Problem 1 (20 Points)

In this True or False problem, no explanation is required for your answer.
If you answer correctly, you will earn 2 points.
However, if your answer is incorrect, you will lose 1 point.
If you choose not to answer at all, you will receive no points.

1. $f(x) = (\sum_{i=1}^{n} x_i^{\frac{1}{2}})^2$ is convex on $\mathbb{R}_{++}^n$.
   - **False**
   - **反例:**
     考虑 $n=2$ 的情况，$f(x) = x_1 + x_2 + \sqrt{x_1 x_2}$
     取 $\begin{cases} x_1 = (1,9) \\ x_2 = (9,1) \end{cases}$，考虑其连线中点 $\frac{1}{2}x_1 + \frac{1}{2}x_2 = (5,5)$:
     $$\begin{aligned} f(\frac{1}{2}x_1 + \frac{1}{2}x_2) - [\frac{1}{2}f(x_1) + \frac{1}{2}f(x_2)] &= f(5,5) - \frac{1}{2}f(1,9) - \frac{1}{2}f(9,1) \\ &= 15 - \frac{13}{2} - \frac{13}{2} \\ &= 2 > 0 \end{aligned}$$
     显然不满足凸性 (凸性要求凸组合的函数值 $\leq$ 函数值的凸组合)
   - **证明:**
     我们记 $x \in \mathbb{R}_{++}^n$ 的逐元素 a 次幂运算的结果为 x^a
     则 $f(x) = (\sum_{i=1}^{n} x_i^{\frac{1}{2}})^2 = (1_n^T x^{\frac{1}{2}})^2 = (x^{\frac{1}{2}})^T (1_n 1_n^T)(x^{\frac{1}{2}})$
     对于任意 $x \in \mathbb{R}_{++}^n$ 计算梯度得:
     $$\begin{aligned} \nabla_x f(x) &= \nabla_x \{x^{\frac{1}{2}}\} \cdot 2(1_n 1_n^T) x^{\frac{1}{2}} \\ &= \text{diag}(\frac{1}{2}x^{-\frac{1}{2}}) \cdot 2(1_n 1_n^T) x^{\frac{1}{2}} \\ &= x^{-\frac{1}{2}} 1_n^T x^{\frac{1}{2}} \\ &= (1_n^T x^{\frac{1}{2}}) x^{-\frac{1}{2}} \end{aligned}$$
     对于任意 $x \in \mathbb{R}_{++}^n$ 计算 Hessian 矩阵得:
     $$\begin{aligned} \nabla_{xx}^2 f(x) &= \nabla_x \{x^{\frac{1}{2}}\} \cdot 1_n (x^{-\frac{1}{2}})^T + (1_n^T x^{\frac{1}{2}}) \cdot \nabla_x \{x^{-\frac{1}{2}}\} \\ &= \text{diag}(\frac{1}{2}x^{-\frac{1}{2}}) \cdot 1_n (x^{-\frac{1}{2}})^T + (1_n^T x^{\frac{1}{2}}) \cdot \text{diag}(-\frac{1}{2}x^{-\frac{3}{2}}) \\ &= \frac{1}{2}x^{-\frac{1}{2}}(x^{-\frac{1}{2}})^T - \frac{1}{2}x^{\frac{1}{2}}(x^{-\frac{3}{2}})^T \end{aligned}$$
     - 对于 $n=1$ 的情况，$f''(x) \equiv 0$，此时 f 在 $\mathbb{R}_{++}^n$ 上为仿射函数，自然是凸的.
     - 对于 $n \geq 2$ 的情况，我们注意到:
       对于任意 $i = 1, \ldots, n$，$\nabla_{xx}^2 f(x)$ 的第 i 个对角元为 $\frac{1}{2}(x_i^{-\frac{1}{2}} x_i^{-\frac{1}{2}} - x^{\frac{1}{2}} x^{-\frac{3}{2}}) = 0$
       因此要使得 $\nabla_{xx}^2 f(x) \succeq 0$，其所有非对角元都必须为 0.

       对于任意 $i \neq j \in \{1, \ldots, n\}$，$\nabla_{xx}^2 f(x)$ 在 (i,j) 位置的非对角元为:
       $\nabla_{xx}^2 f(x)_{(i,j)} = \frac{1}{2}(x_i^{-\frac{1}{2}} x_j^{-\frac{1}{2}} - x_i^{\frac{1}{2}} x_j^{-\frac{3}{2}})$

       显然并不是所有的 $x \in \mathbb{R}_{+}^n$ 都能保证 $\nabla_{xx}^2 f(x)$ 的非对角元均为 0,
       因而不能保证 $\nabla_{xx}^2 f(x)$ 半正定,
       所以在 $n \geq 2$ 的情况下 f 并不是 $\mathbb{R}_{++}^n$ 上的凸函数.

2. Consider the constrained optimization problem：
   $$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \\ & x \in \mathcal{X} \end{aligned}$$

Let $(x_\star, \lambda_\star, \mu_\star)$ be a saddle point.

Then for all $\begin{cases} x \in \mathcal{X} \\ \lambda \in \mathbb{R}^m_+ \\ \nu \in \mathbb{R}^p \end{cases}$ it holds that $L(x_\star, \lambda, \mu) \le L(x, \lambda_\star, \mu_\star)$.

- **True**
- **(江如俊教授 Note 5 定义 1.3)**

  考虑一般优化问题:

  $$(\mathrm{P}) \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \le 0,\ i = 1, \dots, m \\ & h_i(x) = 0,\ i = 1, \cdots, p \\ & x \in \mathcal{X} \end{array}$$

  定义其 Lagrange 函数为:

  $$\begin{cases} L(x, \lambda, \nu) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{i=1}^p \nu_i h_i(x) = f(x) + \lambda^T g(x) + \nu^T h(x) \\ \mathrm{dom}(L) = \mathcal{X} \times \mathbb{R}^m_+ \times \mathbb{R}^p \end{cases}$$

  鞍点的定义:

  若 $(x_\star, \lambda_\star, \nu_\star) \in \mathrm{dom}(L)$ 满足:

  对于任意 $x \in \mathcal{X}$ 和 $(\lambda, \nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$ 都有 $L(x_\star, \lambda, \nu) \le L(x_\star, \lambda_\star, \nu_\star) \le L(x, \lambda_\star, \nu_\star)$ 成立,

  即 $\begin{cases} x_\star = \arg \max\limits_{(\lambda,\nu) \in \mathbb{R}^m_+ \times \mathbb{R}^p} L(x, \lambda_\star, \nu_\star) \\ (\lambda_\star, \nu_\star) = \arg \min\limits_{x \in \mathcal{X}} L(x_\star, \lambda, \nu) \end{cases}$ ($x_\star$ 最小化 $L(x, \lambda_\star, \nu_\star)$ 而 $(\lambda_\star, \nu_\star)$ 最大化 $L(x_\star, \lambda, \nu)$)

  则我们称 $(x_\star, \lambda_\star, \nu_\star) \in \mathrm{dom}(L)$ 是 $L(x, \lambda, \nu)$ 的一个**鞍点**.

3. Consider the unconstrained optimization problem $\min\limits_{x \in \mathbb{R}^n} f(x)$

   If $\begin{cases} \nabla f(\bar{x}) = 0_n \\ \nabla^2 f(\bar{x}) \succeq 0 \end{cases}$, then $\bar{x}$ is a local minimum.

   - **False**
   - 这个命题中的条件只是 (二阶连续可微) 无约束优化问题的二阶必要条件，并不充分.

     考虑反例 $f(x) = x^5$, $x_0 = 0$ 满足 $\begin{cases} f'(x_0) = 0 \\ f''(x_0) = 0 \end{cases}$ 但它不是问题的局部最小点 (实际上是拐点).
   - **(江如俊教授 Note 4 命题 2.5)**

     定理 4.1.5：无约束局部最小点的二阶必要条件 (Second Order Necessary Condition)

     设 $f: \mathbb{R}^n \to \mathbb{R}$ 在 $x_0 \in \mathrm{dom}(f)$ 处**二阶连续可微** (twice continuously differentiable)

     若 x_0 是 f 的一个**局部最小点**,

     即 $f(x_0)$ 是 x_0 某邻域内目标函数的最小值,

     即存在 $\varepsilon > 0$ 使得 $f(x_0) = \inf\limits_{x \in B(x_0, \varepsilon)} f(x)$ (其中Euclid球 $B(x_0, \varepsilon)$ 代表 x_0 的**邻域**)

     则有 $\begin{cases} \nabla f(x_0) = 0_n \\ \nabla^2 f(x_0) \succeq 0 \end{cases}$ 成立 (即 x_0 为**非严格凸的驻点**)

4. The positive semi-definite cone $\mathbb{S}^n_+$ is a self-dual cone.

   - **True**
   - **(江如俊教授 Note 5-2024 2.1 节末尾)**

5. If $f(x)$ is convex, so is $g(x) = (c^T x + d) f(\frac{Ax+b}{c^T x+d})$.

   Note that $\mathrm{dom}(g) = \{x : \begin{cases} c^T x + d > 0 \\ \frac{Ax+b}{c^T x+d} \in \mathrm{dom}(f) \end{cases}\}$

   - **True**

    - 参考这篇笔记 FDU 最优化方法 3. 凸函数 的 4.(3)(Ⅲ) 节，
      那里说明了凸函数 $f(x)$ 的线性分式透视 $g(x)=(c^Tx+d)f(\frac{Ax+b}{c^Tx+d})$ 也是凸函数.
6. A set is convex iff. its convex hull is convex.
    - **False**
    - This is bullshit.
      Whatever the set S is, its convex hull $\mathbf{conv}(S)$ is naturally convex,
      because it represents the smallest convex set that contains S.
7. The largest eigenvalue of symmetric matrix A is the optimal solution for the SDP problem:
   $$\begin{aligned}\min_{\lambda\in\mathbb{R}} \quad & \lambda \\ \text{s.t.} \quad & \lambda I\succeq A\end{aligned}$$
    - **True**
    - **(要素察觉：Prof. smy & Advanced Linear Algebra)**
8. The Lagrange duals of the standard QCQP problem and its SDP relaxation are not equivalent.
    - **False**
    - **(江如俊教授 Note 5 2.3.1节)**
9. Let $f(x)=|\log(a^Tx)|$, then $g(x)=\exp\{f(x)\}$ is convex on $\{x|a^Tx>0\}$.
    - **True**
    - 注意到 $f(x)=|\log(a^Tx)|=\begin{cases}-\log(a^Tx), & \text{if } 0<a^Tx\le 1 \\ \log(a^Tx), & \text{if } a^Tx>1\end{cases}$
      因此 $g(x)=\exp\{f(x)\}=\begin{cases}\frac{1}{a^Tx}, & \text{if } 0<a^Tx\le 1 \\ \log(a^Tx), & \text{if } a^Tx>1\end{cases}$
      引入映射 $\begin{cases}h(u)=\begin{cases}\frac{1}{u}, & \text{if } 0<u\le 1 \\ \log(u), & \text{if } u>1\end{cases} \\ p(x)=a^Tx\end{cases}$ 则 $g(x)=h(p(x))$
      根据 FDU 最优化方法 3. 凸函数 3.(2) 中有关仿射函数参与函数复合的相关结论，
      函数 $g(x)$ 的凸/凹性与 $h(u)$ 的凸/凹性一致.
      而我们将 $h(u)$ 的图像画出来，很容易知道 $h(u)$ 是凸函数，
      因此 $g(x)$ 也是凸函数.
10. $f(x)$ is a convex function iff. $\text{epi}(f)$ is a convex set.
    - **True**
    - 上境图的基本性质参见 FDU 最优化方法 3. 凸函数 1.(4) 节

Problem 2 (15 Points)

Derive the dual of the following **relative entropy** problem:

$$(\text{P})\quad \begin{aligned}\min_{x\in\mathbb{R}_+^n} \quad & \sum_{i=1}^n x_i\log(\frac{x_i}{y_i}) \\ \text{s.t.} \quad & Ax=b \\ & 1_n^Tx=1\end{aligned}$$

where $y\in\mathbb{R}_{++}^n$ and $b\in\mathbb{R}^m$ are given.
Denote a_i as the i-th column of $A\in\mathbb{R}^{m\times n}$.

Solution:

分别为约束 $\begin{cases} Ax = b \\ 1_n^T x = 1 \end{cases}$ 引入 Lagrange 乘子 $\begin{cases} \nu \in \mathbb{R}^m \\ \mu \in \mathbb{R} \end{cases}$

定义 **Lagrange 函数**为:

$$\begin{aligned} L(x,\nu,\mu) &= \sum_{i=1}^n x_i \log(\frac{x_i}{y_i}) + \nu^T(b - Ax) + \mu(1 - 1_n^T x) \\ &= x^T \log(x) - \log(y)^T x - \nu^T Ax - \mu 1_n^T x + b^T \nu + \mu \end{aligned}$$

注意到对于任意 $i = 1, \dots, n$，负熵函数 $x_i \log(x_i)$ 是关于 $x \in \mathbb{R}_+^n$ 的凸函数，
(为严谨起见，我们额外定义 $0 \log(0) = 0$)
因此 $L(x,\nu,\mu)$ 关于 $x \in \mathbb{R}_+^n$ 是凸函数，其驻点 (即为 $\bar{x}$) 一定能够最小化 $L(x,\nu,\mu)$.

求解驻点条件 $\nabla_x L(x,\nu,\mu) = 1_n + \log(x) - \log(y) - A^T\nu - \mu 1_n = 0_n$
我们有 $\begin{cases} \log(\bar{x}) - \log(y) = A^T\nu + \mu 1_n - 1_n \\ \bar{x} = y \circ \exp\{A^T\nu + \mu 1_n - 1_n\} \end{cases}$ 成立.
其中 $\circ$ 代表向量的逐元素乘积 (Hadamard 乘积)

定义 **Lagrange 对偶函数**为:

$$\begin{aligned} d_1(\nu,\mu) &= \inf_{x \in \mathbb{R}_+^n} L(x,\nu,\mu) \\ &= L(\bar{x},\nu,\mu) \\ &= \bar{x}^T(\log(\bar{x}) - \log(y)) - \nu^T A\bar{x} - \mu 1_n^T \bar{x} + b^T\nu + \mu \\ &= \bar{x}^T(A^T\nu + \mu 1_n - 1_n) - \nu^T A\bar{x} - \mu 1_n^T \bar{x} + b^T\nu + \mu \\ &= -1_n^T \bar{x} + b^T\nu + \mu \\ &= -1_n^T(y \circ \exp\{A^T\nu + \mu 1_n - 1_n\}) + b^T\nu + \mu \\ &= -\sum_{i=1}^m y_i e^{a_i^T\nu + \mu - 1} + b^T\nu + \mu \\ &= -e^{\mu-1} \cdot \sum_{i=1}^m y_i e^{a_i^T\nu} + b^T\nu + \mu \end{aligned}$$

因此**对偶问题**为:

$$(\mathrm{D}) \quad \begin{array}{ll} \max & -e^{\mu-1} \cdot \sum_{i=1}^m y_i e^{a_i^T\nu} + b^T\nu + \mu \\ \text{s.t.} & \nu \in \mathbb{R}^m \\ & \mu \in \mathbb{R} \end{array}$$

注意到优化变量 ν, μ 是相互独立的，且 $d_1(\nu,\mu)$ 关于 μ 为凹函数，
因此我们可以单独在 $\mu \in \mathbb{R}$ 上最大化 (D) 的目标函数 $d_1(\nu,\mu)$ 来简化 (D):

由于 $d_1(\nu,\mu)$ 关于 μ 为凹函数，故其驻点 (记为 $\bar{\mu}$) 一定是最大值点.
求解驻点条件:

$$\begin{aligned} \frac{\partial}{\partial\mu} d_1(\nu,\mu) &= \frac{\partial}{\partial\mu}\{-e^{\mu-1} \cdot \sum_{i=1}^m y_i e^{a_i^T\nu} + b^T\nu + \mu\} \\ &= -e^{\mu-1} \cdot \sum_{i=1}^m y_i e^{a_i^T\nu} + 1 \\ &= 0 \end{aligned}$$

于是有 $\begin{cases} e^{\bar{\mu}-1} \cdot \sum_{i=1}^m y_i e^{a_i^T\nu} = 1 \\ \bar{\mu} = -\log(\sum_{i=1}^m y_i e^{a_i^T\nu}) + 1 \end{cases}$

我们定义**新的对偶函数**：

$$
\begin{aligned}
d_2(\nu) &= \sup_{\mu\in\mathbb{R}} d_1(\lambda,\mu) \\
&= d_1(\lambda,\bar{\mu}) \\
&= -e^{\bar{\mu}-1}\cdot\sum_{i=1}^{m} y_i e^{a_i^T\nu} + b^T\nu + \bar{\mu} \\
&= -1 + b^T\nu - \log(\sum_{i=1}^{m} y_i e^{a_i^T\nu-1}) + 1 \\
&= b^T\nu - \log(\sum_{i=1}^{m} y_i e^{a_i^T\nu-1})
\end{aligned}
$$

最终得到**对偶问题的化简形式**：

$$
(\text{simpleD})\begin{array}{ll}
\max & b^T\nu - \log(\sum_{i=1}^{m} y_i e^{a_i^T\nu-1}) \\
\text{s.t.} & \nu\in\mathbb{R}^m
\end{array}
$$

Problem 3 (15 Points)

考虑优化问题：

$$
(\text{P})\begin{array}{ll}
\min & \dfrac{\|Ax-b\|_1^2}{1-\|x\|_\infty} \\
\text{s.t.} & \|x\|_\infty \le \dfrac{1}{2}
\end{array}
$$

其中 $A\in\mathbb{R}^{m\times n}$ 和 $b\in\mathbb{R}^m$ 已经给定.

请将问题 (P) 转化为 SOCP (Second-Order Cone Program) 形式.
(即具有线性目标函数、线性约束和二阶锥约束)

Solution:

首先我们向问题 (P) 引入变量 $t\in\mathbb{R}$ 和约束 $\frac{\|Ax-b\|_1^2}{1-\|x\|_\infty}\le t$，得到：

$$
(\text{P1})\begin{array}{ll}
\min\limits_{x,t} & t \\
\text{s.t.} & \dfrac{\|Ax-b\|_1^2}{1-\|x\|_\infty}\le t \\
& \|x\|_\infty \le \dfrac{1}{2}
\end{array}
$$

考虑为 l_1 范数和 l_∞ 范数引入松弛变量 $\begin{cases} u\in\mathbb{R}^m \\ s\in\mathbb{R}\end{cases}$ 和约束 $\begin{cases} -u\preceq Ax-b\preceq u \\ -s1_n\preceq x\preceq s1_n\end{cases}$，
则我们可以将 $\begin{cases}\|Ax-b\|_1 \\ \|x\|_\infty\end{cases}$ 分别替换为它们的上界 $\begin{cases}1_m^T u \\ s\end{cases}$
于是我们得到：

$$
(\text{P2})\begin{array}{ll}
\min\limits_{x,t,u,s} & t \\
\text{s.t.} & (1_m^T u)^2 \le t(1-s) \\
& s\le \dfrac{1}{2} \\
& -u\preceq Ax-b\preceq u \\
& -s1_n\preceq x\preceq s1_n
\end{array}
$$

注意到 $(1_m^T u)^2 \le t(1-s)$ 实际上是一个**旋转的二阶锥约束** (rotated second-order constraint)
它可以等价变形为：
$(1_m^T u)^2 + \frac{1}{4}[t-(1-s)]^2 \le \frac{1}{4}[t+(1-s)]^2$
根据 $\begin{cases} t \ge 0 \\ 1-s \ge 0 \end{cases}$ 可知 $t+(1-s) \ge 0$，因此上述约束可等价变形为：
$\left\| \begin{bmatrix} 1_m^T u \\ \frac{1}{2}(t+s-1) \end{bmatrix} \right\|_2 \le \dfrac{1}{2}(t-s+1)$
即标准形式的二阶锥约束.

因此 $(\mathrm{P2})$ 即为问题 (P) 的 SOCP 形式.

Problem 4 (15 Points)

设 $f_i(x),\ i=1,\dots,m$ 为有限个 $\mathbb{R}^n \mapsto \mathbb{R}$ 的凸函数.
(1) 证明 $g(x) = \inf\limits_{z_1+\cdots+z_m=x} \{\sum\limits_{i=1}^m f_i(z_i)\}$ 是一个 $\mathbb{R} \mapsto \mathbb{R}$ 的凸函数. (5 Points)

Solution:
任意给定 $x_1, x_2 \in \mathbb{R}$
对于任意 $\alpha \in (0,1)$，我们都有：

$$\begin{aligned} g(\alpha x_1 + (1-\alpha)x_2) &= \inf_{z_1+\cdots+z_m=\alpha x_1+(1-\alpha)x_2} \{\sum_{i=1}^m f_i(z_i)\} \\ &= \inf\{\sum_{i=1}^m f_i(\alpha z_i^{(1)} + (1-\alpha)z_i^{(2)}) : \begin{cases} z_1^{(1)} + \cdots + z_m^{(1)} = x_1 \\ z_1^{(2)} + \cdots + z_m^{(2)} = x_2 \end{cases}\} \\ &\le \inf\{\sum_{i=1}^m [\alpha f_i(z_i^{(1)}) + (1-\alpha) f(z_i^{(2)})] : \begin{cases} z_1^{(1)} + \cdots + z_m^{(1)} = x_1 \\ z_1^{(2)} + \cdots + z_m^{(2)} = x_2 \end{cases}\} \quad (\text{利用 } f_i \text{ 的凸性}) \\ &= \inf\{\alpha \sum_{i=1}^m f_i(z_i^{(1)}) + (1-\alpha)\sum_{i=1}^m f(z_i^{(2)}) : \begin{cases} z_1^{(1)} + \cdots + z_m^{(1)} = x_1 \\ z_1^{(2)} + \cdots + z_m^{(2)} = x_2 \end{cases}\} \\ &= \alpha \inf_{z_1^{(1)}+\cdots+z_m^{(1)}=x_1} \{\sum_{i=1}^m f_i(z_i^{(1)})\} + (1-\alpha) \inf_{z_1^{(2)}+\cdots+z_m^{(2)}=x_2} \{\sum_{i=1}^m f_i(z_i^{(2)})\} \\ &= \alpha g(x_1) + (1-\alpha) g(x_2) \end{aligned}$$

因此 $g(x)$ 是凸函数.

(2) 证明 $g(x) = \inf\limits_{z_1+\cdots+z_m=x} \{\sum\limits_{i=1}^m f_i(z_i)\}$ 的共轭函数 $g^*(y) = \sum\limits_{i=1}^m f_i^*(y)$. (10 Points)

Solution:
任意给定 $y \in \mathbb{R}$，我们都有：

$$\begin{aligned}
g^*(y) &= \sup_{x\in\mathbb{R}}\{yx - g(x)\} \\
&= \sup_{x\in\mathbb{R}}\{yx - \inf_{z_1+\cdots+z_m=x}\{\sum_{i=1}^m f_i(z_i)\}\} \\
&= \sup_{x\in\mathbb{R}}\sup_{z_1+\cdots+z_m=x}\{\sum_{i=1}^m [yz_i - f_i(z_i)]\} \\
&= \sup_{\substack{z_1,\ldots,z_m,x\in\mathbb{R}\\ z_1+\cdots+z_m=x}}\{\sum_{i=1}^m [yz_i - f_i(z_i)]\} \\
&= \sup_{z_1,\ldots,z_m\in\mathbb{R}}\{\sum_{i=1}^m [yz_i - f_i(z_i)]\} \\
&= \sum_{i=1}^m \sup_{z_i\in\mathbb{R}}\{yz_i - f_i(z_i)\} \\
&= \sum_{i=1}^m f_i^*(y)
\end{aligned}$$

因此对于任意 $y \in \mathbb{R}$ 都有 $g^*(y) = \sum\limits_{i=1}^m f_i^*(y)$ 成立.

Problem 5 (10 Points)

考虑二次规划问题：

$$\begin{aligned}
\text{(P)}\ \min_{x\in\mathbb{R}^3}\quad & \frac{1}{2}x^TPx + q^Tx + r \\
\text{s.t.}\quad & -1 \le x_i \le 1,\ \ i = 1,2,3
\end{aligned}$$

其中 $P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, q = \begin{bmatrix} 22 \\ -14.5 \\ 13 \end{bmatrix}, r = 1$

请证明 $x_\star = (1, \frac{1}{2}, -1)$ 是问题 (P) 的全局最优解.

Solution:
首先我们证明 (P) 是一个凸问题 (这是为了使用 KKT 条件作为全局最优解的充分条件)：
由于 (P) 只具有线性约束，故我们只需证明目标函数为凸函数即可，即要证矩阵 P 是半正定的.

- **一阶顺序主子式：** $\det([13]) = 13 > 0$
- **二阶顺序主子式：** $\det\begin{bmatrix} 13 & 12 \\ 12 & 17 \end{bmatrix} = 13\times 17 - 12\times 12 > 0$
- **三阶顺序主子式：**

$$\begin{aligned}
\det\begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix} &= 13\det\begin{bmatrix} 17 & 6 \\ 6 & 12 \end{bmatrix} - 12\det\begin{bmatrix} 12 & 6 \\ -2 & 12 \end{bmatrix} + (-2)\det\begin{bmatrix} 12 & 17 \\ -2 & 6 \end{bmatrix} \\
&= 13(17\times 12 - 36) - 12(12\times 12 + 12) - 2(72 + 34) \\
&= 13\times 168 - 12\times 156 - 212 \\
&= 13\times(168 - 156) - (212 - 156) \\
&= 13\times 12 - 56 \\
&> 0
\end{aligned}$$

因此 P 是半正定阵 (更具体来说，它是正定阵)，
说明问题 (P) 是一个凸优化问题.

根据 **Lemma** (参见后文) 我们知道 KKT 条件是全局最优解的充分条件,
我们只需要验证 $x_\star = (1, \frac{1}{2}, -1)$ 满足 KKT 条件，即可证明 $x_\star$ 是 (P) 的全局最优解.

我们忽略 (P) 的目标函数中的常数项 $r=1$，得到:

$$
(\mathrm{newP}) \begin{array}{ll} \min\limits_{x\in\mathbb{R}^3} & \dfrac{1}{2}x^TPx + q^Tx \\ \text{s.t.} & -1 \le x_i \le 1,\ \ i=1,2,3 \end{array}
$$

我们引入 Lagrange 乘子 $\lambda, \nu \in \mathbb{R}^3$，定义 (newP) 的Lagrange 函数为:

$$
L(x,\lambda,\nu) = \tfrac{1}{2}x^TPx + q^Tx + \sum_{i=1}^3 \lambda_i(x_i-1) + \sum_{i=1}^3 \nu_i(-1-x_i)
$$

我们列出 KKT 条件:

- **可行性条件:** $-1 \le x_i \le 1,\ \ i=1,2,3$
- **一阶必要条件:**
$$
\begin{aligned}
\nabla_x L(x,\lambda,\nu) &= Px + q + \lambda - \nu \\
&= \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} - \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\end{aligned}
$$
即 $\begin{cases} 13x_1 + 12x_2 - 2x_3 + \lambda_1 - \nu_1 = 0 \\ 12x_1 + 17x_2 + 6x_3 + \lambda_2 - \nu_2 = 0 \\ -2x_1 + 6x_2 + 12x_3 + \lambda_3 - \nu_3 = 0 \end{cases}$
- **对偶可行性条件:** $\lambda_1, \lambda_2, \lambda_3, \nu_1, \nu_2, \nu_3 \ge 0$
- **(弱) 互补松弛条件:** $\begin{cases} \lambda_i(x_i-1) = 0, & i=1,2,3 \\ \nu_i(-1-x_i) = 0, & i=1,2,3 \end{cases}$

我们令 $x = x_\star = (1, \frac{1}{2}, -1)$ (注意这是一个可行点),

代入 KKT 系统得到 $\begin{cases} -1 + \lambda_1 - \nu_1 = 0 \\ 0 + \lambda_2 - \nu_2 = 0 \\ 2 + \lambda_3 - \nu_3 = 0 \\ \nu_1 = 0 \\ \lambda_2 = \nu_2 = 0 \\ \lambda_3 = 0 \end{cases}$ 进而解得 $\begin{cases} \lambda = (1,0,0) \\ \nu = (0,0,2) \end{cases}$ (注意它们是对偶可行的)

因此对于 $x_\star = (1, \frac{1}{2}, -1)$,
存在 Lagrange 乘子 $\begin{cases} \lambda_\star = (1,0,0) \\ \nu_\star = (0,0,2) \end{cases}$ 使得 $(x_\star, \lambda_\star, \nu_\star)$ 满足 KKT 条件,
故 $x_\star = (1, \frac{1}{2}, -1)$ 是一个 KKT 点,
结合问题 (P) 的凸性可知 $x_\star = (1, \frac{1}{2}, -1)$ 是全局最优解.

Lemma:

> **定理 4.2.6: (凸优化问题中全局最优解的充分条件, 江如俊教授 Note 4 Theorem 3.7)**
> 考虑凸优化问题:
> $$
> (\mathrm{P}) \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \le 0,\ i=1,\dots,m \\ & h_i(x) = 0,\ i=1,\cdots,p \\ & x \in \mathcal{X} \end{array}
> $$
> 其中 $\mathcal{X} \subseteq \mathbb{R}^n$ 为开凸集，$f, g_1, \dots, g_m$ 为 $\mathcal{X}$ 上的凸函数，$h_1, \dots, h_p$ 是仿射函数.
> 定义 Lagrange 函数为:
> $$
> \begin{cases} L(x,\lambda,\nu) = f(x) + \sum\limits_{i=1}^m \lambda_i g_i(x) + \sum\limits_{i=1}^p \nu_i h_i(x) = f(x) + \lambda^T g(x) + \nu^T h(x) \\ \mathrm{dom}(L) = \mathcal{X} \times \mathbb{R}^m \times \mathbb{R}^p \end{cases}
> $$
> 若 $(x_\star, \lambda_\star, \nu_\star) \in \mathcal{X} \times \mathbb{R}^m \times \mathbb{R}^p$ 是一个 KKT 点，即满足:
>
> - **可行性条件:** $\begin{cases} g(x_\star) \preceq 0_m \\ h(x_\star) = 0_p \end{cases}$

> - **一阶必要条件：** $\nabla_x L(x_\star, \lambda_\star, \nu_\star) = \nabla f(x_\star) + \sum\limits_{i=1}^m \lambda_i^\star \nabla g_i(x_\star) + \sum\limits_{i=1}^p \nu_i^\star h_i(x_\star) = 0_n$
> - **对偶可行性条件：** $\lambda_\star \succeq 0_m$
> - **(弱) 互补松弛条件：** $\lambda_i^\star g_i(x_\star) = 0 \ \ (\forall\, i = 1, \ldots, m)$
>
> 则 $x_\star$ 是凸优化问题 (P) 的全局最优解.

Problem 6 (25 Points)

Consider the following Minimax problem denoted (1):

$$\min_x \max_{i=1,\ldots,m} f_i(x) \tag{1}$$

where $f_1, \ldots, f_m : \mathbb{R}^n \mapsto \mathbb{R}$ are continuously differentiable convex functions.
(1) is equivalent to the one denoted (2):

$$\begin{array}{ll} \min\limits_x \ \ \max\limits_y & \sum\limits_{i=1}^m y_i f_i(x) \\ \quad\ \ \text{s.t.} & y \succeq 0_m \\ & 1_m^T y = 1 \end{array} \tag{2}$$

(a) List the optimality conditions for problem (2). (8 Points)
Hint:
You should fix x first, and derive the KKT conditions for the inner maximization problem.
Apply Lagrange multipliers $\begin{cases} \lambda \in \mathbb{R}^m \\ \nu \in \mathbb{R} \end{cases}$ for the constraints $\begin{cases} y \succeq 0_m \\ 1_m^T = 1 \end{cases}$, respectively.
Then you should fix y, and derive the first-order optimality condtions for the outer minimization problem.

Solution:
首先固定 $x \in \mathbb{R}^n$，考虑关于 $y \in \mathbb{R}^m$ 的最大化问题：

$$(\mathrm{P1}) \quad \begin{array}{ll} \max\limits_y & \sum\limits_{i=1}^m y_i f_i(x) \\ \text{s.t.} & y \succeq 0_m \\ & 1_m^T y = 1 \end{array}$$

显然 $(\mathrm{P1})$ 是一个凸优化问题，因为它只具有线性约束，且目标函数关于 $y \in \mathbb{R}^m$ 仿射 (自然是凸的).
又注意到 $(\mathrm{P1})$ 的任意可行点都是正则的，
结合问题的凸性我们知道 KKT 条件是 $(\mathrm{P1})$ 全局最优解的**充要条件**.

定义 $(\mathrm{P1})$ 的 Lagrange 函数为 $L(y, \lambda, \nu) = \sum\limits_{i=1}^m y_i f_i(x) + \lambda^T(y - 0_m) + \nu(1 - 1_m^T y)$
(注意 $(\mathrm{P1})$ 是最大化问题，定义 Lagrange 函数时需要格外谨慎)
则 **KKT 条件**如下：

- **可行性条件：** $\begin{cases} y \succeq 0_m \\ 1_m^T y = 1 \end{cases}$
- **Lagrange 函数驻点条件：** $\nabla_y L(y, \lambda, \nu) = \begin{bmatrix} f_1(x) + \lambda_1 - \nu \\ \vdots \\ f_m(x) + \lambda_m - \nu \end{bmatrix} = 0_m$
- **对偶可行性条件：** $\lambda \succeq 0_m$
- **(弱) 互补松弛性：** $\lambda_i y_i = 0 \ \ (\forall\, i = 1, \ldots, m)$

现在固定 $y \in \mathbb{R}_+^m$，考虑关于 $x \in \mathbb{R}^n$ 的最小化问题:

$$(\mathrm{P2})\ \min_x \sum_{i=1}^m y_i f_i(x)$$

这是一个无约束凸优化问题，根据**一阶最优化条件**可知:

$x \in \mathbb{R}^n$ 为 $(\mathrm{P2})$ 的全局最优值，当且仅当 $\sum_i^m y_i \nabla f_i(x) = 0_n$ 成立.

综上所述，问题 (2) 的最优性条件为:

$$\begin{cases} y \succeq 0_m \\ 1_m^T y = 1 \\ \lambda \succeq 0_m \\ f_i(x) = \nu - \lambda_i & (\forall\, i = 1, \ldots, m) \\ \lambda_i y_i = 0 & (\forall\, i = 1, \ldots, m) \\ \sum_i^m y_i \nabla f_i(x) = 0_n \end{cases}$$

(b) Denote $\begin{cases} \mathrm{T}(x) = \{i : f_i(x) = \max_{j=1,\ldots,m} f_j(x)\} \\ I(y) = \{i : y_i > 0\} \\ A(y) = \{i : y_i = 0\} \end{cases}$

First show that for all $(x_\star, y_\star, \lambda_\star, \nu_\star)$ satisfying optimality conditions given in (a),
we must have $I(y_\star) \subseteq \mathrm{T}(x_\star)$.
Moreover, show that if $(x_\star, y_\star, \lambda_\star, \nu_\star)$ is strictly complementary (i.e. $\forall\, i \in A(y_\star), \lambda_i^\star > 0$),
then $I(y_\star) = \mathrm{T}(x_\star)$. (7 Points)

Solution:

- **证明第一个命题:**
  假设 $(x_\star, y_\star, \lambda_\star, \nu_\star)$ 满足 (a) 中的最优性条件.

  对于任意 $i \in I(y_\star)$，我们有 $y_i^\star > 0$ 成立.
  根据 (弱) 互补松弛性可知 $\lambda_i^\star = 0$，进而有 $f_i(x_\star) = \nu_\star - \lambda_i^\star = \nu_\star$.

  注意到: 对于任意 $j = 1, \ldots, m$ 都有 $f_j(x_\star) = \nu_\star - \lambda_j^\star \leq \nu_\star$ 成立.
  因此 $f_i(x_\star) = \max_{j=1,\ldots,m} f_j(x_\star)$，说明 $i \in \mathrm{T}(x_\star)$.
  得证 $I(y_\star) \subseteq \mathrm{T}(x_\star)$

- **证明第二个命题:**
  假设 $(x_\star, y_\star, \lambda_\star, \nu_\star)$ 满足 (a) 中的最优性条件，且满足严格互补松弛条件.
  结合第一个命题可知，要证明第二个命题，只需证明 $\mathrm{T}(x_\star) \subseteq I(y_\star)$ 即可.

  对于任意 $i \in \mathrm{T}(x_\star)$，我们有 $f_i(x_\star) = \max_{j=1,\ldots,m} f_j(x_\star)$ 成立.

  (反证法) 假设 $i \notin I(y_\star)$ (即 $y_i^\star = 0$)，
  则根据严格互补松弛条件可知 $\lambda_i^\star > 0$.
  结合 $\begin{cases} f_i(x_\star) = \max_{j=1,\ldots,m} f_j(x_\star) \\ f_j(x_\star) = \nu_\star - \lambda_j^\star \ \ (\forall\, j = 1, \ldots, m) \end{cases}$ 可知对于任意 $j = 1, \ldots, m$ 都有 $\lambda_j^\star \geq \lambda_i^\star > 0$，
  根据 (弱) 互补松弛条件可知，对于任意 $j = 1, \ldots, m$ 都有 $y_j^\star = 0$ 成立，
  进而有 $1_m^T y_\star = 0$ 成立，这与 (a) 中最优性条件里的 "$1_m^T y_\star = 1$" 矛盾.
  因此反证法的假设不成立，我们有 $i \in I(y_\star)$ 成立，
  得证 $\mathrm{T}(x_\star) \subseteq I(y_\star)$

  结合第一个命题，进而得证 $I(y_\star) = \mathrm{T}(x_\star)$

(c) Suppose the strict complementary assumptions holds (i.e. $\forall\, i \in A(y_\star), \lambda_i^\star > 0$).
Denote $\mathrm{T}(x) = \{j_1, \dots, j_{|T(x)|}\}$ where $|\mathrm{T}(x)|$ represents the number of entries in $\mathrm{T}(x)$.
Define $M(x) = \begin{bmatrix} \nabla f_{j_1}(x) & \dots & \nabla f_{j_{|\mathrm{T}(x)|}}(x) \\ 1 & \dots & 1 \end{bmatrix}$

Prove for $(x_\star, y_\star)$ satisfies the KKT conditions in (a),
the matrix $M(x_\star)$ is of full column rank.
Hint:
You may assume that there exists a nonzero vector $c \in \mathbb{R}^{|\mathrm{T}(x_\star)|}$ such that $M(x_\star)c = 0$
and use c and $y^\star$ to construct a counterexample.

Solution:
假设 $(x_\star, y_\star)$ 满足 (a) 中的 KKT 条件，且满足严格互补松弛条件.

(反证法) 假设存在非零向量 $c \in \mathbb{R}^{|\mathrm{T}(x_\star)|}$ 使得 $M(x_\star)c = 0$ 成立.
不失一般性，假设存在 $i \in \mathrm{T}(x_\star)$ 使得 $c_i > 0$ (如果不存在，我们可以使用 $-c$ 代替 c)

根据 (b) 的结论，可知有 $I(y_\star) = \mathrm{T}(x_\star)$ 成立.
定义：

$$
\begin{aligned}
d &= \min\{\frac{y_i^\star}{c_i} : i \in I(y_\star) \text{ such that } c_i > 0\} \\
&= \min\{\frac{y_i^\star}{c_i} : i \in \mathrm{T}(x_\star) \text{ such that } c_i > 0\}
\end{aligned}
$$

定义 $z \in \mathbb{R}^m$：

$$
z_i = \begin{cases} y_i^\star - c_i d & \text{if } i \in I(y_\star) = \mathrm{T}(x_\star) \\ 0 & \text{otherwise} \end{cases}
$$

根据 (b) 的结论，可知有 $I(y_\star) = \mathrm{T}(x_\star)$ 成立.
因此，对于任意 $i \notin \mathrm{T}(x_\star)$，都有 $i \notin I(y_\star)$ 成立，即有 $y_i^\star = 0$ 成立.
而对于任意 $i \in \mathrm{T}(x_\star)$，都有 $i \in I(y_\star)$ 成立，即有 $y_i^\star > 0$ 成立.
于是我们有：

$$
\begin{aligned}
\sum_i^m y_i^\star \nabla f_i(x_\star) &= \sum_{i \notin \mathrm{T}(x_\star)} y_i^\star \nabla f_i(x_\star) + \sum_{i \in \mathrm{T}(x_\star)} y_i^\star \nabla f_i(x_\star) \\
&= \sum_{i \in \mathrm{T}(x_\star)} y_i^\star \nabla f_i(x_\star)
\end{aligned}
$$

(反证法) 假设存在不全为 0 的 $c \in \mathbb{R}^m$ 使得 $M(x_\star)$ 的列线性组合为 0_{n+1}，

即使得 $\begin{cases} \sum\limits_{i \in \mathrm{T}(x_\star)} c_i \nabla f_i(x_\star) = 0_n \\ \sum\limits_{i \in \mathrm{T}(x_\star)} c_i = 0 \end{cases}$ 成立.

一定存在 $i_0 \in \mathrm{T}(x_\star)$ 使得 $c_{i_0} > 0$，

(d) If x^* is an optimal solution of (2), then there exists a unique y, such that (x^*, y, μ, v) solves the system given in (a).