непрерывной релаксации исходной задачи

$$\begin{aligned}
&\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}\\
&\text{s.t. } 0\leqslant x_i\leqslant 1,\ i=1,\dots,n\\
&\mathbf{Ax}\leqslant\mathbf{b}
\end{aligned}\tag{3}$$

Доказать, что решение этой задачи даёт оценку снизу на решение исходной.

Решение. Выпишем вначале двойственную задачу по определению:

$$\begin{aligned}&\max_{\mu}\ \inf_{\mathbf{x}}\left(\mathbf{c}^T\mathbf{x}+\mu^T(\mathbf{A}\mathbf{x}-\mathbf{b})+\sum_{i=1}^{n}\lambda_i x_i(x_i-1)\right)^T\\&\text{s.t. }\mu\succeq 0\end{aligned}\tag{4}$$

Заметим, что её можно упростить, переписав прямую задачу в эквивалентном виде:

$$\begin{aligned}&\min_{\mathbf{x}\in\{0,1\}^n}\mathbf{c}^T\mathbf{x}\\&\text{s.t. }\mathbf{A}\mathbf{x}\leqslant\mathbf{b}\end{aligned}$$

Выпишем Лагранжиан:

$$\begin{aligned}\mathcal{L}(\mathbf{x},\mu)&=\mathbf{c}^T\mathbf{x}+\mu^T(\mathbf{A}\mathbf{x}-\mathbf{b})=\\&=(\mathbf{c}+\mathbf{A}^T\mu)^T\mathbf{x}-\mu^T\mathbf{b}\end{aligned}$$

В силу его выпуклости, двойственная функция Лагранжа запишется так:

$$\begin{aligned}g(\mu)&=\inf_{\mathbf{x}\in\{0,1\}^n}\mathcal{L}(\mathbf{x},\mu)\\&=\inf_{\mathbf{x}\in\{0,1\}^n}(\mathbf{c}+\mathbf{A}^T\mu)^T\mathbf{x}-\mu^T\mathbf{b}=\\&=-\mu^T\mathbf{b}+\sum_{i=1}^{n}\min\{0,c_i+a_i^T\mu\}\end{aligned}$$

Итого двойственная задача запишется так:

$$\begin{aligned}&\max_{\mu}-\mu^T\mathbf{b}+\sum_{i=1}^{n}\min\{0,c_i+a_i^T\mu\}\\&\text{s.t. }\mu\succeq 0\end{aligned}\tag{5}$$

Далее, рассмотрим вторую задачу из условия:

$$\begin{aligned}&\min_{\mathbf{x}}\mathbf{c}^T\mathbf{x}\\&\text{s.t. }0\leqslant x_i\leqslant 1,\ i=1,\dots,n\\&\mathbf{A}\mathbf{x}\leqslant\mathbf{b}\end{aligned}$$

Перепишем её в эквивалентном виде:

$$\begin{aligned}&\min_{\mathbf{x}}\mathbf{c}^T\mathbf{x}\\&\text{s.t. }x_i(x_i-1)\leqslant 0,\ i=1,\dots,n\\&\mathbf{A}\mathbf{x}\leqslant\mathbf{b}\end{aligned}\tag{6}$$

Отсюда видно, что двойственная задача имеет ту же структуру, что и у релаксации Лагранжа:

$$\begin{aligned}&\max_{\mu}\ \inf_{\mathbf{x}}\left(\mathbf{c}^T\mathbf{x}+\mu^T(\mathbf{A}\mathbf{x}-\mathbf{b})+\sum_{i=1}^{n}\nu_i x_i(x_i-1)\right)^T\\&\text{s.t. }\mu\succeq 0,\ \nu\succeq 0\end{aligned}\tag{7}$$

Добавилось только условие $\nu\succeq 0$. Тем не менее, при $\lambda_i<0$ в релаксации Лагранжа получаем $\inf g(\mu,\lambda)=-\infty$ в силу вогнутости целевой функции (видно по Гессиану). Следовательно, релаксация Лагранжа (3) эквивалентна двойственной задаче непрерывной релаксации (6).

Утверждение задачи следовало бы из выполнения условий Слейтера. Релаксация Лагранжа, очевидно, выпуклая, т.к. целевая функция линейна, а ограничения задают пересечение выпуклых множеств: $[0,1]^n$ и $\{\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}$.
Осталось доказать, что для непрерывной релаксации выполнена сильная двойственность. Тем не менее, я в это не особо верю и пока не смог ни оправдать, ни опровергнуть. $\square$

4 Задача 4

Рассмотреть проблему наименьших квадратов при линейных ограничениях:

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$
$$\text{s.t. } \mathbf{G}\mathbf{x} = \mathbf{h}$$

где $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\operatorname{rank} \mathbf{A} = n$, $\mathbf{G} \in \mathbb{R}^{p\times n}$, $\operatorname{rank} \mathbf{G} = p$, $p \leqslant n \leqslant m$. Получить выражение для её решения и решения двойственной задачи. Проверить, выполняется ли сильная двойственность.

Решение. Выпишем Лагранжиан:

$$\mathcal{L}(\mathbf{x}, \lambda) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda^T(\mathbf{G}\mathbf{x} - \mathbf{h})$$

На семинарах мы вывели, что:

$$\nabla_{\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \mathbf{A}^T\mathbf{x}$$
$$\nabla_{\mathbf{x}}(\mathbf{a}^T\mathbf{x}) = \mathbf{a}$$

Поскольку

$$\begin{aligned}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 &= (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}^T\mathbf{b} + \mathbf{b}^T\mathbf{b}\end{aligned}$$

Получаем

$$\mathcal{L}(\mathbf{x}, \lambda) = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} + (G^T\lambda - 2\mathbf{A}^T\mathbf{b})^T\mathbf{x} + \mathbf{b}^T\mathbf{b} - \lambda^T\mathbf{h}$$

Откуда, в силу того, что $\nabla_{\mathbf{x}}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{b}$, имеем:

$$\begin{aligned}\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \lambda) = 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{b} + \mathbf{G}^T\lambda = 0 \\ 2\mathbf{A}^T\mathbf{A}\mathbf{x} = 2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda \\ \mathbf{x}_* = \frac{1}{2}(\mathbf{A}^T\mathbf{A})^{-1}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda)\end{aligned}$$

Лагранжиан выпуклый как сумма выпуклых, следовательно найденная точка — минимум.
Запишем теперь двойственную функцию Лагранжа подстановкой оптимального значения:

$$\begin{aligned}g(\lambda) = &\left(\frac{1}{2}(\mathbf{A}^T\mathbf{A})^{-1}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda)\right)^T \mathbf{A}^T\mathbf{A}\left(\frac{1}{2}(\mathbf{A}^T\mathbf{A})^{-1}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda)\right) \\ &+ (G^T\lambda - 2\mathbf{A}^T\mathbf{b})^T\left(\frac{1}{2}(\mathbf{A}^T\mathbf{A})^{-1}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda)\right) + \mathbf{b}^T\mathbf{b} - \lambda^T\mathbf{h} = \\ &= \frac{1}{4}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda)^T(\mathbf{A}^T\mathbf{A})^{-1}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda) + \\ &- \frac{1}{2}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda)^T(\mathbf{A}^T\mathbf{A})^{-1}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda) + \\ &+ \mathbf{b}^T\mathbf{b} - \lambda^T\mathbf{h} = \\ &= -\frac{1}{4}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda)^T(\mathbf{A}^T\mathbf{A})^{-1}(2\mathbf{A}^T\mathbf{b} - \mathbf{G}^T\lambda) + \mathbf{b}^T\mathbf{b} - \lambda^T\mathbf{h} \\ &= -\left(\frac{1}{2}\lambda^T\mathbf{G} - \mathbf{b}^T\mathbf{A}\right)(\mathbf{A}^T\mathbf{A})^{-1}\left(\frac{1}{2}\mathbf{G}^T\lambda - \mathbf{A}^T\mathbf{b}\right) + \mathbf{b}^T\mathbf{b} - \lambda^T\mathbf{h}\end{aligned}$$

Заметим, что решение прямой задачи зависит от неизвестного параметра λ.
Упростим двойственный Лагранжиан:

$$\begin{aligned}
g(\lambda) &= -\frac{1}{4}\lambda^T \left(\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{G}^T\right)\lambda + \mathbf{b}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{G}^T\lambda + \mathbf{b}^T\mathbf{b} - \lambda^T\mathbf{h} \\
\nabla g(\lambda) &= -\frac{1}{2}\left(\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{G}^T\right)\lambda + \mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} - \mathbf{h} = 0 \\
\lambda_* &= 2\left(\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{G}^T\right)^{-1}\left(\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} - \mathbf{h}\right)
\end{aligned}$$

Подставляя полученное значение λ_* получим выражение для $\mathbf{x}_*$:

$$\mathbf{x}_* = \frac{1}{2}(\mathbf{A}^T\mathbf{A})^{-1}(2\mathbf{A}^T\mathbf{b} - 2\mathbf{G}^T\left(\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{G}^T\right)^{-1}\left(\mathbf{G}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} - \mathbf{h}\right))$$

Условия Слейтера выполнены, т.к. исходная задача выпуклая, ни в ней, ни в двойственной к ней задаче нет условий типа неравенства. □