DATA130026.01 Optimization
Solution of Assignment 10

1. Let f be a convex and continuous differentiable function over $\mathbb{R}^n$. For a fixed $x \in \mathbb{R}^n$, define the functions
$$g_x(y) = f(y) - \nabla f(x)^T y.$$
Suppose ∇f is L Lipschitz continuous, i.e.,
$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$
   (a) Prove that x is a minimizer of g_x over $\mathbb{R}^n$.
   (b) Show that for any $x, y \in \mathbb{R}^n$,
$$g_x(x) \le g_x(y) - \frac{1}{2L}\|\nabla g_x(y)\|^2.$$
   (c) Show that for any $x, y \in \mathbb{R}^n$,
$$f(x) + \nabla f(x)^T(y - x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y).$$

Solution.

   (a) Since $f(y)$ is convex and $\nabla f(x)^T y$ is affine, $g_x(y)$ is also convex. Let $\nabla g_x(y) = \nabla f(y) - \nabla f(x) = 0$, we have $y = x$. Thus, x is a minimizer of g_x over $\mathbb{R}^n$.
   (b) $\forall\ y_1, y_2 \in \mathbb{R}^n$,
$$\begin{aligned}\|\nabla g_x(y_1) - \nabla g_x(y_2)\| &= \|(\nabla f(y_1) - \nabla f(x)) - (\nabla f(y_2) - \nabla f(x))\| \\ &= \|\nabla f(y_1) - \nabla f(y_2)\| \\ &\le L\|y_1 - y_2\|.\end{aligned}$$
Therefore, ∇g_x is also L Lipschitz continuous. Then, $\forall\ y, z \in \mathbb{R}^n$, we have
$$\begin{aligned}& g_x(z) \le g_x(y) + \nabla g_x(y)^T(z - y) + \frac{L}{2}\|z - y\|^2, \\ \Rightarrow\quad & \min_z g_x(z) \le \min_z\{g_x(y) + \nabla g_x(y)^T(z - y) + \frac{L}{2}\|z - y\|^2\}, \\ \Rightarrow\quad & g_x(x) \le g_x(y) + \nabla g_x(y)^T\left(y - \frac{1}{L}\nabla g_x(y) - y\right) + \frac{L}{2}\left\|y - \frac{1}{L}\nabla g_x(y) - y\right\|^2 \\ & \quad = g_x(y) - \frac{1}{2L}\|\nabla g_x(y)\|^2.\end{aligned}$$
   (c) Since $g_x(x) = f(x) - \nabla f(x)^T x$ and $g_x(y) = f(y) - \nabla f(x)^T y$, from (b), we have
$$\begin{aligned}& f(x) - \nabla f(x)^T x \le f(y) - \nabla f(x)^T y - \frac{1}{2L}\|\nabla g_x(y)\|^2, \\ \Rightarrow\quad & f(x) + \nabla f(x)^T(y - x) + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y).\end{aligned}$$

2. For each of the following functions on $\mathbb{R}^n$, explain how to calculate a subgradient at a given x.

   (a) $f(x) = \sup_{0 \le t \le 1} p(t)$, where $p(t) = x_1 + x_2 t + \ldots + x_n t^{n-1}$.

   (b) $f(x) = x_{[1]} + x_{[2]} + \ldots + x_{[k]}$, where $x_{[i]}$ denotes the ith largest elements of x.

   (c) $f(x) = \|Ax - b\|_2 + \|x\|_2$ where $A \in \mathbb{R}^{m \times n}$.

   Solution.

   (a) Define the active set $T := \{t \mid p(t) = f(x)\}$, then the subdifferential of f at x is

   $$\partial f(x) = \operatorname{conv} \bigcup \{\partial_x p(t) \mid t \in T\}.$$

   $\forall\, \tilde{t} \in T$, we have

   $$\partial_x p(\tilde{t}) = (1, \tilde{t}, \ldots, \tilde{t}^{n-1})^T.$$

   Therefore, we can easily obtain a subgradient

   $$g(x) = (1, \tilde{t}, \ldots, \tilde{t}^{n-1})^T.$$

   (b) Define $A_k := \left\{a \in \mathbb{R}^n \mid a_i \in \{0,1\}, i = 1, \ldots, n,\ \sum_{i=1}^n a_i = k\right\}$. Then

   $$f(x) = \sup_{a \in A_k} a^T x.$$

   Define the active set $\tilde{A} := \{a \in A_k \mid a^T x = f(x)\}$, then the subdifferential of f at x is

   $$\partial f(x) = \operatorname{conv} \bigcup \{\tilde{a} \mid \tilde{a} \in \tilde{A}\}.$$

   Therefore, we can easily obtain a subgradient $g(x) = \tilde{a}$.
   Moreover, $\tilde{a}$ satisfies

   $$\tilde{a}_i = \begin{cases} 1, & \text{if } x_i \in \{x_{[1]}, \ldots, x_{[k]}\}, \\ 0, & \text{if } x_i \notin \{x_{[1]}, \ldots, x_{[k]}\}. \end{cases}$$

   (c) We know that

   $$\partial \|Ax - b\|_2 = \begin{cases} \dfrac{A^T(Ax - b)}{\|Ax - b\|_2}, & \text{if } Ax - b \neq 0, \\ \{A^T g \mid \|g\|_2 \le 1\}, & \text{if } Ax - b = 0. \end{cases}$$

   $$\partial \|x\|_2 = \begin{cases} \dfrac{x}{\|x\|_2}, & \text{if } x \neq 0, \\ \{h \mid \|h\|_2 \le 1\}, & \text{if } x = 0. \end{cases}$$

Then the subdifferential of f at x is

$$\partial f(x) = \begin{cases} \dfrac{A^T(Ax-b)}{\|Ax-b\|_2} + \dfrac{x}{\|x\|_2}, & \text{if } Ax-b \neq 0 \text{ and } x \neq 0, \\ \left\{ \dfrac{A^T(Ax-b)}{\|Ax-b\|_2} + h \;\middle|\; \|h\|_2 \leq 1 \right\}, & \text{if } Ax-b \neq 0 \text{ and } x = 0, \\ \left\{ A^T g + \dfrac{x}{\|x\|_2} \;\middle|\; \|g\|_2 \leq 1 \right\}, & \text{if } Ax-b = 0 \text{ and } x \neq 0, \\ \left\{ A^T g + h \mid \|g\|_2 \leq 1, \|h\|_2 \leq 1 \right\}, & \text{if } Ax-b = 0 \text{ and } x = 0. \end{cases}$$

Therefore, we can easily obtain a subgradient by setting $g = h = 0$:

$$g(x) = \begin{cases} \dfrac{A^T(Ax-b)}{\|Ax-b\|_2} + \dfrac{x}{\|x\|_2}, & \text{if } Ax-b \neq 0 \text{ and } x \neq 0, \\ \dfrac{A^T(Ax-b)}{\|Ax-b\|_2}, & \text{if } Ax-b \neq 0 \text{ and } x = 0, \\ \dfrac{x}{\|x\|_2}, & \text{if } Ax-b = 0 \text{ and } x \neq 0, \\ 0, & \text{if } Ax-b = 0 \text{ and } x = 0. \end{cases}$$

3. Under the same notations in lecture slides, prove that with diminishing but non-summable step size $\alpha_i = \frac{R}{G\sqrt{i}}$, we have

$$f_{bs}^k - f^* \leq O(\frac{RG}{\sqrt{k}}).$$

(Hint: Use $f_{bs}^k \leq \bar{f}_{bs}^k$, where $\bar{f}_{bs}^k = \min_{i=k/2,\ldots,k} f(x_k)$.)

Solution. According to leture slides,

$$\begin{aligned} & \|x_{i+1} - x^*\|_2^2 \leq \|x_i - x^*\|_2^2 - 2\alpha_i(f(x_i) - f^*) + \alpha_i^2 \|g_i\|_2^2, \\ \Rightarrow \quad & 2\alpha_i(f(x_i) - f^*) \leq \|x_i - x^*\|_2^2 - \|x_{i+1} - x^*\|_2^2 + \alpha_i^2 \|g_i\|_2^2. \end{aligned} \tag{1}$$

Combine (1) for $i = \frac{k}{2}, \ldots, k$, and define $\bar{f}_{bs}^k = \min_{i=k/2,\ldots,k} f(x_k)$, then

$$\begin{aligned} 2\sum_{i=\frac{k}{2}}^{k} \alpha_i(\bar{f}_{bs}^k - f^*) &\leq 2\sum_{i=\frac{k}{2}}^{k} \alpha_i(f(x_i) - f^*) \\ &\leq \|x_{\frac{k}{2}} - x^*\|_2^2 + \sum_{i=\frac{k}{2}}^{k} \alpha_i^2 \|g_i\|_2^2 \\ &\leq \tilde{R}^2 + \sum_{i=\frac{k}{2}}^{k} \alpha_i^2 G^2, \end{aligned} \tag{2}$$

where $\tilde{R} := \|x_{\frac{k}{2}} - x^*\|_2$. Inserting $\alpha_i = \frac{R}{G\sqrt{i}}$ into (2) gives

$$\bar{f}_{bs}^k - f^* \leq \frac{\frac{\tilde{R}^2}{R}G + RG\sum_{i=\frac{k}{2}}^k \frac{1}{i}}{2\sum_{i=\frac{k}{2}}^k \frac{1}{\sqrt{i}}}. \tag{3}$$

Since

$$\sum_{i=\frac{k}{2}}^k \frac{1}{i} = \sum_{i=\frac{k}{2}}^k \int_{i-1}^i \frac{1}{i} dx \leq \int_{\frac{k}{2}-1}^k \frac{1}{x} dx = \log \frac{k}{\frac{k}{2}-1} = \log \frac{2k}{k-2}, \tag{4}$$

and

$$\sum_{i=\frac{k}{2}}^k \frac{1}{\sqrt{i}} = \sum_{i=\frac{k}{2}}^k \int_i^{i+1} \frac{1}{\sqrt{i}} dx \geq \int_{\frac{k}{2}}^{k+1} \frac{1}{\sqrt{x}} dx = 2(\sqrt{k+1} - \sqrt{\frac{k}{2}}), \tag{5}$$

substituting (4) and (5) into (3) gives

$$\bar{f}_{bs}^k - f^* \leq \frac{\frac{\tilde{R}^2}{R}G + RG\log\frac{2k}{k-2}}{4\left(\sqrt{k+1} - \sqrt{\frac{k}{2}}\right)} = O(\frac{RG}{\sqrt{k}}).$$

4. Write a MATLAB code for solving the Lasso problem using subgradient method:

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

where $\tau > 0$ is a weighting parameter, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^n$ are given data. Choose $x = 0$ as the starting point. Terminate your code after 10000 iterations. Use the following Matlab code to generate the data:

```
m = 100; n = 500; s = 50;
A = randn(m,n);
xs = zeros(n,1); picks = randperm(n);
xs(picks(1:s)) = randn(s,1);
b = A*xs; tau=0.001;
```

Use constant step size, constant step length, diminishing step size and Polyak's step size. Try three different constants or parameter for constant step size, constant step length and diminishing step size. Plot fours figures to show the evolutions for `f(xk)-f*` (the optimal value can be computed by CVX) for the four step size rules.

Solution.
Please run `codes/problem_4.m`.