Math 173A Homework #2 Solutions

1. (a) \( f(x) = \|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_n^2. \)

\[ \nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{pmatrix} = 2x \]

(b) Let \( g(z) = \|z\|_2^2, h(w) = Aw, \) then \( f(x) = g(h(x)). \) By chain rule, \( Df = Dg|_{h(x)} Dh(x) = (2Ax)^T A = 2x^T A^T A \)

Thus \( \nabla f(x) = (Df(x))^T = 2A^T Ax \)

(c) Let \( g(z) = \|z\|_2^2, h(w) = Aw - b, \) then \( f(x) = g(h(x)). \) By chain rule, \( Df = Dg|_{h(x)} Dh(x) = (2Ax - b)^T A \)

Thus \( \nabla f(x) = (Df(x))^T = 2A^T (Ax - b). \)

(d) \[ \nabla f(x) = \nabla \|Ax - b\|_2^2 + \gamma \nabla \|x\|_2^2 = 2A^T (Ax - b) + 2\gamma x \]

2. (a) Choose \( x_0 \in \mathbb{R} \)

For \( t = 1, 2, 3, \cdots \)

\[ x_t = x_{t-1} - \mu_{t-1} 2x \]

(b) Choose \( x_0 \in \mathbb{R} \)

For \( t = 1, 2, 3, \cdots \)

\[ x_t = x_{t-1} - \mu_{t-1} 2A^T Ax \]

(c) Choose \( x_0 \in \mathbb{R} \)

For \( t = 1, 2, 3, \cdots \)

\[ x_t = x_{t-1} - \mu_{t-1} 2A^T (Ax - b) \]

1
Choose $x_0 \in \mathbb{R}$
For $t = 1, 2, 3, \cdots$
\[ x_t = x_{t-1} - \mu_{t-1}(2A^T(Ax - b) + 2\gamma x) \]

3. (a) $f(x) = \|x\|_2^2$, then
\[ \nabla^2 f(x) = 2I_n \succeq 0 \]
where $I_n$ is the $n \times n$ identity matrix. Thus $f(x)$ is convex.

(b) Let $x, y \in \mathbb{R}^n$ and $\tilde{x} = Ax - b, \tilde{y} = Ay - b$. For any $\alpha \in [0, 1]$,
\[ f(\alpha \tilde{x} + (1 - \alpha)\tilde{y}) = f(\alpha(Ax - b) + (1 - \alpha)(Ay - b)) = f(A(\alpha x + (1 - \alpha)y) - b) = g(\alpha x + (1 - \alpha)y) \]
We also have $f(\tilde{x}) = f(Ax - b) = g(x), f(\tilde{y}) = f(Ay - b) = g(y)$. Because $f$ is convex, we have
\[ f(\alpha \tilde{x} + (1 - \alpha)\tilde{y}) \leq \alpha f(\tilde{x}) + (1 - \alpha)f(\tilde{y}) \]
which implies
\[ g(\alpha x + (1 - \alpha)y) \leq \alpha g(x) + (1 - \alpha)g(y) \]
It completes the proof that $g$ is convex.

(c) Let $h(x) = \|x\|_2^2$ which is convex. Then $f(x) = h(Ax - b)$ is also convex by the result of (b).

4. In this question, we need to find $x^*$ such that $\nabla f(x^*) = 0$.

(a) $\nabla f(x^*) = 2x^* = 0 \Rightarrow x^* = 0$.

(b) $\nabla f(x^*) = 2A^TAx^* = 0 \Rightarrow x^* = 0$.

(c) $\nabla f(x^*) = 2A^T(Ax^* - b) = 0 \Leftrightarrow A^TAx^* = A^Tb \Rightarrow x^* = (A^TA)^{-1}A^Tb$.

(d) $\nabla f(x^*) = 2A^T(Ax^* - b) + 2\gamma x = 2A^T(Ax^* - b) + 2\gamma I_n x = 0 \Leftrightarrow (A^TA + \gamma I_n)x^* = A^Tb \Rightarrow x^* = (A^TA + \gamma I_n)^{-1}A^Tb$.

5. (a)
\[ A = \begin{pmatrix} u_1^2 & v_1^2 \\ \vdots & \vdots \\ u_N^2 & v_N^2 \end{pmatrix} b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{N \times 1} \]
(b) \[ x_0 \in \mathbb{R} \]
For \( t = 1, 2, 3, \ldots \)
\[ x_t = x_{t-1} - \mu_{t-1} 2A^T (Ax - b) \]

(c) 
\[
data = \text{csvread('C:C:Users\ziyan\Desktop\myFile.txt');}
A = (data.^2); \]
\[
b = \text{ones}(1000,1); \]
\[
u = 1/(2 * \text{norm}(A' * A)); \]
\[
x = [1;1]; \]
\[
\text{for } i = 1:1000 \]
\[ x = x - 2 * u * A' * (A * x - b); \]
\[
\text{end} \]

(d) We get \( a^* = [0.9558, 0.2468] \).

(e) We get \( \hat{a} = [0.9558, 0.2468] \) from the closed form. It is almost same as what we get from the algorithm.