Solution to Homework 7

Total 30 points

Problem 1. (12 points) For each of the following convex functions, compute the proximal operator:

(a) $f(\mathbf{x}) = \|\mathbf{x}\|_2$;

(b) $f(\mathbf{x}) = -\sum_{i=1}^{d} \log(x_i)$, $\mathbf{x} \in \mathbb{R}^d$;

(c) $f(\mathbf{x}) = \inf_{y\in\mathcal{C}} \|\mathbf{x} - \mathbf{y}\|_2$, where $\mathcal{C}$ is a closed convex set.

(d) $f(\mathbf{x}) = \frac{1}{2}(\inf_{y\in\mathcal{C}} \|\mathbf{x} - \mathbf{y}\|_2)^2$, where $\mathcal{C}$ is a closed convex set.

Solution. According to the definition of proximal operator, we know

$$\mathbf{z} \in \text{prox}_f(\mathbf{x}) \Rightarrow \mathbf{x} - \mathbf{z} \in \partial f(\mathbf{z})$$

(a) Notice that

$$\partial f(\mathbf{z}) = \begin{cases} \frac{\mathbf{z}}{||\mathbf{z}||_2} & \text{if } \mathbf{z} \neq 0 \\ \{\mathbf{g} |\ ||\mathbf{g}||_2 \leq 1\} & \text{if } \mathbf{z} = 0 \end{cases}$$

If $\mathbf{z} = 0$, we have $\|\mathbf{x}\|_2 \leq 1$. If $||\mathbf{z}||_2 \neq 0$, we have

$$\mathbf{z} = \frac{\mathbf{x}||\mathbf{z}||_2}{1 + ||\mathbf{z}||_2},$$

which means

$$||\mathbf{z}||_2 = \frac{||\mathbf{x}||_2||\mathbf{z}||_2}{1 + ||\mathbf{z}||_2}$$

i.e. $1 + ||\mathbf{z}||_2 = ||\mathbf{x}||_2$. Then we can obtain:

$$\mathbf{z} = \frac{\mathbf{x}(||\mathbf{x}||_2 - 1)}{||\mathbf{x}||_2}, ||\mathbf{x}||_2 > 1.$$

Thus,

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}(||\mathbf{x}||_2 - 1)}{||\mathbf{x}||_2} & \text{if } ||\mathbf{x}||_2 > 1 \\ 0 & \text{if } ||\mathbf{x}||_2 \leq 1 \end{cases}$$

(b) By solving $\mathbf{x} - \mathbf{z} = \nabla f(\mathbf{z})$, we can get

$$(\text{prox}_f(\mathbf{x}))_i = \mathbf{z}_i^* = \frac{x_i + \sqrt{x_i^2 + 4}}{2}, \ i = 1, 2, \ldots, d$$

(c) Consider the function $g(\mathbf{z}) = \inf_{y\in\mathcal{C}} \|\mathbf{z}-\mathbf{y}\|_2 + \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2$.

(i) If $\mathbf{z}\in\mathcal{C}$, we have $g(\mathbf{z}) = \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2$, which means $\arg\min_{\mathbf{z}\in\mathcal{C}} g(\mathbf{z}) = \mathcal{P}_{\mathcal{C}}(\mathbf{x})$.

(ii) If $\mathbf{z}\notin\mathcal{C}$, then $g(\mathbf{z}) = \|\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2 - \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2$. Let $\nabla g(\mathbf{z}) = 0$, we have

$$\mathbf{x}-\mathbf{z} = \frac{\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})}{\|\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2}. \tag{1}$$

By the property of projection, we have

$$\langle \mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z}), \mathcal{P}_{\mathcal{C}}(\mathbf{x})-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\rangle \le 0 \tag{2}$$
$$\langle \mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x}), \mathcal{P}_{\mathcal{C}}(\mathbf{z})-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\rangle \le 0 \tag{3}$$

On the other hand, according to the fact that

$$\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{z}) = \mathbf{x}-\mathbf{z}+\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z}) = \frac{\|\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2+1}{\|\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2}(\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})), \tag{4}$$

and inequality (2), we have

$$\langle \mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{z}), \mathcal{P}_{\mathcal{C}}(\mathbf{x})-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\rangle \le 0 \tag{5}$$

Sum up (3) and (5), we have $\|\mathcal{P}_{\mathcal{C}}(\mathbf{x})-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2^2 \le 0$, which means

$$\mathcal{P}_{\mathcal{C}}(\mathbf{x}) = \mathcal{P}_{\mathcal{C}}(\mathbf{z}) \tag{6}$$

Combine (4) and (6), we have

$$\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2 = \|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2 = \|\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2 + 1. \tag{7}$$

Thus,

$$\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x}) = \frac{\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2}{\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2-1}(\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})),$$

which means

$$\mathbf{z} = \frac{\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2-1}{\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2}\mathbf{x} + \frac{1}{\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2}\mathcal{P}_{\mathcal{C}}(\mathbf{x}).$$

Put all things together, we have

$$\mathrm{prox}_f(\mathbf{x}) = \begin{cases} \frac{\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2-1}{\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2}\mathbf{x} + \frac{1}{\|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2}\mathcal{P}_{\mathcal{C}}(\mathbf{x}) & \text{if } \|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2 \ge 1 \\ \mathcal{P}_{\mathcal{C}}(\mathbf{x}) & \text{if } \|\mathbf{x}-\mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2 < 1 \end{cases}$$

(d) Notice that

$$\begin{aligned} \frac{1}{2}(\inf_{y\in\mathcal{C}} \|\mathbf{z}-\mathbf{y}\|_2)^2 + \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 &= \frac{1}{2}\|\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2^2 + \frac{1}{2}\|\mathbf{x}-\mathbf{z}\|_2^2 \\ &\ge \frac{1}{4}\|\mathbf{x}-\mathbf{z}+\mathbf{z}-\mathcal{P}_{\mathcal{C}}(\mathbf{z})\|_2^2 \end{aligned}$$

$$\geq \frac{1}{4} \|\mathbf{x} - \mathcal{P}_{\mathcal{C}}(\mathbf{x})\|_2^2$$

The equations holds when $\mathbf{z}$ is the middle point between $\mathbf{x}$ and $\mathcal{P}_{\mathcal{C}}(\mathbf{x})$, i.e.,

$$\mathrm{prox}_f(\mathbf{x}) = \frac{1}{2}(\mathbf{x} + \mathcal{P}_{\mathcal{C}}(\mathbf{x})).$$

Problem 2. (5 points) (**Extended Moreau decomposition**) Suppose f is closed and convex, and $\lambda > 0$, show that

$$\mathbf{x} = \mathrm{prox}_{\lambda f}(\mathbf{x}) + \lambda\, \mathrm{prox}_{\frac{1}{\lambda} f^*}(\mathbf{x}/\lambda).$$

Solution. By Moreau decomposition, we have

$$\mathrm{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \mathrm{prox}_{(\lambda f)^*}(\mathbf{x}).$$

Notice that $(\lambda f)^*(\mathbf{x}) = \lambda f^*(\mathbf{x}/\lambda)$. In addition, the property of proximal operator tell us

$$f(\mathbf{x}) = \lambda g(\mathbf{x}/\lambda) \Rightarrow \mathrm{prox}_f(\mathbf{x}) = \lambda\, \mathrm{prox}_{\frac{1}{\lambda} g}(\mathbf{x}/\lambda).$$

Thus $\mathrm{prox}_{(\lambda f)^*}(\mathbf{x}) = \lambda\, \mathrm{prox}_{\frac{1}{\lambda} f^*}(\mathbf{x}/\lambda)$, which completes the proof.

Problem 3. (13 points) Given a data matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ and data vector $\mathbf{b} \in \mathbb{R}^n$, the Lasso regression aims to solve the following the regularized least-squares problem:

$$\min_{\mathbf{x} \in \mathbb{R}^d} \frac{1}{2n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1.$$

The homework ZIP file contains two text files, labeled `A_Lasso.txt` and `b_Lasso.txt`, that contains an $n \times d$ matrix $\mathbf{A}$ and an n-dimensional vector $\mathbf{b}$, with $n = 1000$, $d = 500$. The file `xopt.txt` contains the optimal solution to the Lasso problem with regularization parameter $\lambda = 2\sigma\sqrt{\frac{\log d}{n}}$ with $\sigma = 0.2$.

(a) (3 points) Implement the subgradient descent algorithm with both constant ($\eta = 0.2$) and decaying stepsizes ($\eta_t = 1/\sqrt{t+1}$).

(b) (3 points) Now implement the proximal gradient algorithm with constant step size $\eta = 0.2$.

(c) (4 points) For both the proximal gradient algorithm and subgradient algorithm, plot (on the same axes) the logarithm of Euclidean norm errors $\log \|\mathbf{x}_t - \mathbf{x}^*\|_2$ versus the iteration number, where $\mathbf{x}^*$ is the optimal solution from `xopt.txt`.

(d) (3 points) Comment on the convergence behavior that you see for the two methods and the relation to results established in class. Hand in your code and the report.

Solution. See lasso.py