

MSBD 5007 HW4

RONG Shuo

April 20, 2025

Question1

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

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

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

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

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

Answer

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

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

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

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

s.t.

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

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

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

To minimize this, we need:

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

Therefore,

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

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,

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

Combining these two, we get:

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

So, in conclusion,

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

where,

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

Question2

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

(a) The subdifferential of g is given by

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

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

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

Answer

(a)

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

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

Therefore,

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

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

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

According to cs inequality, we know:

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

Therefore, we get:

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

In conclusion,

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

(b)

By definition, we know:

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

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

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

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

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

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

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

So in conclusion,

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

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

$$\begin{aligned}\phi(\mathbf{x}) &= \frac{\|\mathbf{y}\|_2^2}{2} \\ \min \phi(\mathbf{x}) &= \frac{\|\mathbf{y}\|_2^2}{2}\end{aligned}$$

By solving the inequality, we know:

$$\begin{aligned}\frac{\|\mathbf{y}\|_2^2}{2} &< \alpha \|\mathbf{y}\|_2 - \frac{\alpha^2}{2} \\ \|\mathbf{y}\|_2^2 - 2\alpha \|\mathbf{y}\|_2 + \alpha^2 &< 0\end{aligned}$$

It holds when $\alpha > \|\mathbf{y}\|_2$.

Combining these two condition, we know:

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

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

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

where $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, 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 $\text{prox}_{\beta_1\|\cdot\|_1 + \frac{\beta_2}{2}\|\cdot\|_2^2}(\mathbf{y})$.
- (b) We apply the forward-backward splitting (i.e. proximal gradient) algorithm. In particular, we apply a forward step for $\frac{1}{2}\|\mathbf{Ax} - \mathbf{b}\|_2^2$ and a backward step for $\lambda_1\|\mathbf{x}\|_1 + \frac{\lambda_2}{2}\|\mathbf{x}\|_2^2$. Write down the iterative update rule for the resulting algorithm.

Answer

(a)

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

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

$$\begin{aligned}0 &\in \partial \beta_1 |x| + \beta_2 x + x - y_i \\ y_i &\in \partial \beta_1 |x| + (1 + \beta_2)x\end{aligned}$$

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:

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

(b)

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

Question 4

Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex function. Prove the following properties:

(a) For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and for any $\mathbf{u} \in \partial g(\mathbf{x})$ and $\mathbf{v} \in \partial g(\mathbf{y})$, show that

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

Hint: Use the definition of the subdifferential

(b) Prove that the proximity operator of g is nonexpansive; that is, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\|\text{prox}_g(\mathbf{x}) - \text{prox}_g(\mathbf{y})\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2$$

Hint: Apply the result from part (a)

(c) Show that a point \mathbf{x}^* minimizes g if and only if

$$\mathbf{x}^* = \text{prox}_g(\mathbf{x}^*)$$

Answer

(a)

We know $\forall \mathbf{z} \in \mathbb{R}^n$, we get

$$\begin{aligned} g(\mathbf{z}) &\geq g(\mathbf{x}) + \langle \mathbf{u}, \mathbf{z} - \mathbf{x} \rangle \\ g(\mathbf{z}) &\geq g(\mathbf{y}) + \langle \mathbf{v}, \mathbf{z} - \mathbf{y} \rangle \end{aligned}$$

So we choose $\mathbf{z} = \mathbf{y}, \mathbf{z} = \mathbf{x}$ separately.

$$\begin{aligned} g(\mathbf{y}) &\geq g(\mathbf{x}) + \langle \mathbf{u}, \mathbf{y} - \mathbf{x} \rangle \\ g(\mathbf{x}) &\geq g(\mathbf{y}) + \langle \mathbf{v}, \mathbf{x} - \mathbf{y} \rangle \\ g(\mathbf{y}) + g(\mathbf{x}) &\geq g(\mathbf{x}) + g(\mathbf{y}) + \langle \mathbf{u}, \mathbf{y} - \mathbf{x} \rangle + \langle \mathbf{v}, \mathbf{x} - \mathbf{y} \rangle \\ 0 &\geq \langle \mathbf{u} - \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle \\ \langle \mathbf{u} - \mathbf{v}, \mathbf{x} - \mathbf{y} \rangle &\geq 0 \end{aligned}$$

(b)

For $\text{prox}_g(\mathbf{y}) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \phi(\mathbf{x})$. Let's denote $\phi(\mathbf{x}) = g(\mathbf{x}) + \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2$, $\partial \phi(\mathbf{x}) = \partial g(\mathbf{x}) + \mathbf{x} - \mathbf{y}$. The minimizer \mathbf{x}^* of $\phi(\mathbf{x})$ satisfies $\mathbf{0} \in \partial g(\mathbf{x}^*) + \mathbf{x}^* - \mathbf{y}$, and $\text{prox}_g(\mathbf{y}) = \mathbf{x}^*$. Suppose $\text{prox}_g(\mathbf{x}) = \mathbf{p}$, $\text{prox}_g(\mathbf{x}) = \mathbf{q}$

$$\begin{aligned} \mathbf{0} &\in \partial g(\mathbf{p}) + (\mathbf{p} - \mathbf{x}) \\ \mathbf{x} - \mathbf{p} &\in \partial g(\mathbf{p}) \\ \mathbf{0} &\in \partial g(\mathbf{q}) + (\mathbf{q} - \mathbf{y}) \\ \mathbf{y} - \mathbf{q} &\in \partial g(\mathbf{q}) \end{aligned}$$

by (a)

$$\begin{aligned}\langle (\mathbf{x} - \mathbf{p}) - (\mathbf{y} - \mathbf{q}), \mathbf{p} - \mathbf{q} \rangle &\geq 0 \\ \langle (\mathbf{x} - \mathbf{y}) - (\mathbf{p} - \mathbf{q}), \mathbf{p} - \mathbf{q} \rangle &\geq 0 \\ \langle \mathbf{x} - \mathbf{y}, \mathbf{p} - \mathbf{q} \rangle &\geq \|\mathbf{p} - \mathbf{q}\|_2^2\end{aligned}$$

By cs inequality, we know:

$$\langle \mathbf{x} - \mathbf{y}, \mathbf{p} - \mathbf{q} \rangle \leq \|\mathbf{x} - \mathbf{y}\|_2 \|\mathbf{p} - \mathbf{q}\|_2$$

Thus,

$$\begin{aligned}\|\mathbf{x} - \mathbf{y}\|_2 &\geq \|\mathbf{p} - \mathbf{q}\|_2 \\ \|\text{prox}_g(\mathbf{x}) - \text{prox}_g(\mathbf{y})\|_2 &\leq \|\mathbf{x} - \mathbf{y}\|_2\end{aligned}$$

(c)

if \mathbf{x}^* is a minimizer of g , we get

$$\mathbf{0} \in \partial g(\mathbf{x}^*)$$

We know

$$\mathbf{0} \in \partial g(\mathbf{z}) + \mathbf{z} - \mathbf{x}^* \implies \text{prox}_g(\mathbf{x}^*) = \mathbf{z}$$

And we know:

$$\begin{aligned}\mathbf{0} &\in \partial g(\mathbf{x}^*) \\ \mathbf{0} &\in \partial g(\mathbf{x}^*) + \mathbf{x}^* - \mathbf{x}^*\end{aligned}$$

So we can conclude

$$\text{prox}_g(\mathbf{x}^*) = \mathbf{x}^*$$

if $\mathbf{x}^* = \text{prox}_g(\mathbf{x}^*)$

$$\begin{aligned}\mathbf{0} &\in \partial g(\mathbf{x}^*) + \mathbf{x}^* - \mathbf{x}^* \\ \mathbf{0} &\in \partial g(\mathbf{x}^*)\end{aligned}$$

We know

$$g \text{ is a convex function and } \mathbf{0} \in \partial g(\mathbf{x}) \implies \mathbf{x} \text{ is a minimizer of } g.$$

So in conclusion,

$$\mathbf{x}^* \text{ minimizes } g \iff \mathbf{x}^* = \text{prox}_g(\mathbf{x}^*)$$