MSBD 5007 HW4

RONG Shuo

April 20, 2025

Question1

Consider the function $f : \mathbb{R}^d \to \mathbb{R}$ defined by

$$f(\boldsymbol{x}) = \sum_{i=1}^{d} \max(0, 1 - x_i),$$

where $x = [x_1, x_2, \cdots, x_n]^T$. Recall that the proximity operator of a function $g : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\text{prox}_g = \arg\min_{x \in \mathbb{R}^d} \left\{ g(\boldsymbol{x}) + \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \right\}, \boldsymbol{y} \in \mathbb{R}^d.$$

Derive a closed-form expression for $\text{prox}_f(\boldsymbol{y})$.

Answer

Obviously, we can get the $\text{prox}_f(\boldsymbol{y})$ as.

$$\text{prox}_f(\boldsymbol{y}) = \arg\ \min_{\boldsymbol{x} \in \mathbb{R}^d} \{ \sum_{i=1}^{d} \max(0, 1 - x_i) + \frac{1}{2} \sum_{i=1}^{d} (x_i - y_i)^2 \}$$

We can denote $\text{prox}_f(\boldsymbol{y})_i$ as follow:

$$\text{prox}_f(\boldsymbol{y})_i = \arg\ \min_{x \in \mathbb{R}} \{ \max(0, 1 - x) + \frac{1}{2} (x - y_i)^2 \}$$

s.t.

$$\text{prox}_f(\boldsymbol{y}) = \sum_{i=1}^{d} \text{prox}_f(\boldsymbol{y})_i$$

Let $\phi(x) = \max(0, 1 - x) + \frac{1}{2}(x - y_i)^2$
if $x \geq 1$

$$\phi(x) = \{ \frac{1}{2} (x - y_i)^2 \}$$

To minimize this, we need:

$$\begin{aligned} x &= y_i && \text{if } y_i \geq 1 \\ x &= 1 && \text{if } y_i < 1 \end{aligned}$$

Therefore,

$$\begin{aligned} \text{prox}_f(\boldsymbol{y})_i &= y_i && \text{if } y_i \geq 1 \\ \text{prox}_f(\boldsymbol{y})_i &= 1 && \text{if } y_i < 1 \end{aligned}$$

if $x \leq 1$

$$\phi(x) = \{1 - x + \frac{1}{2}(x - y_i)^2\}$$

To minimize this, we need:

$$\begin{aligned} x &= y_i + 1 && \text{if } y_i < 0 \\ x &= 1 && \text{if } y_i \geq 0 \end{aligned}$$

Therefore,

$$\begin{aligned} \text{prox}_f(\boldsymbol{y})_i &= y_i + 1 && \text{if } y_i < 0 \\ \text{prox}_f(\boldsymbol{y})_i &= 1 && \text{if } y_i \geq 0 \end{aligned}$$

Combining these two, we get:

$$\begin{aligned} \text{prox}_f(\boldsymbol{y})_i &= \min\{y_i + 1, 1\} = y_i + 1 && \text{if } y_i < 0 \\ \text{prox}_f(\boldsymbol{y})_i &= \min\{1, 1\} = 1 && \text{if } 0 \leq y_i < 1 \\ \text{prox}_f(\boldsymbol{y})_i &= y_i && \text{if } y_i \geq 1 \end{aligned}$$

So, in conclusion,

$$\text{prox}_f(\boldsymbol{y}) = [\text{prox}_f(\boldsymbol{y})_i]_{i=1}^d$$

where,

$$\begin{aligned} \text{prox}_f(\boldsymbol{y})_i &= \min\{y_i + 1, 1\} = y_i + 1 && \text{if } y_i < 0 \\ \text{prox}_f(\boldsymbol{y})_i &= \min\{1, 1\} = 1 && \text{if } 0 \leq y_i < 1 \\ \text{prox}_f(\boldsymbol{y})_i &= y_i && \text{if } y_i \geq 1 \end{aligned}$$

Question2

In this problem, we study two properties of the 2-norm function $g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2$ defined on $\mathbb{R}^n$. Provide detailed derivations to show that:
(a) The subdifferential of g is given by

$$\partial\|\boldsymbol{x}\|_2 = \begin{cases} \left\{\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right\} & \text{if } \boldsymbol{x} \neq \boldsymbol{0}, \\ \{\boldsymbol{u} \in \mathbb{R}^n | \|\boldsymbol{u}\|_2 \leq 1\} & \text{if } \boldsymbol{x} = \boldsymbol{0}. \end{cases}$$

(b) For any $\alpha > 0$, the proximity operator of $\alpha\|\cdot\|_2$ is

$$\text{prox}_{\alpha\|\cdot\|_2}(\boldsymbol{y}) = \begin{cases} \left(1 - \frac{\alpha}{\|\boldsymbol{y}\|_2}\right)\boldsymbol{y} & \text{if } \|\boldsymbol{y}\|_2 \geq \alpha, \\ \boldsymbol{0} & \text{if } \|\boldsymbol{y}\|_2 \leq \alpha. \end{cases}$$

Answer

(a)

If $\boldsymbol{x} \neq \boldsymbol{0}$, we have:

$$\nabla\|\boldsymbol{x}\|_2 = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}$$

Therefore,

$$\partial\|\boldsymbol{x}\|_2 = \left\{\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right\}, \text{if } \boldsymbol{x} \neq \mathbf{0}.$$

If $\boldsymbol{x} = \mathbf{0}$, we have:

$$\begin{aligned}
&\|\boldsymbol{y}\|_2 \geq \|\mathbf{0}\|_2 + \boldsymbol{v}^T(\boldsymbol{y} - \mathbf{0}) \\
&\|\boldsymbol{y}\|_2 \geq \boldsymbol{v}^T\boldsymbol{y}
\end{aligned}$$

According to cs inequality, we know:

$$\begin{aligned}
&\boldsymbol{v}^T\boldsymbol{y} \leq \|\boldsymbol{v}\|_2\|\boldsymbol{y}\|_2 \\
&\boldsymbol{v}^T\boldsymbol{y} \leq \|\boldsymbol{y}\|_2, \text{ if } \|\boldsymbol{v}\|_2 \leq 1
\end{aligned}$$

Therefore, we get:

$$\partial\|\boldsymbol{x}\|_2 = \{\boldsymbol{u} \in \mathbb{R}^n | \|\boldsymbol{u}\|_2 \leq 1\} \text{ if } \boldsymbol{x} = \mathbf{0}.$$

In conclusion,

$$\partial\|\boldsymbol{x}\|_2 = \begin{cases} \left\{\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2}\right\} & \text{if } \boldsymbol{x} \neq \mathbf{0}, \\ \{\boldsymbol{u} \in \mathbb{R}^n | \|\boldsymbol{u}\|_2 \leq 1\} & \text{if } \boldsymbol{x} = \mathbf{0}. \end{cases}$$

(b)

By definition, we know:

$$\text{prox}_{\alpha\|\cdot\|_2}(\boldsymbol{y}) = \arg\min_{x\in\mathbb{R}^n} \left\{\alpha\|\boldsymbol{x}\|_2 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2\right\}$$

We denote $\phi(\boldsymbol{x}) = \alpha\|\boldsymbol{x}\|_2 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2$, considering the subdifferential of $\|\boldsymbol{x}\|_2$, we have

$$\partial\phi(\boldsymbol{x}) = \begin{cases} \alpha\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_2} + \boldsymbol{x} - \boldsymbol{y} & \text{if } \boldsymbol{x} \neq \mathbf{0}, \\ \alpha\boldsymbol{u} - \boldsymbol{y} & \text{if } \boldsymbol{x} = \mathbf{0}, \text{ where } \|\boldsymbol{u}\|_2 \leq 1 \end{cases}$$

If $\boldsymbol{x} \neq \mathbf{0}$, we get the minimizer $\boldsymbol{x}^*$

$$\begin{aligned}
&\alpha\frac{\boldsymbol{x}^*}{\|\boldsymbol{x}^*\|_2} + \boldsymbol{x}^* - \boldsymbol{y} = \mathbf{0} \\
&\boldsymbol{y} = \alpha\frac{\boldsymbol{x}^*}{\|\boldsymbol{x}^*\|_2} + \boldsymbol{x}^*
\end{aligned}$$

Obviously, $t\boldsymbol{y} = \boldsymbol{x}^*, t \neq 0$, therefore,

$$\begin{aligned}
&\boldsymbol{y} = \alpha\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_2} + t\boldsymbol{y} \\
&t = 1 - \frac{\alpha}{\|\boldsymbol{y}\|_2} \\
&\boldsymbol{x}^* = (1 - \frac{\alpha}{\|\boldsymbol{y}\|_2})\boldsymbol{y}, \text{ where } \alpha \neq \|\boldsymbol{y}\|_2
\end{aligned}$$

So in conclusion,

$$\begin{aligned}
&\min\phi(\boldsymbol{x}) = \alpha\|\boldsymbol{y}\|_2 - \frac{\alpha^2}{2} \text{ if } \boldsymbol{x} \neq \mathbf{0} \\
&\boldsymbol{x}^* = (1 - \frac{\alpha}{\|\boldsymbol{y}\|_2})\boldsymbol{y}
\end{aligned}$$

If $\boldsymbol{x} = \boldsymbol{0}$, we get

$$\phi(\boldsymbol{x}) = \frac{\|\boldsymbol{y}\|_2^2}{2}$$
$$\min \phi(\boldsymbol{x}) = \frac{\|\boldsymbol{y}\|_2^2}{2}$$

By solving the inequality, we know:

$$\frac{\|\boldsymbol{y}\|_2^2}{2} < \alpha\|\boldsymbol{y}\|_2 - \frac{\alpha^2}{2}$$
$$\|\boldsymbol{y}\|_2^2 - 2\alpha\|\boldsymbol{y}\|_2 + \alpha^2 < 0$$

It holds when $\alpha > \|\boldsymbol{y}\|_2^2$.
Combining these two condition, we know:

$$\operatorname{prox}_{\alpha\|\cdot\|_2}(\boldsymbol{y}) = \begin{cases} \left(1 - \frac{\alpha}{\|\boldsymbol{y}\|_2}\right)\boldsymbol{y} & \text{if } \|\boldsymbol{y}\|_2 \geq \alpha, \\ \boldsymbol{0} & \text{if } \|\boldsymbol{y}\|_2 \leq \alpha. \end{cases}$$

Question 3

In this problem, we consider the elastic net regression model, which is widely used in statistics for regularized linear regression. The optimization problem is given by

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \lambda_1\|\boldsymbol{x}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{x}\|_2^2,$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{b} \in \mathbb{R}^m$, and $\lambda_1, \lambda_2 > 0$ are regularization parameters. Answer the following:
(a) For any $\beta_1, \beta_2 > 0$, find a closed-form expression for proximity operator $\operatorname{prox}_{\beta_1\|\cdot\|_1 + \frac{\beta_2}{2}\|\cdot\|_2^2}(\boldsymbol{y})$.
(b) We apply the forward-backward splitting (i.e. proximal gradient) algorithm. In particular, we apply a forward step for $\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2$ and a backward step for $\lambda_1\|\boldsymbol{x}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{x}\|_2^2$. Write down the iterative update rule for the resulting algorithm.

Answer

(a)

$$\operatorname{prox}_{\beta_1\|\cdot\|_1 + \frac{\beta_2}{2}\|\cdot\|_2^2}(\boldsymbol{y}) = \arg\min_{\boldsymbol{x}\in\mathbb{R}^n} \left(\frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \beta_1\|\boldsymbol{x}\|_1 + \frac{\beta_2}{2}\|\boldsymbol{x}\|_2^2\right)$$

We can reconstruct the problem as minimize the following problem:

$$\phi(x) = \beta_1|x| + \frac{\beta_2}{2}x^2 + \frac{1}{2}(x - y_i)^2$$

And we know, we need

$$0 \in \partial\beta_1|x| + \beta_2 x + x - y_i$$
$$y_i \in \partial\beta_1|x| + (1 + \beta_2)x$$

Therefore, we get:

$$x = \begin{cases} \frac{y_i - \beta_1}{1+\beta_2} & \text{if } y_i > \beta_1, \\ 0 & \text{if } |y_i| \leq \beta_1, \\ \frac{y_i + \beta_1}{1+\beta_2} & \text{if } y_i < -\beta_1, \end{cases}$$

Therefore:

$$
\begin{aligned}
\operatorname{prox}_{\beta_1\|\cdot\|_1+\frac{\beta_2}{2}\|\cdot\|_2^2}(\boldsymbol{y})_i &= \begin{cases} \frac{y_i-\beta_1}{1+\beta_2} & \text{if } y_i > \beta_1, \\ 0 & \text{if } |y_i| \leq \beta_1, \\ \frac{y_i+\beta_1}{1+\beta_2} & \text{if } y_i < -\beta_1, \end{cases} \\
\operatorname{prox}_{\beta_1\|\cdot\|_1+\frac{\beta_2}{2}\|\cdot\|_2^2}(\boldsymbol{y}) &= [\operatorname{prox}_{\beta_1\|\cdot\|_1+\frac{\beta_2}{2}\|\cdot\|_2^2}(\boldsymbol{y})_i]_{i=1}^n
\end{aligned}
$$

(b)

$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= T_{\alpha_k\lambda_1\lambda_2}\left(x^{(k)} - \alpha_k \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{x}^{(k)} - \boldsymbol{b})\right) \\
T_{\alpha_k\lambda_1\lambda_2}(\boldsymbol{x}) &= \operatorname{prox}_{\alpha_k\lambda_1\|\cdot\|_1+\frac{\alpha_k\lambda_2}{2}\|\cdot\|_2^2}(\boldsymbol{x})
\end{aligned}
$$

Question 4

Let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Prove the following properties:
(a) For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and for any $\boldsymbol{u} \in \partial g(\boldsymbol{x})$ and $\boldsymbol{v} \in \partial g(\boldsymbol{y})$, show that

$$\langle \boldsymbol{u} - \boldsymbol{v}, \boldsymbol{x} - \boldsymbol{y} \rangle \geq 0.$$

Hint: Use the definition of the subdifferential
(b) Prove that the proximity operator of g is nonexpansive; that is, for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$\|\operatorname{prox}_g(\boldsymbol{x}) - \operatorname{prox}_g(\boldsymbol{y})\|_2 \leq \|\boldsymbol{x} - \boldsymbol{y}\|_2$$

Hint: Apply the result from part (a)
(c) Show that a point $\boldsymbol{x}^*$ minimizes g if and only if

$$\boldsymbol{x}^* = \operatorname{prox}_g(\boldsymbol{x}^*)$$