Optimization Methods **Fall 2023**

Homework 2

Instructor: Lijun Zhang *Name:* Ji Qiankun, *StudentId:* 221300066

Notice

- The submission email is: optfall2023@163.com.
- Please use the provided LaTeX file as a template.
- If you are not familiar with LaTeX, you can also use Word to generate a **PDF** file.

Problem 1: Convexity

a) Suppose $f : \mathbb{R} \to \mathbb{R}$ is convex and differentiable. Show that its running average F, i.e.,

$$F(x) = \frac{1}{x}\int_0^x f(t)dt$$

is convex over $\mathbf{dom}(F) = \mathbb{R}_{++}$.

b) Suppose f and g are both convex, nondecreasing (or nonincreasing), and positive real-valued functions defined on $\mathbb{R}$, prove that fg is convex on $\mathbf{dom}(f) \cap \mathbf{dom}(g)$.

c) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Its perspective transform $g : \mathbb{R}^{n+1} \to \mathbb{R}$ is defined by

$$g(x,t) = tf(\frac{x}{t}),$$

with domain $\mathbf{dom}(g) = \{(x,t) \in \mathbb{R}^{n+1} : x \in \mathbf{dom}(f), t > 0\}$. Use the definition of convexity to prove that if f is convex, then so is its perspective transform g.

Solution. a) To show that F is convex, we need to show that for any $x, y \in \mathbf{dom}(F)$ and $\lambda \in [0,1]$, we have

$$F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y)$$

Using the definition of F, we can get

$$\begin{aligned}
F(\lambda x + (1-\lambda)y) &= \frac{1}{\lambda x + (1-\lambda)y}\int_0^{\lambda x + (1-\lambda)y} f(t)dt \\
&= \frac{1}{\lambda x + (1-\lambda)y}\left(\int_0^{\lambda x} f(t)dt + \int_{\lambda x}^{\lambda x + (1-\lambda)y} f(t)dt\right) \\
&= \frac{\lambda x}{\lambda x + (1-\lambda)y}\frac{1}{\lambda x}\int_0^{\lambda x} f(t)dt + \frac{(1-\lambda)y}{\lambda x + (1-\lambda)y}\frac{1}{(1-\lambda)y}\int_{\lambda x}^{\lambda x + (1-\lambda)y} f(t)dt \\
&= \frac{\lambda x}{\lambda x + (1-\lambda)y}F(x) + \frac{(1-\lambda)y}{\lambda x + (1-\lambda)y}F(y) \\
&\leq \lambda F(x) + (1-\lambda)F(y)
\end{aligned}$$

where the last inequality follows from the fact that $\frac{\lambda x}{\lambda x + (1-\lambda)y} \leq \lambda$ and $\frac{(1-\lambda)y}{\lambda x + (1-\lambda)y} \leq (1-\lambda)$, since $x, y > 0$ and $\lambda \in [0,1]$. Therefore, F is convex.

b) To show that fg is convex, we need to show that for any $x, y \in \mathbf{dom}(f) \cap \mathbf{dom}(g)$ and $\lambda \in [0,1]$, we have

$$(fg)(\lambda x + (1-\lambda)y) \leq \lambda(fg)(x) + (1-\lambda)(fg)(y)$$

Using the fact that f and g are convex, we have

$$\begin{aligned}(fg)(\lambda x+(1-\lambda)y) &= f(\lambda x+(1-\lambda)y)g(\lambda x+(1-\lambda)y)\\ &\leq (\lambda f(x)+(1-\lambda)f(y))\left(\lambda g(x)+(1-\lambda)g(y)\right)\\ &= \lambda^2 f(x)g(x)+\lambda(1-\lambda)f(x)g(y)+\lambda(1-\lambda)f(y)g(x)+(1-\lambda)^2 f(y)g(y)\end{aligned}$$

Now, if f and g are both nondecreasing, then we have $f(x)g(y)\leq f(y)g(y)$ and $f(y)g(x)\leq f(y)g(y)$, so we can get

$$\begin{aligned}(fg)(\lambda x+(1-\lambda)y) &\leq \lambda^2 f(x)g(x)+2\lambda(1-\lambda)f(y)g(y)+(1-\lambda)^2 f(y)g(y)\\ &= \lambda f(x)g(x)+(1-\lambda)f(y)g(y)\\ &= \lambda(fg)(x)+(1-\lambda)(fg)(y)\end{aligned}$$

Similarly, if f and g are both nonincreasing, then we have $f(x)g(y)\geq f(y)g(y)$ and $f(y)g(x)\geq f(y)g(y)$, so we can get

$$\begin{aligned}(fg)(\lambda x+(1-\lambda)y) &\leq \lambda^2 f(x)g(x)+2\lambda(1-\lambda)f(x)g(x)+(1-\lambda)^2 f(y)g(y)\\ &= \lambda f(x)g(x)+(1-\lambda)f(y)g(y)\\ &= \lambda(fg)(x)+(1-\lambda)(fg)(y)\end{aligned}$$

Therefore, in either case, fg is convex.
c) To show that g is convex, we need to show that for any $(x,t),(y,s)\in \mathbf{dom}(g)$ and $\lambda\in[0,1]$, we have

$$g(\lambda(x,t)+(1-\lambda)(y,s))\leq \lambda g(x,t)+(1-\lambda)g(y,s)$$

Using the definition of g, we can get

$$\begin{aligned}g(\lambda(x,t)+(1-\lambda)(y,s)) &= (\lambda t+(1-\lambda)s)f\left(\frac{\lambda x+(1-\lambda)y}{\lambda t+(1-\lambda)s}\right)\\ &= (\lambda t+(1-\lambda)s)f\left(\frac{\lambda t}{\lambda t+(1-\lambda)s}\frac{x}{t}+\frac{(1-\lambda)s}{\lambda t+(1-\lambda)s}\frac{y}{s}\right)\\ &\leq (\lambda t+(1-\lambda)s)\left(\frac{\lambda t}{\lambda t+(1-\lambda)s}f\left(\frac{x}{t}\right)+\frac{(1-\lambda)s}{\lambda t+(1-\lambda)s}f\left(\frac{y}{s}\right)\right)\\ &= \lambda t f\left(\frac{x}{t}\right)+(1-\lambda)sf\left(\frac{y}{s}\right)\\ &= \lambda g(x,t)+(1-\lambda)g(y,s)\end{aligned}$$

where the inequality follows from the fact that f is convex and $\frac{\lambda t}{\lambda t+(1-\lambda)s},\frac{(1-\lambda)s}{\lambda t+(1-\lambda)s}\in[0,1]$. Therefore, g is convex.

$\square$

Problem 2: Convex Functions

a) Let A is a positive semidefinite symmetric $n\times n$ matrix and β is a positive scalar. Prove that the function $f:\mathbb{R}^n\mapsto\mathbb{R}$, defined as
$$f(x)=e^{x^\top Ax},$$
is convex.

b) Show that the function
$$f(x,u,v)=-\log\left(uv-x^Tx\right)$$
is convex over $\mathbf{dom}f=\left\{(x,u,v)\mid uv>x^Tx, u,v>0\right\}$ (*Hint: consider the function* $f(x,u)=x^\top x/u$)

c) Suppose $A\in\mathbb{R}^{n\times n}, b\in\mathbb{R}^n$. Show that the function
$$f(x)=\frac{\|Ax-b\|_2^2}{1-x^\top x}$$
is convex on $\mathbf{dom}(f)=\{x\in\mathbb{R}^n|\|x\|_2\leq 1\}$.

Solution. a) To show that f is convex, we need to show that its Hessian matrix is positive semidefinite, i.e., for any $z \in \mathbb{R}^n$, we have

$$z^\top \nabla^2 f(x) z \geq 0$$

Using the chain rule, compute the gradient and the Hessian of f as follows:

$$\begin{aligned}
\nabla f(x) &= f(x)\nabla(x^\top Ax) \\
&= 2f(x)Ax \\
\nabla^2 f(x) &= 2f(x)\nabla(Ax) + 2f(x)Ax\nabla^\top(Ax) \\
&= 4f(x)A + 4f(x)Axx^\top A
\end{aligned}$$

Then, we have

$$\begin{aligned}
z^\top \nabla^2 f(x) z &= 4f(x)z^\top Az + 4f(x)z^\top Axx^\top Az \\
&= 4f(x)\left(z^\top Az + (x^\top Az)^2\right) \\
&\geq 0
\end{aligned}$$

where the last inequality follows from the fact that $f(x) > 0$ and A is positive semidefinite, so $z^\top Az \geq 0$ and $(x^\top Az)^2 \geq 0$. Therefore, f is convex.
b) To show that f is convex, we need to show that its Hessian matrix is positive semidefinite, i.e., for any $z \in \mathbb{R}^{n+2}$, we have

$$z^\top \nabla^2 f(x,u,v) z \geq 0$$

Using the chain rule, compute the gradient and the Hessian of f as follows:

$$\begin{aligned}
\nabla f(x,u,v) &= -\frac{1}{uv - x^\top x}\nabla(uv - x^\top x) \\
&= -\frac{1}{uv - x^\top x}\begin{bmatrix} -2x \\ v \\ u \end{bmatrix} \\
\nabla^2 f(x,u,v) &= -\frac{1}{uv - x^\top x}\nabla\begin{bmatrix} -2x \\ v \\ u \end{bmatrix} + \frac{1}{(uv - x^\top x)^2}\nabla(uv - x^\top x)\nabla^\top(uv - x^\top x) \\
&= -\frac{1}{uv - x^\top x}\begin{bmatrix} -2I & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} + \frac{1}{(uv - x^\top x)^2}\begin{bmatrix} -2x \\ v \\ u \end{bmatrix}\begin{bmatrix} -2x \\ v \\ u \end{bmatrix}^\top \\
&= \frac{1}{uv - x^\top x}\begin{bmatrix} 2I + \frac{4xx^\top}{uv - x^\top x} & -\frac{2x}{uv - x^\top x} & -\frac{2x}{uv - x^\top x} \\ -\frac{2x}{uv - x^\top x} & 1 + \frac{v^2}{uv - x^\top x} & \frac{u^2 - v^2}{uv - x^\top x} \\ -\frac{2x}{uv - x^\top x} & \frac{u^2 - v^2}{uv - x^\top x} & 1 + \frac{u^2}{uv - x^\top x} \end{bmatrix}
\end{aligned}$$

Then, we have

$$\begin{aligned}
z^\top \nabla^2 f(x,u,v) z &= \frac{1}{uv - x^\top x}\left(2z_1^\top z_1 + \frac{4(z_1^\top x)^2}{uv - x^\top x} - \frac{4z_1^\top x(z_2 + z_3)}{uv - x^\top x} + z_2^2 + \frac{v^2 z_2^2}{uv - x^\top x} + z_3^2 + \frac{u^2 z_3^2}{uv - x^\top x} + 2\frac{u^2 - v^2}{uv - x^\top x}z_2 z_3\right) \\
&= \frac{1}{uv - x^\top x}\left(2\|z_1\|^2 + \frac{(z_2 + z_3)^2}{uv - x^\top x}\left((uv - x^\top x)^2 - 4(x^\top z_1)^2\right) + \frac{u^2 + v^2}{uv - x^\top x}(z_2^2 + z_3^2) + 2\frac{u^2 - v^2}{uv - x^\top x}z_2 z_3\right)
\end{aligned}$$

where $z_1 \in \mathbb{R}^n$ and $z_2, z_3 \in \mathbb{R}$ are the components of z. Now, we can use the following facts to show that the expression inside the parentheses is nonnegative: - $2\|z_1\|^2 \geq 0$ since z_1 is a vector. - $(uv - x^\top x)^2 - 4(x^\top z_1)^2 \geq 0$ since it is equivalent to $(uv - x^\top x - 2x^\top z_1)(uv - x^\top x + 2x^\top z_1) \geq 0$, which holds because $uv - x^\top x > 0$ and $uv - x^\top x \pm 2x^\top z_1$ have the same sign for any z_1. - $\frac{u^2+v^2}{uv - x^\top x}(z_2^2 + z_3^2) + 2\frac{u^2 - v^2}{uv - x^\top x}z_2 z_3 \geq 0$ since it is equivalent to $\left(\frac{uz_2 + vz_3}{\sqrt{uv - x^\top x}}\right)^2 \geq 0$, which holds because $u, v > 0$ and $uv - x^\top x > 0$.

Therefore, we have

$$z^\top \nabla^2 f(x,u,v) z \geq 0$$

for any $z \in \mathbb{R}^{n+2}$, which implies that f is convex.
c) To show that f is convex, we need to show that its Hessian matrix is positive semidefinite, i.e., for any $z \in \mathbb{R}^n$, we have

$$z^\top \nabla^2 f(x) z \geq 0$$

Using the chain rule, compute the gradient and the Hessian of f as follows:

$$\begin{aligned}
\nabla f(x) &= \frac{1}{1-x^\top x}\nabla(\|Ax-b\|_2^2) - \frac{\|Ax-b\|_2^2}{(1-x^\top x)^2}\nabla(x^\top x) \\
&= \frac{2}{1-x^\top x}(A^\top Ax - A^\top b) - \frac{2\|Ax-b\|_2^2}{(1-x^\top x)^2}x \\
\nabla^2 f(x) &= \frac{2}{1-x^\top x}\nabla(A^\top Ax - A^\top b) - \frac{2\|Ax-b\|_2^2}{(1-x^\top x)^2}\nabla x - \frac{2}{(1-x^\top x)^2}(A^\top Ax - A^\top b)\nabla^\top(x^\top x) \\
&= \frac{2}{1-x^\top x}A^\top A - \frac{2\|Ax-b\|_2^2}{(1-x^\top x)^2}I - \frac{4}{(1-x^\top x)^2}(A^\top Ax - A^\top b)x^\top - \frac{4\|Ax-b\|_2^2}{(1-x^\top x)^3}xx^\top
\end{aligned}$$

Then, we have

$$\begin{aligned}
z^\top \nabla^2 f(x) z &= \frac{2}{1-x^\top x}z^\top A^\top Az - \frac{2\|Ax-b\|_2^2}{(1-x^\top x)^2}z^\top z - \frac{4}{(1-x^\top x)^2}z^\top(A^\top Ax - A^\top b)x^\top z - \frac{4\|Ax-b\|_2^2}{(1-x^\top x)^3}z^\top xx^\top z \\
&= \frac{2}{1-x^\top x}\|Az\|_2^2 - \frac{2\|Ax-b\|_2^2}{(1-x^\top x)^2}\|z\|_2^2 - \frac{4}{(1-x^\top x)^2}(z^\top A^\top Ax - z^\top A^\top b)z^\top x - \frac{4\|Ax-b\|_2^2}{(1-x^\top x)^3}(z^\top x)^2
\end{aligned}$$

Now, use the following facts to show that the expression is nonnegative: - $\frac{2}{1-x^\top x}\|Az\|_2^2 \geq 0$ since A is a matrix and z is a vector. - $\frac{2\|Ax-b\|_2^2}{(1-x^\top x)^2}\|z\|_2^2 \geq 0$ since $\|Ax-b\|_2^2 \geq 0$, $\|z\|_2^2 \geq 0$, and $(1-x^\top x)^2 > 0$ for $\|x\|_2 \leq 1$. - $\frac{4}{(1-x^\top x)^2}(z^\top A^\top Ax - z^\top A^\top b)z^\top x \geq 0$ since $(z^\top A^\top Ax - z^\top A^\top b)z^\top x = z^\top A^\top(Ax-b)x \leq 0$ by the Cauchy-Schwarz inequality, with equality if and only if $Ax-b$ and x are linearly dependent, and $(1-x^\top x)^2 > 0$ for $\|x\|_2 \leq 1$. - $\frac{4\|Ax-b\|_2^2}{(1-x^\top x)^3}(z^\top x)^2 \geq 0$ since $\|Ax-b\|_2^2 \geq 0$, $(z^\top x)^2 \geq 0$, and $(1-x^\top x)^3 > 0$ for $\|x\|_2 \leq 1$.

Therefore, we have

$$z^\top \nabla^2 f(x) z \geq 0$$

for any $z \in \mathbb{R}^n$, which implies that f is convex.

□

Problem 3: Concave Function

Show that the function

$$f(x) = \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}}$$

with $\mathbf{dom}(f) = \mathbb{R}^n_{++}$ is concave.

Solution. My idea to show that f is concave is to use the arithmetic-geometric mean inequality, which states that for any positive numbers $x_1, x_2, \ldots, x_n$, we have

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq (x_1 x_2 \ldots x_n)^{\frac{1}{n}}$$

with equality if and only if $x_1 = x_2 = \cdots = x_n$. Taking the logarithm of both sides, we get

$$\log\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) \geq \frac{1}{n}\log(x_1 x_2 \ldots x_n)$$

with equality if and only if $x_1 = x_2 = \cdots = x_n$. Now, let $x, y \in \mathbf{dom}(f)$ and $\lambda \in [0,1]$. Then, we have

$$
\begin{aligned}
\log f(\lambda x + (1-\lambda)y) &= \frac{1}{n} \log \left(\prod_{i=1}^{n} (\lambda x_i + (1-\lambda) y_i) \right) \\
&\leq \frac{1}{n} \log \left(\frac{\lambda \sum_{i=1}^{n} x_i + (1-\lambda) \sum_{i=1}^{n} y_i}{n} \right) \\
&= \frac{1}{n} \log \left(\lambda \frac{\sum_{i=1}^{n} x_i}{n} + (1-\lambda) \frac{\sum_{i=1}^{n} y_i}{n} \right) \\
&\leq \frac{1}{n} \left(\lambda \log \left(\frac{\sum_{i=1}^{n} x_i}{n} \right) + (1-\lambda) \log \left(\frac{\sum_{i=1}^{n} y_i}{n} \right) \right) \\
&= \lambda \frac{1}{n} \log \left(\prod_{i=1}^{n} x_i \right) + (1-\lambda) \frac{1}{n} \log \left(\prod_{i=1}^{n} y_i \right) \\
&= \lambda \log f(x) + (1-\lambda) \log f(y)
\end{aligned}
$$

where the first inequality follows from the arithmetic-geometric mean inequality applied to each term $\lambda x_i + (1-\lambda) y_i$, and the second inequality follows from the concavity of the logarithm function. Taking the exponential of both sides, we get

$$
f(\lambda x + (1-\lambda)y) \leq f(x)^{\lambda} f(y)^{1-\lambda}
$$

which shows that f is concave.

$\square$

Problem 4: Conjugate Function

a) Prove that the conjugate of the conjugate of a closed convex function is itself, i.e., $f^{**} = f$, if f is closed and convex.

b) Show that the conjugate of $f(x) = \max_{i=1,\cdots,n} x_i$ over $\mathbb{R}^n$.

c) Show that the conjugate of $f(x) = x^p$ over $\mathbb{R}_{++}$, where $p > 1$.

Solution. a)To prove that conjugate of the conjugate of a closed convex function is itself, we need to show that for any $x \in \mathbf{dom}(f)$, we have

$$
f^{**}(x) = f(x)
$$

Using the definition of the conjugate function, we can get

$$
\begin{aligned}
f^{**}(x) &= \sup_{y \in \mathbb{R}^n} (y^\top x - f^*(y)) \\
&= \sup_{y \in \mathbb{R}^n} (y^\top x - \sup_{z \in \mathbb{R}^n} (z^\top y - f(z))) \\
&= \sup_{y \in \mathbb{R}^n} \inf_{z \in \mathbb{R}^n} (y^\top x - z^\top y + f(z))
\end{aligned}
$$

Now, use the following facts to simplify the expression:
- The supremum and the infimum can be interchanged by using the min-max theorem.
- The infimum over z is attained at $z = x$.
- The supremum over y is attained at $y = \nabla f(x)$.

Therefore, we have

$$\begin{aligned}
f^{**}(x) &= \sup_{y\in\mathbb{R}^n} \inf_{z\in\mathbb{R}^n} (y^\top x - z^\top y + f(z)) \\
&= \inf_{z\in\mathbb{R}^n} \sup_{y\in\mathbb{R}^n} (y^\top x - z^\top y + f(z)) \\
&= \sup_{y\in\mathbb{R}^n} (y^\top x - x^\top y + f(x)) \\
&= \sup_{y\in\mathbb{R}^n} (y - \nabla f(x))^\top x + f(x) \\
&= f(x) + \sup_{y\in\mathbb{R}^n} (y - \nabla f(x))^\top x \\
&= f(x) + \|x\|^2 \sup_{\|y\|=1} (y - \nabla f(x)/\|x\|)^\top x/\|x\| \\
&= f(x) + \|x\|^2 \sup_{\|y\|=1} (y^\top x/\|x\| - \nabla f(x)^\top x/\|x\|^2) \\
&= f(x) + \|x\|^2 \sup_{\|y\|=1} (y^\top x/\|x\| - 1) \\
&= f(x) + \|x\|^2(\|x\|/\|x\| - 1) \\
&= f(x)
\end{aligned}$$

where the second last equality follows from the fact that $y^\top x/\|x\| \le \|y\|\|x\|/\|x\| = 1$ by the Cauchy-Schwarz inequality, with equality if and only if $y = x/\|x\|$. Therefore, we have proved that $f^{**} = f$so conjugate of the conjugate of a closed convex function is itself.

b)To show that the conjugate of $f(x) = \max_{i=1,\cdots,n} x_i$ over $\mathbb{R}^n$ is

$$f^*(y) = \begin{cases} 0 & \text{if } y \in \Delta_n \\ \infty & \text{otherwise} \end{cases}$$

where $\Delta_n = \{y \in \mathbb{R}^n : y_i \ge 0, \sum_{i=1}^n y_i = 1\}$ is the standard simplex, we need to use the definition of the conjugate function, which is

$$f^*(y) = \sup_{x\in\mathbb{R}^n} (y^\top x - f(x))$$

Now, use the following facts to find the expression for $f^*(y)$:

- The function $y^\top x - f(x)$ is linear in x, so it is unbounded above unless $y \in \Delta_n$, in which case it is constant and equal to zero.
- The supremum over x is either zero or infinity, depending on whether $y \in \Delta_n$ or not.

Therefore, we have

$$f^*(y) = \begin{cases} 0 & \text{if } y \in \Delta_n \\ \infty & \text{otherwise} \end{cases}$$

which is the conjugate of $f(x) = \max_{i=1,\cdots,n} x_i$ over $\mathbb{R}^n$.

c)To show that the conjugate of $f(x) = x^p$ over $\mathbb{R}_{++}$, where $p > 1$, is

$$f^*(y) = \begin{cases} \left(\frac{y}{p-1}\right)^{\frac{p}{p-1}} & \text{if } y \ge 0 \\ \infty & \text{otherwise} \end{cases}$$

Using the definition of the conjugate function, which is

$$f^*(y) = \sup_{x\in\mathbb{R}_{++}} (yx - x^p)$$

Now, we can use the following facts to find the expression for $f^*(y)$:

- The function $yx - x^p$ is concave in x.
- The global maximum value is $yx - x^p = \left(\frac{y}{p}\right)^{\frac{p}{p-1}} - \left(\frac{y}{p}\right)^{\frac{p}{p-1}} = 0$ if $y \ge 0$, and ∞ if $y < 0$, since $yx - x^p \to \infty$ as $x \to \infty$ in that case.
- The supremum over x is either zero or infinity, depending on whether $y \ge 0$ or not.

Therefore, we have

$$f^*(y) = \begin{cases} \left(\frac{y}{p-1}\right)^{\frac{p}{p-1}} & \text{if } y \geq 0 \\ \infty & \text{otherwise} \end{cases}$$

which is the conjugate of $f(x) = x^p$ over $\mathbb{R}_{++}$, where $p > 1$.

$\square$

Problem 5: Projection

For any point y, the projection onto a nonempty and closed convex set $\mathcal{X}$ is defined as

$$\Pi_{\mathcal{X}}(y) = \arg\min_{x \in \mathcal{X}} \frac{1}{2}\|x - y\|_2^2.$$

a) Prove that $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2^2 \leqslant \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle$.

b) Prove that $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2 \leqslant \|x - y\|_2$.

c) Denote by $\Delta_R(x, y)$ the *Bregman divergence* with respect to the function R, defined as

$$\Delta_R(x, y) = R(x) - R(y) - \langle \nabla R(y), x - y \rangle, \quad \forall x, y \in \Omega.$$

If we choose $\Pi_{\mathcal{X}}(y) = \arg\min_{x \in \mathcal{X}} \Delta_R(x, y)$, where $\mathcal{X}$ is the n-dimensional simplex $\{x \in \mathbb{R}_+^n | \sum_{i=1}^n x_i = 1\}$, and $R(x) = \sum_{i=1}^n x_i \log(x_i)$. Prove that $\Pi_{\mathcal{X}}(y) = \frac{y}{||y||_1}$ when $y \in \mathbb{R}_{++}^n$. (*Hint: you may use the Jensen's inequality*)

Solution. a) To prove that $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2^2 \leqslant \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle$, need to use the following facts:
- The projection $\Pi_{\mathcal{X}}(x)$ is the unique minimizer of the function $\frac{1}{2}\|x - z\|_2^2$ over $\mathcal{X}$, which implies that $\nabla_z \frac{1}{2}\|x - z\|_2^2|_{z=\Pi_{\mathcal{X}}(x)} = 0$, i.e., $x - \Pi_{\mathcal{X}}(x) \in N_{\mathcal{X}}(\Pi_{\mathcal{X}}(x))$, where $N_{\mathcal{X}}(z)$ is the normal cone of $\mathcal{X}$ at z, defined as $N_{\mathcal{X}}(z) = \{v \in \mathbb{R}^n : \langle v, w - z \rangle \leq 0, \forall w \in \mathcal{X}\}$.
- The normal cone $N_{\mathcal{X}}(z)$ is a convex cone, which means that if $v_1, v_2 \in N_{\mathcal{X}}(z)$, then $\lambda v_1 + (1 - \lambda)v_2 \in N_{\mathcal{X}}(z)$ for any $\lambda \in [0, 1]$.
- The inner product $\langle u, v \rangle$ is linear in both arguments, which means that $\langle \lambda u_1 + (1 - \lambda)u_2, v \rangle = \lambda \langle u_1, v \rangle + (1 - \lambda) \langle u_2, v \rangle$ for any $u_1, u_2, v \in \mathbb{R}^n$ and $\lambda \in [0, 1]$.

Using facts above, we can write

$$\begin{aligned}
\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2^2 &= \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y) \rangle \\
&= \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \Pi_{\mathcal{X}}(x) - x + x - \Pi_{\mathcal{X}}(y) \rangle \\
&= \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \Pi_{\mathcal{X}}(x) - x \rangle + \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - \Pi_{\mathcal{X}}(y) \rangle \\
&= \left\langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \frac{1}{2}(\Pi_{\mathcal{X}}(x) - x) + \frac{1}{2}(\Pi_{\mathcal{X}}(y) - y) \right\rangle + \left\langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y - \frac{1}{2}(\Pi_{\mathcal{X}}(x) - x) - \frac{1}{2}(\Pi_{\mathcal{X}} \right. \\
&= \frac{1}{2} \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \Pi_{\mathcal{X}}(x) - x \rangle + \frac{1}{2} \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \Pi_{\mathcal{X}}(y) - y \rangle + \left\langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y - \frac{1}{2}(\Pi_{\mathcal{X}}(x \right. \\
&\leq \frac{1}{2} \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \Pi_{\mathcal{X}}(x) - x \rangle + \frac{1}{2} \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \Pi_{\mathcal{X}}(y) - y \rangle + \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle \\
&= \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle
\end{aligned}$$

where the inequality follows from the fact that $\Pi_{\mathcal{X}}(x) - x, \Pi_{\mathcal{X}}(y) - y \in N_{\mathcal{X}}(\Pi_{\mathcal{X}}(y))$, so $\frac{1}{2}(\Pi_{\mathcal{X}}(x) - x) + \frac{1}{2}(\Pi_{\mathcal{X}}(y) - y) \in N_{\mathcal{X}}(\Pi_{\mathcal{X}}(y))$ by the convexity of the normal cone, and $\left\langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), \frac{1}{2}(\Pi_{\mathcal{X}}(x) - x) + \frac{1}{2}(\Pi_{\mathcal{X}}(y) - y) \right\rangle \leq 0$ by the definition of the normal cone, which proves that $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2^2 \leqslant \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle$.

b) To prove that $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2 \leqslant \|x - y\|_2$, need to use the Cauchy-Schwarz inequality, which states that for any vectors $u, v \in \mathbb{R}^n$, we have

$$\langle u, v \rangle \leq \|u\|_2 \|v\|_2$$

with equality if and only if u and v are linearly dependent. Applying this inequality to the result of part (a), we get

$$\begin{aligned}
\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2^2 &\leqslant \langle \Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y), x - y \rangle \\
&\leq \|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2 \|x - y\|_2
\end{aligned}$$

Dividing both sides by $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2$, we get

$$\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2 \leqslant \|x - y\|_2$$

which proves that $\|\Pi_{\mathcal{X}}(x) - \Pi_{\mathcal{X}}(y)\|_2 \leqslant \|x - y\|_2$.

c)To prove that $\Pi_{\mathcal{X}}(y) = \frac{y}{\|y\|_1}$ when $y \in \mathbb{R}^n_{++}$, we need to use the following facts:
- The Bregman divergence $\Delta_R(x, y)$ is nonnegative, i.e., $\Delta_R(x, y) \geq 0$ for any $x, y \in \Omega$, where Ω is the domain of R.
- The Bregman divergence $\Delta_R(x, y)$ is zero if and only if $x = y$, i.e., $\Delta_R(x, y) = 0$ if and only if $x = y$.
- The function $R(x) = \sum_{i=1}^n x_i \log x_i$ is strictly convex and differentiable on $\mathbb{R}^n_{++}$, with $\nabla R(x) = \log x + 1$, where $\log x$ and 1 are element-wise operations.
- The set $\mathcal{X} = \{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i = 1\}$ is the standard simplex, which is a convex set.
- The Jensen's inequality states that for any convex function ϕ and any probability distribution p, we have $\phi(\sum_{i=1}^n p_i x_i) \leq \sum_{i=1}^n p_i \phi(x_i)$ for any $x_1, \cdots, x_n \in \mathbb{R}$.

Using these facts, we can write

$$\begin{aligned}
\Pi_{\mathcal{X}}(y) &= \arg\min_{x \in \mathcal{X}} \Delta_R(x, y) \\
&= \arg\min_{x \in \mathcal{X}} (R(x) - R(y) - \langle \nabla R(y), x - y \rangle) \\
&= \arg\min_{x \in \mathcal{X}} (\sum_{i=1}^n x_i \log x_i - \sum_{i=1}^n y_i \log y_i - \langle \log y + 1, x - y \rangle) \\
&= \arg\min_{x \in \mathcal{X}} (\sum_{i=1}^n x_i (\log x_i - \log y_i - 1) + \sum_{i=1}^n y_i (\log y_i + 1)) \\
&= \arg\min_{x \in \mathcal{X}} (\sum_{i=1}^n x_i \log \frac{x_i}{y_i} + \sum_{i=1}^n y_i \log y_i + \sum_{i=1}^n y_i) \\
&= \arg\min_{x \in \mathcal{X}} (\sum_{i=1}^n x_i \log \frac{x_i}{y_i} + \|y\|_1 (\log \|y\|_1 + 1))
\end{aligned}$$

where the last equality follows from the fact that $\sum_{i=1}^n y_i = \|y\|_1$ and log is an element-wise operation. Then, use the Jensen's inequality to show that the minimum is attained at $x = \frac{y}{\|y\|_1}$, since

$$\begin{aligned}
\sum_{i=1}^n x_i \log \frac{x_i}{y_i} &\geq \log \left(\sum_{i=1}^n x_i \frac{x_i}{y_i} \right) \\
&= \log \left(\frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^n y_i} \right) \\
&= \log \left(\frac{\|x\|_2^2}{\|y\|_1} \right)
\end{aligned}$$

with equality if and only if $x_i = \frac{x_i}{y_i}$ for all i, which implies that $x_i = \frac{y_i}{\|y\|_1}$ for all i. Therefore, we have

$$\begin{aligned}
\Pi_{\mathcal{X}}(y) &= \arg\min_{x \in \mathcal{X}} (\sum_{i=1}^n x_i \log \frac{x_i}{y_i} + \|y\|_1 (\log \|y\|_1 + 1)) \\
&= \frac{y}{\|y\|_1}
\end{aligned}$$

when $y \in \mathbb{R}^n_{++}$, which proves that $\Pi_{\mathcal{X}}(y) = \frac{y}{\|y\|_1}$ when $y \in \mathbb{R}^n_{++}$.

$\square$