Biomath 210 Homework 2

Huwenbo Shi (603-778-363) shihuwenbo@ucla.edu

October 12, 2015

Problem 2.10

To show $f(\mathbf{x})_+$ is convex, we notice that

$$\begin{aligned} \text{epi } f(\mathbf{x})_+ &= \{(\mathbf{x}, y) : \mathbf{x} \in \text{dom } f, \max\{f(\mathbf{x}), 0\} \leqslant y\} \\ &= \{(\mathbf{x}, y) : \mathbf{x} \in \text{dom } f, f(\mathbf{x}) \leqslant y\} \cap \{(\mathbf{x}, y) : \mathbf{x} \in \text{dom } f, 0 \leqslant y\}. \end{aligned} \tag{1}$$

Because $f(\mathbf{x})$ is a convex function, the set $\{(\mathbf{x}, y) : \mathbf{x} \in \text{dom } f, f(\mathbf{x}) \leqslant y\}$ is a convex set. The set $\{(\mathbf{x}, y) : \mathbf{x} \in \text{dom } f, 0 \leqslant y\}$ is also a convex set because it's the epigraph of the function $f(\mathbf{x}) = 0$. Because the intersection of convex sets is convex, we conclude that epi $f(\mathbf{x})_+$ is a convex set. And therefore, the function $f(\mathbf{x})_+$ is a convex function.

Let $h(\mathbf{x}) = \sqrt{f(\mathbf{x})^2 + \epsilon}$. To show the convexity of $h(\mathbf{x})$, we show that $h(\mathbf{x})$ satisfies the following definition of convex functions

$$h(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leqslant \alpha h(\mathbf{x}) + (1-\alpha)h(\mathbf{y}). \tag{2}$$

Because $f(\mathbf{x})$ is convex and non-negative, we have

$$\begin{aligned} h(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) &= \sqrt{f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})^2 + \epsilon} \\ &\leqslant \sqrt{[\alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})]^2 + \epsilon}. \end{aligned} \tag{3}$$

Taking the square of the right hand side of the inequality, we have

$$\begin{aligned} &\left(\sqrt{[\alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})]^2 + \epsilon}\right)^2 \\ &= [\alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y})]^2 + \epsilon \\ &= \alpha^2 f(\mathbf{x})^2 + (1-\alpha)^2 f(\mathbf{y})^2 + 2\alpha(1-\alpha)f(\mathbf{x})f(\mathbf{y}) + \epsilon. \end{aligned} \tag{4}$$

Taking the square of $(\alpha h(\mathbf{x}) + (1-\alpha)h(\mathbf{y}))$, we get

$$\begin{aligned} &(\alpha h(\mathbf{x}) + (1-\alpha)h(\mathbf{y}))^2 \\ &= \left(\alpha\sqrt{f(\mathbf{x})^2 + \epsilon} + (1-\alpha)\sqrt{f(\mathbf{y})^2 + \epsilon)}\right)^2 \\ &= \alpha^2 f(\mathbf{x})^2 + \alpha\epsilon + (1-\alpha)^2 f(\mathbf{y})^2 + (1-\alpha)\epsilon + 2\alpha(1-\alpha)\sqrt{f(\mathbf{x})^2 + \epsilon}\sqrt{f(\mathbf{y})^2 + \epsilon)} \\ &= \alpha^2 f(\mathbf{x})^2 + (1-\alpha)^2 f(\mathbf{y})^2 + 2\alpha(1-\alpha)\sqrt{f(\mathbf{x})^2 + \epsilon}\sqrt{f(\mathbf{y})^2 + \epsilon)} + \epsilon. \end{aligned} \tag{5}$$

Because ϵ is positive, and

$$\begin{aligned}\sqrt{f(\mathbf{x})^2+\epsilon} &\geqslant f(\mathbf{x}) \\ \sqrt{f(\mathbf{y})^2+\epsilon} &\geqslant f(\mathbf{y})\end{aligned}, \tag{6}$$

we have

$$(\alpha h(\mathbf{x})+(1-\alpha)h(\mathbf{y}))^2 \geqslant \left(\sqrt{[\alpha f(\mathbf{x})+(1-\alpha)f(\mathbf{y})]^2+\epsilon}\right)^2, \tag{7}$$

which implies

$$\alpha h(\mathbf{x})+(1-\alpha)h(\mathbf{y}) \geqslant \sqrt{[\alpha f(\mathbf{x})+(1-\alpha)f(\mathbf{y})]^2+\epsilon} \geqslant h(\alpha\mathbf{x}+(1-\alpha)\mathbf{y}). \tag{8}$$

Thus, we conclude that $\sqrt{f(\mathbf{x})^2+\epsilon}$ is a convex function.

Problem 2.11

Because $f(\mathbf{x})$ is Lipschitz with constant L, we have

$$|f(\mathbf{x})-f(\mathbf{y})| \leqslant L\|\mathbf{x}-\mathbf{y}\| \tag{9}$$

for all $\mathbf{x}$ and $\mathbf{y}$ in the domain of f.

First, we show that $|f(\mathbf{x})|$ is Lipschitz with constant L. By reverse triangle inequality, we have

$$||f(\mathbf{x})|-|f(\mathbf{y})|| \leqslant |f(\mathbf{x})-f(\mathbf{y})| \leqslant L\|\mathbf{x}-\mathbf{y}\|. \tag{10}$$

And therefore, $|f(\mathbf{x})|$ is Lipschitz with constant L.

Then, we show that $f(\mathbf{x})_+$ is Lipschitz with constant L. We notice that

$$\begin{aligned}|f(\mathbf{x})_+-f(\mathbf{y})_+| &= |\max\{f(\mathbf{x})_+,0\}-\max\{f(\mathbf{y})_+,0\}| \\ &= \begin{cases}|f(\mathbf{x})| & \text{if } f(\mathbf{x})\geqslant 0 \text{ and } f(\mathbf{y})\leqslant 0 \\ |f(\mathbf{y})| & \text{if } f(\mathbf{x})\leqslant 0 \text{ and } f(\mathbf{y})\geqslant 0 \\ |f(\mathbf{x})-f(\mathbf{y})| & \text{if } f(\mathbf{x})\geqslant 0 \text{ and } f(\mathbf{y})\geqslant 0 \\ 0 & \text{if } f(\mathbf{x})\leqslant 0 \text{ and } f(\mathbf{y})\leqslant 0\end{cases}\end{aligned} \tag{11}$$

For $f(\mathbf{x})\geqslant 0$ and $f(\mathbf{y})\leqslant 0$, $|f(\mathbf{x})| \leqslant |f(\mathbf{x})-f(\mathbf{y})| \leqslant L\|\mathbf{x}-\mathbf{y}\|$.
For $f(\mathbf{x})\leqslant 0$ and $f(\mathbf{y})\geqslant 0$, $|f(\mathbf{y})| \leqslant |f(\mathbf{x})-f(\mathbf{y})| \leqslant L\|\mathbf{x}-\mathbf{y}\|$.
For $f(\mathbf{x})\geqslant 0$ and $f(\mathbf{y})\geqslant 0$, $|f(\mathbf{x})-f(\mathbf{y})| \leqslant L\|\mathbf{x}-\mathbf{y}\|$ by assumption.
For $f(\mathbf{x})\leqslant 0$ and $f(\mathbf{y})\leqslant 0$, obviously $0 \leqslant L\|\mathbf{x}-\mathbf{y}\|$.
Therefore, $f(\mathbf{x})_+$ is Lipschitz with constant L.

To show $\sqrt{f(\mathbf{x})^2+\epsilon}$ is Lipschitz, we notice that for positive ϵ

$$\begin{aligned}&|\sqrt{f(\mathbf{x})^2+\epsilon}-\sqrt{f(\mathbf{y})^2+\epsilon}|^2 \\ &= f(\mathbf{x})^2+f(\mathbf{y})^2+2\epsilon-2\sqrt{f(\mathbf{x})^2+\epsilon}\sqrt{f(\mathbf{y})^2+\epsilon} \\ &\leqslant f(\mathbf{x})^2+f(\mathbf{y})^2+2\epsilon-2(\sqrt{f(\mathbf{x})^2}+\sqrt{\epsilon})(\sqrt{f(\mathbf{y})^2}+\sqrt{\epsilon}) \\ &\leqslant f(\mathbf{x})^2+f(\mathbf{y})^2-2\sqrt{f(\mathbf{x})^2}\sqrt{f(\mathbf{y})^2}-2\sqrt{f(\mathbf{x})^2}\sqrt{\epsilon}-2\sqrt{f(\mathbf{y})^2}\sqrt{\epsilon} \\ &\leqslant f(\mathbf{x})^2+f(\mathbf{y})^2-2\sqrt{f(\mathbf{x})^2}\sqrt{f(\mathbf{y})^2} \\ &= |f(\mathbf{x})-f(\mathbf{y})|^2,\end{aligned} \tag{12}$$

where the first inequality follows from

$$(\sqrt{f(\mathbf{x})^2} + \sqrt{\epsilon})^2 = f(\mathbf{x})^2 + \epsilon + 2\sqrt{f(\mathbf{x})^2}\sqrt{\epsilon} \geqslant f(\mathbf{x})^2 + \epsilon = (\sqrt{f(\mathbf{x})^2 + \epsilon})^2. \tag{13}$$

Therefore,

$$|\sqrt{f(\mathbf{x})^2 + \epsilon} - \sqrt{f(\mathbf{y})^2 + \epsilon}| \leqslant |f(\mathbf{x}) - f(\mathbf{y})| \leqslant L\|\mathbf{x} - \mathbf{y}\|. \tag{14}$$

And so $\sqrt{f(\mathbf{x})^2 + \epsilon}$ is Lipschitz as well.

Problem 2.14

First, notice that

$$\frac{1}{b-a}\int_a^b f(x)dx = \int_a^b f(x)\frac{1}{b-a}dx = \mathrm{E}[f(x)], \tag{15}$$

where $x \sim \text{Uniform}(a, b)$. By the probabilistic version of Jensen's inequality for convex functions, we have

$$f(\mathrm{E}[x]) = f\left(\frac{a+b}{2}\right) \leqslant \mathrm{E}[f(x)] = \frac{1}{b-a}\int_a^b f(x)dx. \tag{16}$$

By the convexity of $f(x)$ we have

$$f(x) \leqslant f(a) + \frac{f(b) - f(a)}{b-a}(x-a), \tag{17}$$

for $x \in [a, b]$, which implies

$$\begin{aligned}
\int_a^b f(x)dx &\leqslant \int_a^b f(a) + \frac{f(b)-f(a)}{b-a}(x-a)dx \\
&= f(a)(b-a) + \frac{1}{2}[f(b) - f(a)](b+a) - a[f(b) - f(a)] \\
&= f(a)(b-a) + \frac{1}{2}[f(b) - f(a)](b-a) \\
&= \frac{1}{2}[f(b) + f(a)](b-a).
\end{aligned} \tag{18}$$

Therefore, we have the inequality

$$f\left(\frac{a+b}{2}\right) \leqslant \frac{1}{b-a}\int_a^b f(x)dx \leqslant \frac{1}{2}[f(b) + f(a)]. \tag{19}$$

Problem 2.15

Let $g(x,y) = f(y)$. For each fixed y, $g(x,y)$ becomes a constant and is convex in x. Then,

$$h(x) = \frac{1}{x}\int_0^x g(x,y)dy = \int_0^x g(x,y)d\mu(y), \tag{20}$$

where $\mu(y) = \frac{1}{x}$ is a measure for $y \sim \text{Uniform}(0,x)$, is a convex function by Proposition 2.3.4. Since $\frac{1}{x}\int_0^x g(x,y)dy = \frac{1}{x}\int_0^x f(y)dy$, we conclude that the running average $\frac{1}{x}\int_0^x f(y)dy$ is also convex.

Problem 2.21

Let $g(\mathbf{x}) = f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$, because $f(\mathbf{x})$ is strongly convex with parameter μ, $g(\mathbf{x})$ is then convex, and satisfies the following inequalities

$$\begin{aligned} f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2 &\geqslant f(\mathbf{y}) - \frac{\mu}{2}\|\mathbf{y}\|^2 + (\nabla f(\mathbf{y}) - \mu\mathbf{y})^*(\mathbf{x}-\mathbf{y}) \\ f(\mathbf{y}) - \frac{\mu}{2}\|\mathbf{y}\|^2 &\geqslant f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2 + (\nabla f(\mathbf{x}) - \mu\mathbf{x})^*(\mathbf{y}-\mathbf{x}). \end{aligned} \tag{21}$$

Summing both sides of the above inequalities gives

$$\begin{aligned} &0 \geqslant (\nabla f(\mathbf{x}) - \mu\mathbf{x})^*(\mathbf{y}-\mathbf{x}) + (\nabla f(\mathbf{y}) - \mu\mathbf{y})^*(\mathbf{x}-\mathbf{y}) \\ &0 \geqslant df(\mathbf{x})(\mathbf{y}-\mathbf{x}) - \mu\mathbf{x}^*(\mathbf{y}-\mathbf{x}) + df(\mathbf{y})(\mathbf{x}-\mathbf{y}) - \mu\mathbf{y}^*(\mathbf{x}-\mathbf{y}) \\ &\mu[\mathbf{x}^*(\mathbf{y}-\mathbf{x}) + \mathbf{y}^*(\mathbf{x}-\mathbf{y})] \geqslant df(\mathbf{x})(\mathbf{y}-\mathbf{x}) + df(\mathbf{y})(\mathbf{x}-\mathbf{y}) \\ &-\mu\|\mathbf{y}-\mathbf{x}\|^2 \geqslant -[df(\mathbf{y}) - df(\mathbf{x})](\mathbf{y}-\mathbf{x}) \\ &[df(\mathbf{y}) - df(\mathbf{x})](\mathbf{y}-\mathbf{x}) \geqslant \mu\|\mathbf{y}-\mathbf{x}\|^2. \end{aligned} \tag{22}$$

To show $d^2f(\mathbf{x}) - \mu\mathbf{I}$ is positive semidefinite, let $\mathbf{y} = \mathbf{x} + t\mathbf{v}$, for sufficiently small t. From the previous inequality

$$[df(\mathbf{x}+t\mathbf{v}) - df(\mathbf{x})]t\mathbf{v} \geqslant \mu t^2\|\mathbf{v}\|^2. \tag{23}$$

Simplifying terms, gives

$$\begin{aligned} &[df(\mathbf{x}+t\mathbf{v}) - df(\mathbf{x})]\mathbf{v} - \mu t\|\mathbf{v}\|^2 \geqslant 0 \\ &(s^2(\mathbf{x}+t\mathbf{v},\mathbf{x})t\mathbf{v})^*\mathbf{v} - \mu t\|\mathbf{v}\|^2 \geqslant 0 \\ &\mathbf{v}^*s^2(\mathbf{x}+t\mathbf{v},\mathbf{x})\mathbf{v} - \mu\|\mathbf{v}\|^2 \geqslant 0 \\ &\mathbf{v}^*[s^2(\mathbf{x}+t\mathbf{v},\mathbf{x}) - \mu\mathbf{I}]\mathbf{v} \geqslant 0. \end{aligned} \tag{24}$$

Sending t to 0, gives

$$\mathbf{v}^*[d^2f(\mathbf{x}) - \mu\mathbf{I}]\mathbf{v} \geqslant 0. \tag{25}$$

In other words, $d^2f(\mathbf{x}) - \mu\mathbf{I}$ is positive semidefinite.

Problem 2.22

Let $g(\mathbf{x}) = f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$. Because $f(\mathbf{x})$ is strongly convex with parameter $\mu > 0$, $g(\mathbf{x})$ is convex and satisfies the inequality

$$g(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leqslant \alpha g(\mathbf{x}) + (1-\alpha)g(\mathbf{y}) \tag{26}$$

for $\alpha \in [0,1]$, which implies

$$\begin{aligned} f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) - \frac{\mu}{2}\|\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\|^2 &\leqslant \alpha[f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2] + (1-\alpha)[f(\mathbf{y}) - \frac{\mu}{2}\|\mathbf{y}\|^2] \\ f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) &\leqslant \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}) + \frac{\mu}{2}[\|\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\|^2 - \alpha\|\mathbf{x}\|^2 - (1-\alpha)\|\mathbf{y}\|^2] \end{aligned} \tag{27}$$

Let $c = \frac{\mu}{2}[\|\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\|^2 - \alpha\|\mathbf{x}\|^2 - (1-\alpha)\|\mathbf{y}\|^2]$. Because the function $\mathbf{x}^*\mathbf{x}$ has positive definite second differential $2\mathbf{I}$, it's strictly convex. Therefore, for $x \neq y$ and μ positive, $c < 0$, which implies the strict inequality

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) < \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{y}). \tag{28}$$

In other words, $f(\mathbf{x})$ is strictly convex.

To show $f(\mathbf{x})$ has a unique global minimum, we use proof by contradiction. Assume there exist two local minimum $\mathbf{x}_1$ and $\mathbf{x}_2$ ($\mathbf{x}_1 \neq \mathbf{x}_2$). Without loss of generality, let $f(\mathbf{x}_1) \leqslant f(\mathbf{x}_2)$. By strict convexity of $f(\mathbf{x})$, we have

$$\begin{aligned} f(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) &< \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \\ &\leqslant \alpha f(\mathbf{x}_2) + (1-\alpha)f(\mathbf{x}_2) = f(\mathbf{x}_2). \end{aligned} \tag{29}$$

For α sufficiently small, the above inequality contradicts the assumption that $\mathbf{x}_2$ is a local minimum. Therefore, there exists a unique global minimum for $f(\mathbf{x})$.

To show $f(\mathbf{x}) \geqslant f(\mathbf{y}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{y}\|^2$, we apply the supporting hyperplane inequality on $f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2$,

$$\begin{aligned} f(\mathbf{x}) - \frac{\mu}{2}\|\mathbf{x}\|^2 &\geqslant f(\mathbf{y}) - \frac{\mu}{2}\|\mathbf{y}\|^2 + (\nabla f(\mathbf{y}) - \mu\mathbf{y})^*(\mathbf{x} - \mathbf{y}) \\ f(\mathbf{x}) &\geqslant f(\mathbf{y}) + \frac{\mu}{2}(\mathbf{x}^*\mathbf{x} - \mathbf{y}^*\mathbf{y}) + df(\mathbf{y})(\mathbf{x} - \mathbf{y}) - \mu\mathbf{y}^*(\mathbf{x} - \mathbf{y}) \\ &= f(\mathbf{y}) + \frac{\mu}{2}(\mathbf{x}^*\mathbf{x} - \mathbf{y}^*\mathbf{y} - 2\mathbf{x}^*\mathbf{y} + 2\mathbf{y}^*\mathbf{y}) + df(\mathbf{y})(\mathbf{x} - \mathbf{y}) \\ &= f(\mathbf{y}) + \frac{\mu}{2}(\mathbf{x}^*\mathbf{x} + \mathbf{y}^*\mathbf{y} - 2\mathbf{x}^*\mathbf{y}) + df(\mathbf{y})(\mathbf{x} - \mathbf{y}) \\ &= f(\mathbf{y}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{y}\|^2 + df(\mathbf{y})(\mathbf{x} - \mathbf{y}). \end{aligned} \tag{30}$$

For the stationary point $\mathbf{y}$, $df(\mathbf{y}) = 0$. Thus the inequality $f(\mathbf{x}) \geqslant f(\mathbf{y}) + \frac{\mu}{2}\|\mathbf{x} - \mathbf{y}\|^2$.

Problem 2.23

Assume $f(\mathbf{x})$ is strongly convex with parameter μ. Let $g(\mathbf{x}) = f(\mathbf{x}) - \mathbf{y}^*\mathbf{x} - \frac{\mu}{2}\|\mathbf{x}\|^2$. Then, by the strong convexity of $f(\mathbf{x})$,

$$\begin{aligned}
&g(\alpha\mathbf{x} + (1-\alpha)\mathbf{z}) \\
&= f(\alpha\mathbf{x} + (1-\alpha)\mathbf{z}) - \mathbf{y}^*(\alpha\mathbf{x} + (1-\alpha)\mathbf{z}) - \frac{\mu}{2}\|\alpha\mathbf{x} + (1-\alpha)\mathbf{z}\|^2 \\
&\leqslant \alpha f(\mathbf{x}) + (1-\alpha)f(\mathbf{z}) - (\alpha\mathbf{y}^*\mathbf{x} + (1-\alpha)\mathbf{y}^*\mathbf{z}) - \frac{\mu}{2}\alpha\mathbf{x} - \frac{\mu}{2}(1-\alpha)\mathbf{z} \\
&= \alpha g(\mathbf{x}) + (1-\alpha)g(\mathbf{z}).
\end{aligned} \tag{31}$$

Therefore, $g(\mathbf{x})$ is convex, and $f(\mathbf{x}) - \mathbf{y}^*\mathbf{x}$ is strongly convex. From problem 22, we know that $h(\mathbf{x}) = f(\mathbf{x}) - \mathbf{y}^*\mathbf{x}$ possesses a unique global minimum. Since $h(\mathbf{x})$ is differentiable, the global minimum occurs at the stationary point at which

$$\nabla h(\mathbf{x}) = \nabla f(\mathbf{x}) - \mathbf{y} = 0. \tag{32}$$

Therefore, the equation $\nabla f(\mathbf{x}) = \mathbf{y}$ is unique solvable for all $\mathbf{y}$.

Problem 2.26

First, we show that C is a convex set. Let $\mathbf{x} = \sum_{i=1}^m a_i\mathbf{u}_i$, $a_i \geqslant 0$ and $\mathbf{y} = \sum_{i=1}^m b_i\mathbf{u}_i$, $b_i \geqslant 0$, i.e. $\mathbf{x} \in C$ and $\mathbf{y} \in C$. Let $\mathbf{z} = \gamma\mathbf{x} + (1-\gamma)\mathbf{y}$, where $\gamma \in [0,1]$. Then

$$\mathbf{z} = \sum_{i=1}^m c_i\mathbf{u}_i = \sum_{i=1}^m [\gamma a_i + (1-\gamma)b_i]\mathbf{u}_i, \tag{33}$$

where $c_i = \gamma a_i + (1-\gamma)b_i \geqslant 0$. Thus, $\mathbf{z} \in C$, and C is a convex set.

To show the set C is closed, we first assume the vectors $\mathbf{u}_i$ are linearly independent. Then, for each $\mathbf{z}_j \in C$, we can represent it as $\sum_{i=1}^m c_{ji}\mathbf{u}_i$, where the coefficients $c_{ji} \geqslant 0$, $i = \{1, \ldots, m\}$ are unique for each $\mathbf{z}_j$. Let $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$ be the matrix where $\mathbf{u}_i$ are the columns of $\mathbf{U}$. Then, the coefficient vector $\mathbf{c}_j = (c_1, \ldots, c_m)$ that constructs $\mathbf{z}_j$ from $\mathbf{u}_i$ can be uniquely represented as

$$\mathbf{c}_j = (\mathbf{U}^*\mathbf{U})^{-1}\mathbf{U}^*\mathbf{z}_j, \tag{34}$$

which follows from $\mathbf{z}_j = \mathbf{U}\mathbf{c}_j$ and $\mathbf{U}^*\mathbf{z}_j = \mathbf{U}^*\mathbf{U}\mathbf{c}_j$. Now, assume the sequence $\mathbf{z}_j \in C$ converge to a point $\mathbf{z}$. By Equation (34), the sequence $\mathbf{c}_j$ converges to $\mathbf{c}$ with all entries positive. Therefore, the point $\mathbf{z} = \mathbf{U}\mathbf{c} \in C$. And thus, the set C is closed.

When the vectors $\mathbf{u}_i$ are linearly dependent, there exists $\beta = (\beta_1, \ldots, \beta_m)$ such that not all β_i are 0 and $\sum_{i=1}^m \beta_i\mathbf{u}_i = 0$. Then the point $\mathbf{z} \in C$ can be expressed as

$$\mathbf{z} = \sum_{i=1}^m c_i\mathbf{u}_i = \sum_{i=1}^m c_i\mathbf{u}_i + t\sum_{i=1}^m \beta_i\mathbf{u}_i = \sum_{i=1}^m (c_i + t\beta_i)\mathbf{u}_i. \tag{35}$$

By taking the smallest $|t|$, we can render $c_j + t\beta_j = 0$ for the j-th coefficient, while keeping all other coefficients non-negative. Then $\mathbf{z}$ can be expressed as $\mathbf{z} = \sum_{i=1,i\neq j}^{m}(c_i + t\beta_i)\mathbf{u}_i$, which implies

$$C = \bigcup_{j=1}^{m} \left\{ \sum_{i=1,i\neq j}^{m} a_i \mathbf{u}_i : \; i \neq j, \; a_i \geqslant 0. \right\} \tag{36}$$

In other words, C can be expressed as a union of the span of the linearly independent subset of $\mathbf{u}_i$ by non-negative coefficients. Because the sets $\left\{\sum_{i=1,i\neq j}^{m} a_i \mathbf{u}_i : \; i \neq j, \; a_i \geqslant 0\right\}$ are closed and convex, the finite union of them is also closed and convex.

The above proof assume the matrix $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$ has rank $m-1$. By recursively applying the same argument in Equation (35), one can generalize the proof for $\mathbf{U}$ with any rank.

Problem 2.29

Finding the projection of $\mathbf{y}$ onto the set U_k^n is equivalent to finding $\underset{\mathbf{z}}{\operatorname{argmin}} \|\mathbf{z} - \mathbf{y}\|$ with the constraint $\mathbf{z} \in U_k^n$. To show that the projection can be achieved by replacing the k largest entries by 1 and the remaining entries by 0, we use proof by induction.

Base case:

For the case when $n = 1$ and $k = 0$, $U_k^n = \{0\}$, and the projection of $\mathbf{y}$ onto U_k^n is indeed 0. For test case when $n = 1$ and $k = 1$, $U_k^n = \{1\}$, and the projection of $\mathbf{y}$ onto U_k^n is indeed 1.

Induction assumption:

Assume for n and k, we can obtain the projection by setting the k largest entries of $\mathbf{y}$ to 1 and the remaining to 0.

Induction k to $k+1$:

We first show that for n, the projection algorithm is correct as k increases to $k+1$.

Without loss of generality, assume $\mathbf{y} = (y_1, \ldots, y_n)$ with $y_1 \geqslant \ldots \geqslant y_n$. Then by assumption $(\mathbf{1}_k, \mathbf{0}_{n-k})$ minimizes $\|\mathbf{z} - \mathbf{y}\|$ with the constraint $\mathbf{z} \in U_k^n$, with distance

$$\|(\mathbf{1}_k, \mathbf{0}_{n-k}) - \mathbf{y}\|^2 = \sum_{i=1}^{k}(1 - y_i)^2 + \sum_{i=k+1}^{n} y_i^2. \tag{37}$$

For $k+1$, the projection algorithm yields $(\mathbf{1}_{k+1}, \mathbf{0}_{n-k-1})$ with

$$\begin{aligned}
\|(\mathbf{1}_{k+1}, \mathbf{0}_{n-k-1}) - \mathbf{y}\|^2 &= \sum_{i=1}^{k+1}(1 - y_i)^2 + \sum_{i=k+2}^{n} y_i^2 \\
&= \sum_{i=1}^{k}(1 - y_i)^2 + \sum_{i=k+1}^{n} y_i^2 + [(1 - y_{k+1})^2 - y_{k+1}^2] \qquad (38) \\
&= \sum_{i=1}^{k}(1 - y_i)^2 + \sum_{i=k+1}^{n} y_i^2 - [2y_{k+1} - 1]
\end{aligned}$$

From Equation (40), it's clear that the entry y_{k+1} results in the largest reduction in distance between $\mathbf{z}$ and $\mathbf{y}$ from $\|(\mathbf{1}_k, \mathbf{0}_{n-k}) - \mathbf{y}\|^2$. Therefore, for n and $k+1$, setting the $k+1$ largest entries of $\mathbf{y}$ to 1 and the remaining to 0 minimizes $\|\mathbf{z} - \mathbf{y}\|$.

Induction n **to** $n+1$:

Then we show that for k, the projection algorithm is correct as n increases to $n+1$.

As shown previously, by assumption $(\mathbf{1}_k, \mathbf{0}_{n-k})$ minimizes $\|\mathbf{z}-\mathbf{y}\|$ with the constraint $\mathbf{z} \in U_k^n$, with distance

$$\|(\mathbf{1}_k, \mathbf{0}_{n-k}) - \mathbf{y}\|^2 = \sum_{i=1}^{k}(1 - y_i)^2 + \sum_{i=k+1}^{n} y_i^2. \tag{39}$$

With k fixed and n increasing to $n+1$, the algorithm yields $(\mathbf{1}_k, \mathbf{0}_{n+1-k})$, with distance

$$\|(\mathbf{1}_k, \mathbf{0}_{n+1-k}) - \mathbf{y}\|^2 = \sum_{i=1}^{k}(1 - y_i)^2 + \sum_{i=k+1}^{n} y_i^2 + y_{n+1}^2 \tag{40}$$

Clearly, y_{n+1} yields the smallest increment in $\|(\mathbf{1}_k, \mathbf{0}_{n+1-k}) - \mathbf{y}\|^2$. So the algorithm is correct when k is fixed and n increases to $n+1$.

Conclusion:

Having shown the induction in the k and the n direction, we can conclude that projection of $\mathbf{y}$ onto U_k^n replaces the k largest entries of $\mathbf{y}$ by 1 and the remaining entries by 0.

Problem 2.37

$$\|\mathbf{x}\|_1 \|\mathbf{y}\|_\infty = \max_i |y_i| \sum_i |x_i| \geqslant \sum_i |x_i||y_i| \geqslant \sum_i x_i y_i = \mathbf{x}^* \mathbf{y} \tag{41}$$

Equality holds when $x_i y_i \geqslant 0$ for all i and $y_i = c$ for all i.

Problem 2.40

From Von Neumann-Fan inequality, it follows that

$$\text{tr } \mathbf{A} = \sum_{i=1}^{n} \lambda_i \text{ , tr } \mathbf{A}^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i}, \tag{42}$$

by letting $\mathbf{B} = \mathbf{I}_n$ in the inequality (2.16) and then applying the equality condition. Then tr $\mathbf{A}$ + tr $\mathbf{A}^{-1} = \sum_{i=1}^{n} \lambda_i + 1/\lambda_i$. The minimum of $\lambda_i + 1/\lambda_i$ can be found by setting the derivative $1 - \lambda_i^{-2}$ to 0, from which we get $\lambda_i = 1$, and $\lambda_i + 1/\lambda_i = 2$. Therefore, tr $\mathbf{A}$ + tr $\mathbf{A}^{-1} \geqslant 2n$. Equality is attained when $\lambda_i = 1$ for $i = 1, \ldots, n$, which implies $\mathbf{A} = \mathbf{I}_n$.