MAS374 Optimization Theory
Homework #7

20150597 Jeonghwan Lee*

Department of Mathematical Sciences, KAIST

November 25, 2021

Problem 1 (*Exercise 10.1* in [1]: Squaring SOCP constraints)**.**

Let

$$\mathcal{C}_n := \left\{ \mathbf{x} = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n : \|\mathbf{x}\|_2^2 \leq (x_1 + 2x_2)^2 \right\}, \ \forall n \geq 2.$$

We claim that $\mathcal{C}_n$ is not convex in $\mathbb{R}^n$ for every $n \geq 2$. First, we consider the case where $n = 2$. Then, the subset $\mathcal{C}_2 \subseteq \mathbb{R}^2$ can be expressed in the following simpler way:

$$\mathcal{C}_2 = \left\{ (x_1, x_2) \in \mathbb{R}^2 : x_2 (4x_1 + 3x_2) \geq 0 \right\}.$$

Then $\mathcal{C}_2$ is not convex in $\mathbb{R}^2$ from the following reason: It's clear that $(3, 0), (3, -4) \in \mathcal{C}_2$. However, their midpoint $(3, -2)$ does not belong to $\mathcal{C}_2$ and this shows that $\mathcal{C}_2$ is not convex in $\mathbb{R}^2$.

Hereafter, we consider the case for which $n \geq 3$ and assume towards a contradiction that $\mathcal{C}_n$ is a convex subset of $\mathbb{R}^n$. Consider two vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, where

$$\begin{aligned} \mathbf{x} &:= \mathbf{e}_1 + \mathbf{e}_2 + \sqrt{\frac{7}{n-2}} \left(\sum_{j=3}^{n} \mathbf{e}_j \right); \\ \mathbf{y} &:= -\mathbf{e}_1 - \mathbf{e}_2 + \sqrt{\frac{7}{n-2}} \left(\sum_{j=3}^{n} \mathbf{e}_j \right), \end{aligned}$$

where $\mathbf{e}_j \in \mathbb{R}^n$ denotes the j-th unit vector in $\mathbb{R}^n$ for $j \in [n]$. Then, it's straightforward to see that

$$\begin{aligned} \|\mathbf{x}\|_2^2 &= 1^2 + 1^2 + (n-2) \cdot \left(\sqrt{\frac{7}{n-2}} \right)^2 = 9 = (x_1 + 2x_2)^2; \\ \|\mathbf{y}\|_2^2 &= (-1)^2 + (-1)^2 + (n-2) \cdot \left(\sqrt{\frac{7}{n-2}} \right)^2 = 9 = (y_1 + 2y_2)^2, \end{aligned} \tag{1.1}$$

thereby $\mathbf{x}, \mathbf{y} \in \mathcal{C}_n$. Due to the convexity of $\mathcal{C}_n$ in $\mathbb{R}^n$, it holds that $(1-\theta)\mathbf{x} + \theta\mathbf{y} \in \mathcal{C}_n$ for every $\theta \in [0, 1]$:

$$\begin{aligned} \|(1-\theta)\mathbf{x} + \theta\mathbf{y}\|_2^2 &\leq \left[\{(1-\theta) x_1 + \theta y_1\} + 2\{(1-\theta) x_2 + \theta y_2\} \right]^2 \\ &= \{(1-\theta)(x_1 + 2x_2) + \theta (y_1 + 2y_2)\}^2. \end{aligned} \tag{1.2}$$

*E-mail: sa8seung@kaist.ac.kr

However, by doing some straightforward algebra, we arrive at

$$\begin{aligned}
&\|(1-\theta)\,\mathbf{x}+\theta\mathbf{y}\|_2^2-\{(1-\theta)\,(x_1+2x_2)+\theta\,(y_1+2y_2)\}^2\\
=\;&(1-\theta)^2\left\{\|\mathbf{x}\|_2^2-(x_1+2x_2)^2\right\}+\theta^2\left\{\|\mathbf{y}\|_2^2-(y_1+2y_2)^2\right\}+2\theta\,(1-\theta)\left\{\mathbf{x}^\top\mathbf{y}-(x_1+2x_2)\,(y_1+2y_2)\right\}\\
\overset{\text{(a)}}{=}\;&2\theta\,(1-\theta)\left\{\mathbf{x}^\top\mathbf{y}-(x_1+2x_2)\,(y_1+2y_2)\right\}\\
=\;&28\theta\,(1-\theta)>0
\end{aligned}\tag{1.3}$$

for every $\theta \in (0,1)$, where the step (a) follows from the equation (1.1). Therefore, the conclusion (1.3) yields a contradiction against the convexity of $\mathcal{C}_n$ in $\mathbb{R}^n$, which proves that $\mathcal{C}_n$ is not convex in $\mathbb{R}^n$ as desired.

Problem 2 (*Exercise 10.6* in [1]: A trust-region problem)**.**

We consider the following primal convex quadratic-constrained quadratic program (QCQP):

$$\begin{aligned}
p^* = \min_{\mathbf{x}\in\mathbb{R}^n}\;&\left(\frac{1}{2}\mathbf{x}^\top\mathbf{H}\mathbf{x}+\mathbf{c}^\top\mathbf{x}+d\right)\\
\text{subject to }&\mathbf{x}^\top\mathbf{x}\le r^2,
\end{aligned}\tag{2.1}$$

where $\mathbf{H}\in\mathcal{S}_{++}^n$, $\mathbf{c}\in\mathbb{R}^n$, $d\in\mathbb{R}$, and $r\in(0,+\infty)$, where $\mathcal{S}_{++}^n$ is the set of all $n\times n$ positive definite real symmetric matrices. Then the Lagrangian function of (2.1) is given by $\mathcal{L}(\cdot,\cdot):\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$, where

$$\mathcal{L}(\mathbf{x},\lambda) := \left(\frac{1}{2}\mathbf{x}^\top\mathbf{H}\mathbf{x}+\mathbf{c}^\top\mathbf{x}+d\right)+\lambda\left(\mathbf{x}^\top\mathbf{x}-r^2\right)=\frac{1}{2}\mathbf{x}^\top\left(\mathbf{H}+2\lambda\mathbf{I}_n\right)\mathbf{x}+\mathbf{c}^\top\mathbf{x}+\left(d-r^2\lambda\right).$$

So one can see that

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x},\lambda)=\left(\mathbf{H}+2\lambda\mathbf{I}_n\right)\mathbf{x}+\mathbf{c},$$

which implies that for every $\lambda\ge 0$,

$$\operatorname{argmin}\left\{\mathcal{L}(\mathbf{x},\lambda):\mathbf{x}\in\mathbb{R}^n\right\}=\{\mathbf{x}(\lambda)\},\ \text{ where } \mathbf{x}(\lambda):=-\left(\mathbf{H}+2\lambda\mathbf{I}_n\right)^{-1}\mathbf{c}.\tag{2.2}$$

Here, we note that $\mathbf{H}+2\lambda\mathbf{I}_n\in\mathcal{S}_{++}^n$ and the function $\mathbf{x}\in\mathbb{R}^n\mapsto\mathcal{L}(\mathbf{x},\lambda)\in\mathbb{R}$ is convex for every $\lambda\ge 0$. So if $g(\lambda):\mathbb{R}\to[-\infty,+\infty)$ is the Lagrange dual function associated to the primal convex QCQP (2.1), then for every $\lambda\ge 0$,

$$g(\lambda)=\inf\left\{\mathcal{L}(\mathbf{x},\lambda):\mathbf{x}\in\mathbb{R}^n\right\}=\mathcal{L}(\mathbf{x}(\lambda),\lambda)=-\frac{1}{2}\mathbf{c}^\top\left(\mathbf{H}+2\lambda\mathbf{I}_n\right)^{-1}\mathbf{c}+d-r^2\lambda.$$

Hence, the dual problem associated to the primal convex QCQP (2.1) is formulated by

$$\begin{aligned}
d^* = \max_{\lambda\in\mathbb{R}}\;&\left\{-\frac{1}{2}\mathbf{c}^\top\left(\mathbf{H}+2\lambda\mathbf{I}_n\right)^{-1}\mathbf{c}+d-r^2\lambda\right\}\\
\text{subject to }&\lambda\ge 0.
\end{aligned}\tag{2.3}$$

We note that since the primal convex QCQP (2.1) is strictly feasible, the strong duality holds between the primal convex QCQP (2.1) and its dual problem (2.3), *i.e.*, $p^*=d^*$, by the Slater's condition for convex programs (*Proposition 8.7* in [1]).

Now, we let $\mathcal{P}_{\mathsf{opt}}\subseteq\mathbb{R}^n$ and $\mathcal{D}_{\mathsf{opt}}\subseteq\mathbb{R}$ denote the sets of optimal solutions to the primal convex QCQP (2.1) and its dual problem (2.3), respectively. Note that $\mathcal{P}_{\mathsf{opt}}\neq\varnothing$ since the feasible set of the primal problem (2.1) is a compact subset of $\mathbb{R}^n$ and the primal objective function is convex.

Claim 1. $\mathcal{P}_{\mathsf{opt}} \subseteq \{\mathbf{x}(\lambda^*) : \lambda^* \in \mathcal{D}_{\mathsf{opt}}\}$.

Proof of Claim 1. Choose any $(\mathbf{x}^*, \lambda^*) \in \mathcal{P}_{\mathsf{opt}} \times \mathcal{D}_{\mathsf{opt}}$. Since the strong duality between the primal convex QCQP (2.1) and its dual problem (2.3) holds, the Karush-Kuhn-Tucker (KKT) conditions hold for the pair $(\mathbf{x}^*, \lambda^*) \in \mathcal{P}_{\mathsf{opt}} \times \mathcal{D}_{\mathsf{opt}}$:

(i) Lagrangian stationarity: $\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda^*)|_{\mathbf{x}=\mathbf{x}^*} = \mathbf{0}$;

(ii) Complementary slackness: $\lambda^* \left((\mathbf{x}^*)^\top \mathbf{x}^* - r^2 \right) = 0$;

(iii) Primal feasibility: $(\mathbf{x}^*)^\top \mathbf{x}^* \leq r^2$;

(iv) Dual feasibility: $\lambda^* \geq 0$.

The condition (i) implies

$$(\mathbf{H} + 2\lambda^* \mathbf{I}_n)\, \mathbf{x}^* + \mathbf{c} = \mathbf{0},$$

thereby $\mathbf{x}^* = -(\mathbf{H} + 2\lambda^* \mathbf{I}_n)^{-1} \mathbf{c} = \mathbf{x}(\lambda^*)$. This establishes Claim 1.

□

In view of Claim 1, it suffices to find dual optimal solutions explicitly! The following result characterizes the set $\mathcal{D}_{\mathsf{opt}}$ of dual optimal solutions:

Claim 2. *It holds that* $\mathcal{D}_{\mathsf{opt}} = \{\lambda^*\}$, *where*

$$\lambda^* := \begin{cases} 0 & \text{if } \left\|\mathbf{H}^{-1}\mathbf{c}\right\|_2 \leq r; \\ \theta^* & \text{otherwise,} \end{cases} \tag{2.4}$$

where $\theta^* \in \mathbb{R}_+$ *is the unique value in* $(0, +\infty)$ *such that* $\left\| (\mathbf{H} + 2\theta^* \mathbf{I}_n)^{-1} \mathbf{c} \right\|_2 = r$ *provided that* $\left\|\mathbf{H}^{-1}\mathbf{c}\right\|_2 > r$.

Proof of Claim 2. Let $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\top = \sum_{i=1}^n \lambda_i(\mathbf{H}) \mathbf{u}_i \mathbf{u}_i^\top$ be the spectral decomposition of $\mathbf{H} \in \mathcal{S}_{++}^n$, where $\mathbf{U} := \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \in \mathcal{O}(n)$, $\mathbf{\Sigma} := \mathsf{diag}\left(\lambda_1(\mathbf{H}), \lambda_2(\mathbf{H}), \cdots, \lambda_n(\mathbf{H})\right) \in \mathbb{R}^{n \times n}$, and

$$\lambda_1(\mathbf{H}) \geq \lambda_2(\mathbf{H}) \geq \cdots \geq \lambda_n(\mathbf{H}) > 0.$$

Here, $\mathcal{O}(n)$ denotes the orthogonal group of dimension n. Then for every $\lambda \geq 0$,

$$\begin{aligned} g(\lambda) &= -\frac{1}{2}\mathbf{c}^\top (\mathbf{H} + 2\lambda \mathbf{I}_n)^{-1} \mathbf{c} + d - r^2\lambda \\ &= -\frac{1}{2}\mathbf{c}^\top \left\{ \mathbf{U}\mathsf{diag}\left(\frac{1}{\lambda_1(\mathbf{H}) + 2\lambda}, \frac{1}{\lambda_2(\mathbf{H}) + 2\lambda}, \cdots, \frac{1}{\lambda_n(\mathbf{H}) + 2\lambda} \right) \mathbf{U}^\top \right\} \mathbf{c} + d - r^2\lambda \\ &= -\frac{1}{2} \sum_{i=1}^n \left(\left[\mathbf{U}^\top \mathbf{c} \right]_i \right)^2 \cdot \frac{1}{\lambda_i(\mathbf{H}) + 2\lambda} + d - r^2\lambda \\ &\overset{\text{(a)}}{=} -\frac{1}{2} \sum_{i=1}^n \frac{(\mathbf{u}_i^\top \mathbf{c})^2}{\lambda_i(\mathbf{H}) + 2\lambda} + d - r^2\lambda, \end{aligned} \tag{2.5}$$

where the step (a) holds since

$$\left[\mathbf{U}^\top \mathbf{c} \right]_i = \mathbf{e}_i^\top \mathbf{U}^\top \mathbf{c} = (\mathbf{U}\mathbf{e}_i)^\top \mathbf{c} = \mathbf{u}_i^\top \mathbf{c},\ \forall i \in [n].$$

From the equation (2.5), one has

$$\begin{aligned} g'(\lambda) &= \sum_{i=1}^{n} \frac{\left(\mathbf{u}_i^\top \mathbf{c}\right)^2}{\left\{\lambda_i(\mathbf{H}) + 2\lambda\right\}^2} - r^2 \\ &= \mathbf{c}^\top \left\{ \sum_{i=1}^{n} \frac{1}{\left\{\lambda_i(\mathbf{H}) + 2\lambda\right\}^2} \mathbf{u}_i \mathbf{u}_i^\top \right\} \mathbf{c} - r^2 \\ &= \mathbf{c}^\top \left\{ \mathbf{U}\mathsf{diag}\left(\frac{1}{\left\{\lambda_1(\mathbf{H}) + 2\lambda\right\}^2}, \frac{1}{\left\{\lambda_2(\mathbf{H}) + 2\lambda\right\}^2}, \cdots, \frac{1}{\left\{\lambda_n(\mathbf{H}) + 2\lambda\right\}^2} \right) \mathbf{U}^\top \right\} \mathbf{c} - r^2 \\ &= \mathbf{c}^\top \left\{ \left(\mathbf{H} + 2\lambda \mathbf{I}_n\right)^{-1} \right\}^\top \left(\mathbf{H} + 2\lambda \mathbf{I}_n\right)^{-1} \mathbf{c} - r^2 \\ &= \left\| \left(\mathbf{H} + 2\lambda \mathbf{I}_n\right)^{-1} \mathbf{c} \right\|_2^2 - r^2 \end{aligned} \tag{2.6}$$

and

$$g''(\lambda) = -2 \sum_{i=1}^{n} \frac{\left(\mathbf{u}_i^\top \mathbf{c}\right)^2}{\left\{\lambda_i(\mathbf{H}) + 2\lambda\right\}^3}$$

for every $\lambda \geq 0$.

Case #1. $\mathbf{c} = \mathbf{0}$: For this case, it's clear that the dual optimal solution is unique and given by $\lambda^* = 0$. This completes the proof of Claim 2 for the case where $\mathbf{c} = \mathbf{0}$.

Case #2. $\mathbf{c} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$: Then we know that $g''(\lambda) < 0$ for all $\lambda \geq 0$. So, the function $\lambda \in \mathbb{R}_+ \mapsto g'(\lambda) \in \mathbb{R}$ is strictly decreasing. This guarantees that if $\left\|\mathbf{H}^{-1}\mathbf{c}\right\|_2 > r$, there exists a unique value $\theta^* \in (0, +\infty)$ such that $\left\| \left(\mathbf{H} + 2\theta^* \mathbf{I}_n\right)^{-1} \mathbf{c} \right\|_2^2 = r^2$ since $g'(0) = \left\|\mathbf{H}^{-1}\mathbf{c}\right\|_2^2 - r^2 > 0$.

(i) $\left\|\mathbf{H}^{-1}\mathbf{c}\right\|_2 \leq r$: We have $g'(0) = \left\|\mathbf{H}^{-1}\mathbf{c}\right\|_2^2 - r^2 \leq 0$. Thus, $g'(\lambda) < 0$ for every $\lambda \in (0, +\infty)$ and this ensures that

$$\operatorname{argmax} \left\{ g(\lambda) : \lambda \in \mathbb{R}_+ \right\} = \mathcal{D}_{\mathsf{opt}} = \{0\}, \tag{2.7}$$

since the function $\lambda \in \mathbb{R}_+ \mapsto g'(\lambda) \in \mathbb{R}$ is strictly decreasing;

(ii) $\left\|\mathbf{H}^{-1}\mathbf{c}\right\|_2 > r$: We have $g'(0) = \left\|\mathbf{H}^{-1}\mathbf{c}\right\|_2^2 - r^2 > 0$. For this case, one can see that

$$\left\{ \lambda \in \mathbb{R}_+ : g'(\lambda) = 0 \right\} = \{\theta^*\}.$$

Since the function $\lambda \in \mathbb{R}_+ \mapsto g'(\lambda) \in \mathbb{R}$ is strictly decreasing, we may conclude that

$$\operatorname{argmax} \left\{ g(\lambda) : \lambda \in \mathbb{R}_+ \right\} = \mathcal{D}_{\mathsf{opt}} = \{\theta^*\}. \tag{2.8}$$

By taking two pieces (2.7) and (2.8) collectively, we finish the proof of Claim 2 for the case where $\mathbf{c} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$.

□

Due to Claim 1 & 2 and from the fact that $\mathcal{P}_{\mathsf{opt}}$ is non-empty, it holds that

$$\mathcal{P}_{\mathsf{opt}} = \left\{ \mathbf{x}\left(\lambda^*\right) : \lambda^* \in \mathcal{D}_{\mathsf{opt}} \right\} = \left\{ -\left(\mathbf{H} + 2\lambda^* \mathbf{I}_n\right)^{-1} \mathbf{c} : \lambda^* \in \mathcal{D}_{\mathsf{opt}} \right\}. \tag{2.9}$$

Hence, $\mathbf{x}\left(\lambda^*\right) = -\left(\mathbf{H} + 2\lambda^* \mathbf{I}_n\right)^{-1} \mathbf{c}$ is the unique optimal solution to the primal convex QCQP (2.1), where $\lambda^* \in \mathbb{R}_+$ is the value given by (2.4). By substituting λ^* to $\frac{\lambda^*}{2}$, the value λ^* equals to the one in Problem 2 and this completes the proof.

Problem 3 (*Exercise 10.8* in [1]: Proving convexity via duality)**.**

(1) To begin with, we define a function $\Phi(\cdot,\cdot) : \mathbb{R}^n \times \mathbb{R} \to (-\infty, +\infty]$ by

$$\Phi(\mathbf{x}, t) := \begin{cases} 2\left(t - \sum_{i=1}^n \sqrt{x_i + t^2}\right) & \text{if } \mathbf{x} \in \mathbb{R}^n_{++}; \\ +\infty & \text{otherwise,} \end{cases}$$

and $f(\cdot) : \mathbb{R}^n \to (-\infty, +\infty]$ by

$$f(\mathbf{x}) := \sup\left\{\Phi(\mathbf{x}, t) : t \in \mathbb{R}\right\}.$$

Now we fix any $\mathbf{x} \in \mathbb{R}^n_{++}$. Then we have

$$\frac{\partial}{\partial t}\Phi(\mathbf{x}, t) = 2 - 2t\sum_{i=1}^n \frac{1}{\sqrt{x_i + t^2}},$$

and

$$\frac{\partial^2}{\partial t^2}\Phi(\mathbf{x}, t) = -2\sum_{i=1}^n \frac{x_i}{(x_i + t^2)^{\frac{3}{2}}}$$

for every $t \in \mathbb{R}$. Thus, one can see that $\frac{\partial^2}{\partial t^2}\Phi(\mathbf{x}, t) < 0$ for every $t \in \mathbb{R}$. Hence, the function $\Phi(\mathbf{x}, \cdot) : \mathbb{R} \to \mathbb{R}$ is concave for every fixed $\mathbf{x} \in \mathbb{R}^n_{++}$, which shows that the following optimization problem which defines the value of $f(\mathbf{x}) \in \mathbb{R}$:

$$f(\mathbf{x}) = \max_{t \in \mathbb{R}} \Phi(\mathbf{x}, t) = \max_{t \in \mathbb{R}} 2\left(t - \sum_{i=1}^n \sqrt{x_i + t^2}\right), \tag{3.1}$$

is a convex optimization problem in the variable $t \in \mathbb{R}$ for each $\mathbf{x} \in \mathbb{R}^n_{++}$.

Hereafter, we provide you an equivalent second-order cone program (SOCP) formulation to the optimization problem (3.1). Note that the optimization problem (3.1) is equivalent with the following formulation with additional variable $\mathbf{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}^\top \in \mathbb{R}^n$:

$$\begin{aligned} -f(\mathbf{x}) = \min_{(t,\mathbf{u}) \in \mathbb{R} \times \mathbb{R}^n} \ & 2\left(\sum_{i=1}^n u_i - t\right) \\ \text{subject to } & \sqrt{x_i + t^2} \leq u_i, \ i \in [n]. \end{aligned} \tag{3.2}$$

Let $\mathbf{e}_j^{(n+1)} \in \mathbb{R}^{n+1}$ denote the j-th unit vector for every $j \in [n+1]$. By setting $\mathbf{A}_i := \mathbf{e}_{i+1}^{(n+1)}\left(\mathbf{e}_{i+1}^{(n+1)}\right)^\top \in \mathbb{R}^{(n+1)\times(n+1)}$, $\mathbf{b}_i := \sqrt{x_i} \cdot \mathbf{e}_{i+1}^{(n+1)} \in \mathbb{R}^{n+1}$, $\mathbf{c}_i := \mathbf{e}_{i+1}^{n+1} \in \mathbb{R}^{(n+1)}$, and $d_i := 0 \in \mathbb{R}$ for $i \in [n]$, one can realize that the inequality constraint $\sqrt{x_i + t^2} \leq u_i$ in the equivalent formulation (3.2) of the optimization problem (3.1) is equivalent to

$$\left\|\mathbf{A}_i \begin{bmatrix} t \\ \mathbf{u} \end{bmatrix} + \mathbf{b}_i\right\|_2 \leq \mathbf{c}_i^\top \begin{bmatrix} t \\ \mathbf{u} \end{bmatrix} + d_i$$

for every $i \in [n]$. Hence, the optimization problem (3.2) can be equivalently formulated into the following SOCP:

$$\begin{aligned} -f(\mathbf{x}) = \min_{(t,\mathbf{u}) \in \mathbb{R} \times \mathbb{R}^n} \ & \mathbf{c}^\top \begin{bmatrix} t \\ \mathbf{u} \end{bmatrix} \\ \text{subject to } & \left\|\mathbf{A}_i \begin{bmatrix} t \\ \mathbf{u} \end{bmatrix} + \mathbf{b}_i\right\|_2 \leq \mathbf{c}_i^\top \begin{bmatrix} t \\ \mathbf{u} \end{bmatrix} + d_i, \ i \in [n], \end{aligned} \tag{3.3}$$

where $\mathbf{c} := -2\mathbf{e}_1^{(n+1)} + \sum_{i=2}^{n+1} 2\mathbf{e}_i^{(n+1)} \in \mathbb{R}^{n+1}$, which gives an SOCP formulation which is equivalent with the original problem (3.1).

(2) We first prove that the function $\Phi(\cdot, t) : \mathbb{R}^n \to (-\infty, +\infty]$ is convex for every $t \in \mathbb{R}$. By doing some straightforward calculations, we obtain

$$\frac{\partial}{\partial x_i}\Phi(\mathbf{x}, t) = -\frac{1}{\sqrt{x_i + t^2}}, \ \forall i \in [n],$$

and

$$\frac{\partial^2}{\partial x_j x_i}\Phi(\mathbf{x}, t) = \begin{cases} \frac{1}{2}\left(x_i + t^2\right)^{-\frac{3}{2}} & \text{if } i = j; \\ 0 & \text{otherwise,} \end{cases}$$

for every $\mathbf{x} \in \mathbb{R}^n_{++} = \mathsf{dom}(\Phi(\cdot, t))$. Therefore, one has

$$\nabla^2_{\mathbf{x}}\Phi(\mathbf{x}, t) = \mathsf{diag}\left(\frac{1}{2}\left(x_1 + t^2\right)^{-\frac{3}{2}}, \frac{1}{2}\left(x_2 + t^2\right)^{-\frac{3}{2}}, \cdots, \frac{1}{2}\left(x_n + t^2\right)^{-\frac{3}{2}}\right) \in \mathcal{S}^n_{++}$$

for every $\mathbf{x} \in \mathbb{R}^n_{++} = \mathsf{dom}(\Phi(\cdot, t))$. Note that the effective domain of the function $\Phi(\cdot, t) : \mathbb{R}^n \to (-\infty, +\infty]$, $\mathsf{dom}(\Phi(\cdot, t)) = \mathbb{R}^n_{++}$, is a convex subset of $\mathbb{R}^n$ for every $t \in \mathbb{R}$. So the second-order condition for convexity implies that $\Phi(\cdot, t) : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex function for every $t \in \mathbb{R}$. Hence, their pointwise supremum over $t \in \mathbb{R}$:

$$f(\cdot) := \sup\{\Phi(\cdot, t) : t \in \mathbb{R}\} : \mathbb{R}^n \to (-\infty, +\infty],$$

is also a convex function.

(3) We may observe that

$$\sup\left\{-\mathbf{y}^\top\mathbf{x} - f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n_{++}\right\} = \sup\left\{\inf\left\{-\mathbf{y}^\top\mathbf{x} - \Phi(\mathbf{x}, t) : t \in \mathbb{R}\right\} : \mathbf{x} \in \mathbb{R}^n_{++}\right\}. \tag{3.4}$$

On the other hand, we know that

$$\nabla_{\mathbf{x}}\Phi(\mathbf{x}, t) = -\mathbf{y} + \begin{bmatrix} \frac{1}{\sqrt{x_1+t^2}} & \frac{1}{\sqrt{x_2+t^2}} & \cdots & \frac{1}{\sqrt{x_n+t^2}} \end{bmatrix}^\top.$$

So we obtain

$$\operatorname{argmax}\left\{\Phi(\mathbf{x}, t) : \mathbf{x} \in \mathbb{R}^n_{++}\right\} = \begin{cases} \{\mathbf{x}^*(t)\} & \text{if } t^2 < \min\left\{\frac{1}{y_i^2} : i \in [n]\right\}; \\ \varnothing & \text{otherwise,} \end{cases}$$

where $\mathbf{x}^*(t) := \begin{bmatrix} x_1^*(t) & x_2^*(t) & \cdots & x_n^*(t) \end{bmatrix}^\top \in \mathbb{R}^n_{++}$ is given by $x_i^*(t) := \frac{1}{y_i^2} - t^2$ for $i \in [n]$. Thus,

$$\begin{aligned} \sup\left\{-\mathbf{y}^\top\mathbf{x} - f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n_{++}\right\} &= \begin{cases} -\mathbf{y}^\top\mathbf{x}^*(t) - \Phi(\mathbf{x}^*(t), t) & \text{if } t^2 < \min\left\{\frac{1}{y_i^2} : i \in [n]\right\}; \\ +\infty & \text{otherwise.} \end{cases} \\ &= \begin{cases} \sum_{i=1}^n \frac{1}{y_i} - 2t + t^2\left(\sum_{i=1}^n y_i\right) & \text{if } t^2 < \min\left\{\frac{1}{y_i^2} : i \in [n]\right\}; \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Therefore, we arrive at

$$\inf\left\{\sup\left\{-\mathbf{y}^\top\mathbf{x} - f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n_{++}\right\} : t \in \mathbb{R}\right\} = \sum_{i=1}^n \frac{1}{y_i} - \frac{1}{\sum_{i=1}^n y_i} = g(\mathbf{y}) \tag{3.5}$$

for every $\mathbf{y} \in \mathbb{R}^n_{++}$. At this point, we assume that the following minimax principle holds:

$$\inf \left\{ \sup \left\{ -\mathbf{y}^\top \mathbf{x} - f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n_{++} \right\} : t \in \mathbb{R} \right\} = \sup \left\{ \inf \left\{ -\mathbf{y}^\top \mathbf{x} - \Phi\left(\mathbf{x}, t\right) : t \in \mathbb{R} \right\} : \mathbf{x} \in \mathbb{R}^n_{++} \right\}. \tag{3.6}$$

By taking three pieces (3.4), (3.5), and (3.6) collectively, we eventually get

$$g(\mathbf{y}) = \sup \left\{ -\mathbf{y}^\top \mathbf{x} - f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^n_{++} \right\}.$$

Since the function $\mathbf{y} \in \mathbb{R}^n_{++} \mapsto -\mathbf{y}^\top \mathbf{x} - f(\mathbf{x}) \in \mathbb{R}$ is an affine function for every $\mathbf{x} \in \mathbb{R}^n_{++}$, $g(\cdot) : \mathbb{R}^n_{++} \to \mathbb{R}$ is a convex function. However, I still don't know how to prove the proposed minimax principle (3.6) rigorously..

References

[1] Giuseppe C Calafiore and Laurent El Ghaoui. *Optimization models*. Cambridge university press, 2014.