Homework 3 – Due by 11:59 PM on Sunday March 2

(1) Consider the following function for a given paramete $\mu > 0$:

$$f_\mu(x) = \begin{cases} \dfrac{\|x\|^2}{2\mu} & \|x\| \le \mu \\ \|x\| - \dfrac{\mu}{2} & \text{otherwise} \end{cases}$$

(a) Show that the gradient of $f_\mu(x)$ is

$$\nabla f_\mu(x) = \begin{cases} \dfrac{x}{\mu} & \|x\| \le \mu \\ \dfrac{x}{\|x\|} & \text{otherwise} \end{cases}$$

(b) Show that the Lipschitz constant of the gradient is $1/\mu$.

Hint: For part (a), use the fact that $\|x\| = \sqrt{x_1^2 + x_2^2 + \ldots, x_n^2}$. For part (b), use the definition of the projection onto the closed ball $B[0, \mu]$ and nonexpansivness of the projection operator.

(2) Consider $f(x) = \log(1 + e^x)$ for $x \in \mathbb{R}$. Show that the Lipschitz constant of the gradient is $1/4$.

(3) Consider the following quadratic optimization problem

$$\min_x \quad x^T Ax + 2b^T x, \quad \text{s.t.} \ \sum_{i=1}^{n} x_i \ge -10$$

Write a code in MATLAB/Python to implement projected gradient method with initial point $x_0 = \mathbf{0}$, for solving the above problem with stopping criteria $\|x_{k+1} - x_k\| \le \epsilon$ where $\epsilon = 1e-3$, and step size of $\alpha = 1/L$, where L is the Lipschitz constant of the gradient. Fix the seed and generate the problem parameters as follows:

```
rng(123);
n = 10;
R = rand(n,n);
A = R'*R;
b = rand(n,1);
```

Display the function value at the last iteration point.

(4) Consider the data-fitting problem, i.e., $\min_{x\in\mathbb{R}^{n+1}} \frac{1}{2}\|Sx - t\|^2$, and implement gradient method with stopping criteria $\|\nabla f(x)\| \le \epsilon$, where $\epsilon = 1e-4$, step size of $\alpha = 1/L$, where L is the Lipschitz constant of the gradient, and $x_0 = \mathbf{0}$. Fix the seed and generate the problem parameters as follows:

```
rng(123);
d = 1;
n = 100;
xbar = rand(d,1); % use to generate data points
s = rand(n,1)*10-5;
t = s*xbar+randn(n,1);
S = [s,ones(n,1)];
x = zeros(d+1,1); % initial point x0
```

```
np.random.seed(123)
d = 1
n = 100
xbar = np.random.rand(d)
s = np.random.rand(n)*10-5
t = s*xbar+np.random.randn(n)
S = np.column_stack((s,np.ones(n)))
x = np.zeros(2)
```

Plot the data points and the line corresponding to the algorithm's output, which fits the data.

Students enrolled in SIE 549 must solve the following problem.
Students in SIE 449 will get extra credit by solving it.

(5) Let function f be convex and continuously differentiable whose gradient is Lipschitz with constant L. Also, let $\{x_k\}_{k\geq 0}$ be the sequence generated by Projected Gradient method with step-size $\alpha = 1/L$. Show that $f(x_T) - f(x^*) \leq \frac{L}{2T}\|x_0 - x^*\|^2$, for all $T \geq 1$.

Hint: Follow the steps of the proof for convergence rate of the gradient method. To show that $\langle \nabla f(x_k), x_{k+1} - x^* \rangle \leq \frac{1}{\alpha}(x_k - x_{k+1})^T(x_{k+1} - x^*)$, *use the fact that* $x_{k+1} = P_C(x_k - \alpha\nabla f(x_k))$ *and use the second theorem of the projection.*

Hint for Q4: Plotting Data Points and Lines in MATLAB and Python

MATLAB:

(1) How to plot the data set (s, t): `plot(s,t,'*');`

(2) How to plot a line if the slope is a and intercept is b, i.e. $y = ax + b$?

```
vec = linspace(-1,1);  % choose the interval based on your data points
plot(vec,vec*a+b);
```

(3) How plot the data set and the line in the same figure?

```
plot(s,t,'*');
hold on
vec = linspace(-1,1);
plot(vec,vec*a+b);
```

Python:

```
import matplotlib.pyplot as plt
```

(1) How to plot the data set (s, t):

```
plt.plot(s, t, '*')
```

(2) How to plot a line if the slope is a and intercept is b, i.e. $y = ax + b$?

```
vec = np.linspace(-1, 1)
plt.plot(vec, vec * a + b)
plt.show()
```

1) $$f_\mu(x) = \begin{cases} \frac{\|x\|^2}{2\mu} & \|x\| \le \mu \\ \|x\| - \frac{\mu}{2} & \text{otherwise} \end{cases}$$

a) for $\|x\| \le \mu$:

$$\nabla f_\mu(x) = \frac{1}{2\mu} * \frac{df}{dx}(\Sigma x^2)$$
$$= \frac{1}{2\mu} * 2x$$
$$= \frac{x}{\mu}$$

for $\|x\| > \mu$:

$$\nabla f_\mu(x) = \frac{df}{dx}(\|x\|)$$
$$= \frac{df}{dx}(\sqrt{\Sigma x^2})$$

Chain rule

$$= \frac{df}{dx}((\Sigma x^2)^{1/2})$$
$$= 1/2(\Sigma x^2)^{-1/2} * 2x$$
$$= (\Sigma x^2)^{-1/2} * x$$
$$= \frac{x}{\|x\|}$$

Hint: Projection onto closed ball $B[0, \mu]$

b) Show that lipshitz const. of gradient is: $1/\mu$

Projection of x onto ball $[0, \mu]$: if $\|x\| > \mu$, then $x = \frac{x}{\|x\|}\mu$, else $x = x$ (inside ball)

Meaning for $\|x\| \le \mu$, $x = x$ and for $\|x\| > \mu$, $x = \frac{x}{\|x\|}\mu$

rewrite gradient in terms of $P_\mu(x)$:

$$\nabla f_\mu(x) = \begin{cases} \frac{x}{\mu} = \frac{P_\mu(x)}{\mu} & \|x\| \le \mu \\ \frac{x}{\|x\|} = \frac{P_\mu(x)}{\mu} & \text{otherwise} \end{cases}$$

$$\nabla f_\mu(x) = \frac{P_c(x)}{\mu}$$

Ball $[0, \mu]$ is non-expansive:

$$\|P_\mu(x) - P_\mu(y)\| \le \|x - y\|$$

Known

Assume:

$$\|\nabla f_\mu(x) - \nabla f_\mu(y)\| \le 1/\mu \|x - y\|$$
$$= \left\|\frac{P_\mu(x)}{\mu} - \frac{P_\mu(y)}{\mu}\right\| \le 1/\mu \|x - y\|$$
$$= \mu\left\|\frac{P_\mu(x)}{\mu} - \frac{P_\mu(y)}{\mu}\right\| \le \mu(1/\mu)\|x - y\|$$
$$= \|P_\mu(x) - P_\mu(y)\| \le \|x - y\|$$

∎

2) $f(x) = \log(1+e^x)$ for $x \in \mathbb{R}$. Show Lipshitz const. is $1/4$ for gradient.

$$\nabla f(x) = \frac{e^x}{(1+e^x)}$$

$$\nabla^2 f(x) = \frac{(1+e^x)e^x - e^x(e^x)}{(1+e^x)^2} = \frac{e^x + e^{2x} - e^{2x}}{(1+e^x)^2} = \frac{e^x}{(1+e^x)^2}$$

$f(x)$

$g'(x) = 2(1+e^x)(e^x)$

$$\frac{d}{dx}\left(\nabla^2 f(x)\right) = \frac{(1+e^x)^2 e^x - 2e^x(1+e^x)e^x}{(1+e^x)^4} = \frac{e^x(1+e^x)\left((1+e^x) - 2e^x\right)}{(1+e^x)^4} = \frac{e^x(1-e^x)}{(1+e^x)^3}$$

Critical point: $x = 0$

$$\nabla^2 f(0) = \frac{e^0}{(1+e^0)^2} = \frac{1}{(1+1)^2} = \frac{1}{4}$$

max point of $\nabla^2 f(x)$, equalling Limpschitz constant.

$\blacksquare$

3) $$\min_x \; x^T A x + 2b^T x, \quad \text{s.t.} \sum_{i=1}^{n} x_i \geq -10$$

initial: $x_0 = 0$

stopping criteria: $\|x_{k+1} - x_k\| \leq \epsilon$, $\epsilon = 1e\text{-}3$

Step size: $\alpha = 1/L = 1/\|2A\|$

Lipschitz: $\nabla f(x) = 2Ax + 2b^T$

$$\|(2Ax+2b) - (2Ay+2b)\| \leq L\|x-y\|$$

$$\|2Ax - 2Ay\| \leq L\|x-y\|$$

$$\|A(x-y)\| \leq L/2 \|x-y\|$$

matrix property: $\|A(x-y)\| \leq \|A\|\|x-y\|$

$$L/2 = \|A\|$$

$$L = \|2A\|$$

Code/hw3_projectedgradient.py

```python
#Initialize variables
x = np.zeros((n, 1))
e = 1e-3
L = 2 * la.norm(A,2)
alpha = 1 / L

#returns projected value
def projection(x, c = -10):
    sum = np.sum(x)

    if sum >= c:
        return x
    else:
        diff = c - x
        adj = diff / len(x)
        projected = x + adj
        return projected

#Begin algorithm
for _ in range(100000): # range used to prevent infinite loop
    grad = 2 * (A@x + b)
    x_temp = x - alpha * (grad)
    x = projection(x_temp)

    if(la.norm(x - x_temp, 2) <= e):
        fun_val = x.T @ A @ x + 2*(b.T @ x)
        min_val = x
        break
```