CMSE890 Homework#2

Haiyang Yu

October 10, 2019

1.

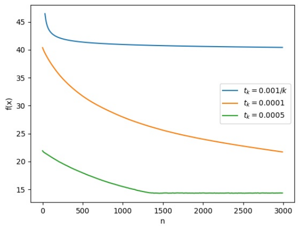

Figure 1: Iteration 3000

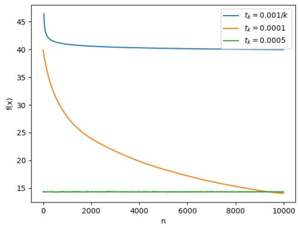

Figure 2: Iteration 10000

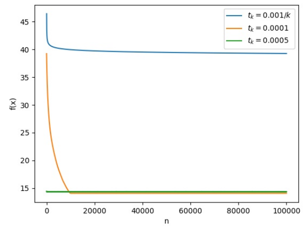

Figure 3: Iteration 100000

```
import numpy as np
import matplotlib.pyplot as plt

def partial(x):
    res=np.zeros((len(x),1))
    for i in range(len(x)):

        if abs(x[i])<eps:
                res[i]=0
        else:
            res[i]=x[i]/abs(x[i])

    tmp=A.dot(x)-b
    res2=A.T.dot(tmp)
    return res+res2

def f(x):
    s1=0
    for i in range(len(x)):
        s1+=abs(x[i])
    tmp=A.dot(x)-b
    s2=np.dot(tmp.T,tmp)
    return s1+s2


np.random.seed(1000)
eps=1e-7
m=200
n=1000
A=np.random.normal(0,1,(m,n))
x_bar=np.zeros((n,1))
for i in range(20):
    k=np.random.randint(0,1000)
    x_bar[k]=np.random.normal(0,1)
b=A.dot(x_bar)

x_pre=np.zeros((n,1))
x_now=x_pre
y1=[]
```

```
n1=[]
y2=[]
n2=[]
y3=[]
n3=[]
iteration=100000
for i in range(iteration):
    tk=0.001/(i+1)
    x_now=x_pre-tk*partial(x_pre)
    x_pre=x_now
    if i%10==0 and i>20:
        y1.append(f(x_now)[0][0])
        n1.append(i)
for i in range(iteration):
    tk=0.0001
    x_now=x_pre-tk*partial(x_pre)
    x_pre=x_now
    if i%10==0:
        y2.append(f(x_now)[0][0])
        n2.append(i)
for i in range(iteration):
    tk=0.0005
    x_now=x_pre-tk*partial(x_pre)
    x_pre=x_now
    if i%10==0:
        y3.append(f(x_now)[0][0])
        n3.append(i)
plt.plot(n1,y1,label=r'$t_{k}=0.001/k$')
plt.plot(n2,y2,label=r'$t_{k}=0.0001$')
plt.plot(n3,y3,label=r'$t_{k}=0.0005$')
plt.xlabel("n")
plt.ylabel("f(x)")
plt.legend()
plt.savefig("subgradient100000.jpg")
plt.show()
```

2.

Let $\boldsymbol{a} = \boldsymbol{x}^{k-1} - \boldsymbol{x}^{k-2}$,$\boldsymbol{b} = \nabla f(\boldsymbol{x}^{k-1}) - \nabla f(\boldsymbol{x}^{k-2})$

$$f(\frac{1}{t}) = ||\boldsymbol{a}/t - \boldsymbol{b}||_2^2$$

Let

$$\frac{\partial f}{\partial 1/t} = 0$$

we get $\hat{t} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}$.

In the same way, we get $\tilde{t} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}$

Because $\nabla f(\boldsymbol{x})$ is Lipschitz continuous with L, $\frac{L}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - f(\boldsymbol{x})$ is convex. $L\boldsymbol{x} - \nabla f(\boldsymbol{x})$ is monotone, which is equivalent to $\hat{t} \geq 1/L$.Because of the co-coercivity, we get $\tilde{t} \geq 1/L$. So $\min\{\hat{t}, \tilde{t}\} \geq 1/L$

3.

Suppose $\boldsymbol{y} = \mathrm{prox}_f(\boldsymbol{x}), \boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}, \boldsymbol{y} = (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$.

1

$$y_i = \begin{cases} 1 & \text{if} \quad x_i > 2 \\ x_i - 1 & \text{if} \quad 1 \leq x_i \leq 2 \\ 0 & \text{if} \quad -1 < x_i < 1 \\ x_i + 1 & \text{if} \quad -2 \leq x_i \leq -1 \\ -1 & \text{if} \quad x_i < -2 \end{cases}$$

2

Consider $g(\boldsymbol{x}) = ||\boldsymbol{x}||_1$. $f(\boldsymbol{x}) = g(A\boldsymbol{x} - \boldsymbol{b})$. Suppose $AA^{\mathrm{T}} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, where $C = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})$. We have $C^2 = AA^{\mathrm{T}}$.

$\boldsymbol{u} = \mathrm{prox}_f(\boldsymbol{x})$ is the solution of the optimization problem

$$\min_{\boldsymbol{u},\boldsymbol{v}} g(\boldsymbol{v}) + \frac{1}{2}||\boldsymbol{u} - \boldsymbol{x}||_2^2 \quad \text{s.t.} \;\; A\boldsymbol{u} - \boldsymbol{b} = \boldsymbol{v}$$

Lagrangian function:

$$L(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\lambda}) = g(\boldsymbol{v}) + \frac{1}{2}||\boldsymbol{u} - \boldsymbol{x}||_2^2 + \boldsymbol{\lambda}^{\mathrm{T}}(A\boldsymbol{u} - \boldsymbol{b} - \boldsymbol{v})$$

KKT conditions:

$$\begin{aligned} \boldsymbol{u} - \boldsymbol{x} + A^{\mathrm{T}}\boldsymbol{\lambda} &= \boldsymbol{0} \\ \boldsymbol{\lambda} &\in \partial g(\boldsymbol{v}) \\ A\boldsymbol{u} - \boldsymbol{b} &= \boldsymbol{v} \end{aligned}$$

So

$$\boldsymbol{0} \in \partial g(\boldsymbol{v}) + (AA^{\mathrm{T}})^{-1}(\boldsymbol{v} - (A\boldsymbol{x} - \boldsymbol{b}))$$

, which means $\boldsymbol{v}$ is the minimizer of $h(\boldsymbol{v}) = g(\boldsymbol{v}) + 1/2||C^{-1}\boldsymbol{v} - C^{-1}(A\boldsymbol{x} - b)||_2^2$. Because C is diagonal matrix, $g(\boldsymbol{v}) = ||\boldsymbol{v}||_1$. It is separable for

$$h(\boldsymbol{v}) = \sum_{i=1}^{n} |v_i| + \frac{1}{2}(\frac{v_i}{\sqrt{\lambda_i}} - C^{-1}(A\boldsymbol{x} - b)[i])$$

So we can get the minimizer $\hat{\boldsymbol{v}}$.

$$\begin{aligned} \mathrm{prox}_f(\boldsymbol{x}) &= \boldsymbol{u} \\ &= \boldsymbol{x} - A^{\mathrm{T}}\boldsymbol{\lambda} \\ &= \boldsymbol{x} - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}(A\boldsymbol{x} - \hat{\boldsymbol{v}} - \boldsymbol{b}) \end{aligned}$$

3

For any $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$.Note that x_{k_i} is the i-th largest component in the vector $\boldsymbol{x}$, k_i is the index of the i-th largest component in the vector $\boldsymbol{x}$. For example, if $\boldsymbol{x} = (2, 4, 1, 5)^{\mathrm{T}}$, then $x_{k_1} = 5, k_1 = 4$.

$\boldsymbol{y} = \mathrm{prox}_f(\boldsymbol{x})$. We have $y_{k_1} = x_{k_1} - 1$, $y_{k_i} = x_{k_i}, i > 1$.

4

$$\boldsymbol{y} = \begin{cases} (1 - 1/||\boldsymbol{x}||)\boldsymbol{x} & \text{if} \quad ||\boldsymbol{x}|| < 1 \\ \boldsymbol{0} & \text{if} \quad 1 \le ||\boldsymbol{x}|| \le 3 \\ (1 + 1/||\boldsymbol{x}||)\boldsymbol{x} & \text{if} \quad ||\boldsymbol{x}|| > 3 \end{cases}$$

5

Suppose X has SVD $X = P\mathrm{diag}(\lambda_1, \cdots, \lambda_n)Q^{\mathrm{T}}$, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_n)^{\mathrm{T}}$.

Let

$$\hat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{u}} ||\boldsymbol{u}||_1 + \frac{1}{2}||\boldsymbol{u} - \lambda||_2^2$$

The proximal mapping is

$$\mathrm{prox}_f(X) = P\mathrm{diag}(\hat{\lambda_1}, \cdots, \hat{\lambda_n})Q^{\mathrm{T}}$$

4.

Let

$$g(\boldsymbol{x}) = \sup_{\boldsymbol{u}}(-f(\boldsymbol{u}) - \frac{1}{2\lambda}||\boldsymbol{u}||_2^2 + \frac{1}{\lambda}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x})$$

$\boldsymbol{u}^*$ is the maximizer, $\mathbf{0} \in \partial\lambda f(\boldsymbol{u}^*) + \boldsymbol{u}^* - \boldsymbol{x}$, so

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}} \lambda f(\boldsymbol{u}) + \frac{1}{2}||\boldsymbol{u} - \boldsymbol{x}||_2^2 = \mathrm{prox}_{\lambda f}(\boldsymbol{x})$$

Thus,

$$\begin{aligned}\nabla f_\lambda(\boldsymbol{x}) &= \frac{\boldsymbol{x}}{\lambda} - \nabla g(\boldsymbol{x}) \\ &= \frac{\boldsymbol{x}}{\lambda} - \mathrm{prox}_{\lambda f}(\boldsymbol{x})\end{aligned}$$