最优化11

加兴华 18300290007

1

1. 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 $A \in \mathbb{R}^{m\times n}$.

1.

(a) $p(t) = [1, t, \cdots, t^{n-1}]\cdot[x_1, \cdots, x_n]' =: A(t)x$

$\therefore \partial p(t) = \{A(t)'\}$.

$\because f(x) = \max_{0\le t\le 1} p(t)$

$\therefore \partial f(x) = \mathrm{Conv}\bigcup_{0\le t\le 1}\{A(t)' \mid p(t)=f(x)\}$

(b) $f(x) = \max_{1\le i_1<i_2<\cdots<i_k} \sum_{j=1}^{k} x_{i_j} = \max_{1\le i_1<i_2<\cdots<i_k\le n} [I_1,\cdots,I_n]\cdot x$. 式中 $I_m = \begin{cases}1 & m\in\{i_1,\cdots,i_k\}\\ 0 & m\notin\{i_1,\cdots,i_k\}\end{cases}$

记 $p(i_1,\cdots,i_k) = [I_1,\cdots,I_n]\cdot x$.

s.t. $\partial p(i_1,\cdots,i_k) = \{[I_1,\cdots,I_n]'\}$

$\therefore \partial f(x) = \mathrm{Conv}\bigcup\{[I_1,\cdots,I_n]'(i_1,\cdots,i_k) \mid p(i_1,\cdots,i_k)=f(x)\}$.

(c) 记 $g(x)=\|Ax-b\|_2$; $h(x)=\|x\|_2$.

s.t. $\partial f(x) = \partial g(x)+\partial h(x) = A^T\partial h(Ax-b)+\partial h(x)$

而 $\partial h(x) = \begin{cases} x/\|x\|_2 & x\ne 0\\ \{g \mid \|g\|_2\le 1\} & x=0.\end{cases}$

$\partial h(Ax-b) = \begin{cases}(Ax-b)/\|Ax-b\|_2 & Ax\ne b\\ \{g\mid \|g\|_2\le 1\}. & Ax=b\end{cases}$

$\therefore \partial f(x) = \begin{cases} \frac{A^T(Ax-b)}{\|Ax-b\|_2}+\frac{x}{\|x\|_2} & Ax\ne b \,\&\, x\ne 0\\ \{\frac{A^T(Ax-b)}{\|Ax-b\|_2}+g_2 \mid \|g_2\|\le 1\}. & Ax\ne b\,\&\,x=0\\ \{A^Tg_1+\frac{x}{\|x\|_2}\mid \|g_1\|_2\le 1\}. & Ax=b\,\&\,x\ne 0\\ \{A^Tg_1+g_2\mid \|g_1\|_2,\|g_2\|_2\le 1\}. & Ax=b\,\&\,x=0.\end{cases}$

#已知 $\partial f(x)$ 后，从 $\partial f(x)$ 中特取一个作为 $g(x)$ 即可

(a): $g(x) = A(t)'$ s.t. $p(t)=f(x)$

(b): $g(x)=[I_1,\cdots,I_n](i_1,\cdots,i_k)$ s.t. $p(i_1\cdots i_k)=f(x)$

(c): 令 $g_1,g_2=0$. $\partial f(x)\to g(x)$.

2

2. 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

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

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

2. $\|x_{i+1} - x^*\|_2^2 \le \|x_i + x^*\|_2^2 - 2\alpha_i(f(x_i) - f^*) + \alpha_i^2\|g_i\|^2$

$\Rightarrow 2\alpha_i(f(x_i) - f^*) \le \|x_i + x^*\|_2^2 - \|x_{i+1} - x^*\|_2^2 + \alpha_i^2\|g_i\|_2^2$

$\therefore 2\sum_{k/2}^{k}\alpha_i \cdot (\bar{f}_{bs}^k - f^*) \le \sum_{k/2}^{k} 2\alpha_i(f(x_i) - f^*) \le \|x_{\frac{k}{2}} - x^*\|_2^2 - \|x_{k+1} - x^*\|_2^2 + \sum_{k/2}^{k}\alpha_i^2\|g_i\|_2^2$

$\le \underbrace{\|x_{\frac{k}{2}} - x^*\|_2^2}_{= R^2} + \sum_{k/2}^{k}\alpha_i^2 \underbrace{\|g_i\|_2^2}_{\le G^2}$

代入 $\alpha_i = \frac{R}{G\sqrt{i}}$

$\Rightarrow \bar{f}_{bs}^k - f^* \le \dfrac{R^2 + G^2 \cdot \frac{R^2}{G^2} \cdot \sum_{k/2}^{k}\frac{1}{i}}{2 \cdot \frac{R}{G}\sum_{k/2}^{k}\frac{1}{\sqrt{i}}}$

数学上: $\begin{cases} \sum_{k/2}^{k}\frac{1}{i} \le \int_{\frac{k}{2}-1}^{k}\frac{1}{x}dx = \ln\frac{2k}{k-2} \\ \sum_{k/2}^{k}\frac{1}{\sqrt{i}} \ge \int_{\frac{k}{2}}^{k+1}\frac{1}{\sqrt{x}}dx = 2(\sqrt{k+1} - \sqrt{\frac{k}{2}}) \end{cases}$

$\therefore f_{bs}^k - f^* \le \bar{f}_{bs}^k - f^* \le \dfrac{R^2 \cdot \frac{G}{R} + RG \cdot \log\frac{2k}{k-2}}{4(\sqrt{k+1} - \sqrt{\frac{k}{2}})} = O(\frac{RG}{\sqrt{k}})$ as $\begin{cases} \log\frac{2k}{k-2} \to \log 2 \\ \frac{\sqrt{k+1} - \sqrt{\frac{k}{2}}}{\sqrt{k}} \to 1 - \sqrt{\frac{1}{2}} \end{cases}$

3

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

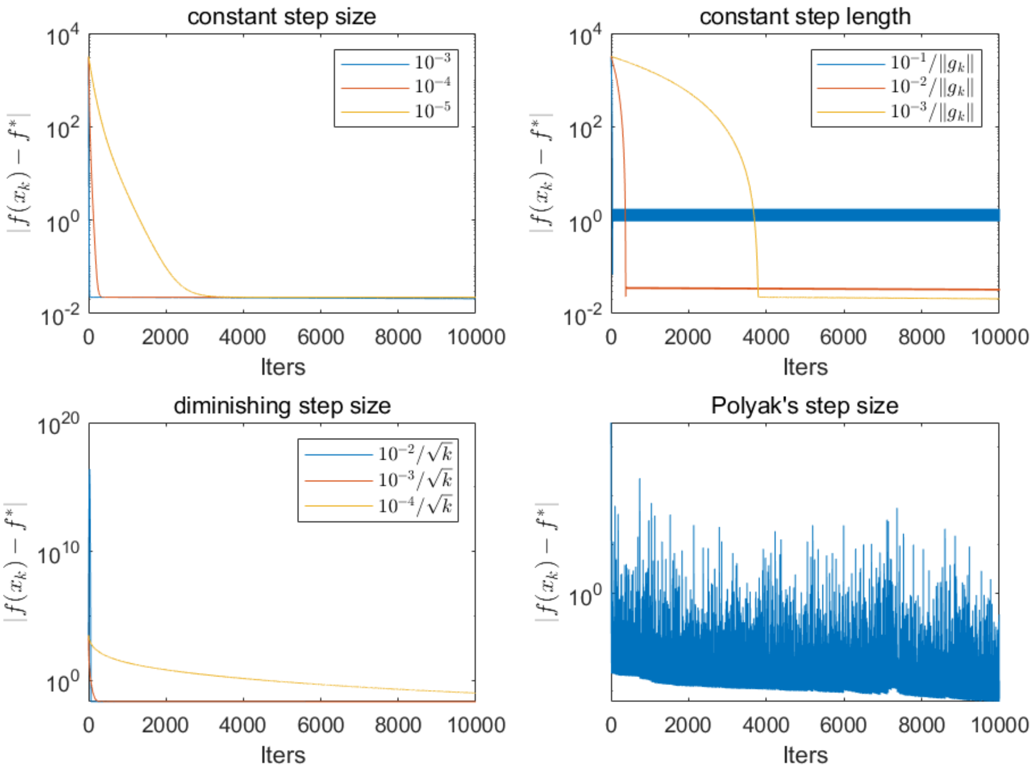
$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|_2^2 + \tau\|x\|_1$$

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, dimnishing step size and Polyak's step size. Try three different constants or parameter for constant step size, constant step length and dimnishing 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.

运行文件夹中question3.m文件，结果如下：



说明：图2中蓝线和红线水平部分更粗是视觉效果，放大看发现是密度很大的上下波动。