THE UNIVERSITY OF TEXAS
AT AUSTIN

EE381V LARGE SCALE OPTIMIZATION

Problem Set 1

Edited by LaTeX

Department of Computer Science

STUDENT

Jimmy Lin

xl5224

COURSE COORDINATOR

Sujay Sanghavi

UNIQUE NUMBER

17350

RELEASE DATE

September 15, 2014

DUE DATE

September 23, 2014

TIME SPENT

13 hours

September 16, 2014

Table of Contents

List of Figures

Part I
Matlab and Computational Assignment

1 Gradient Descent on three matrices

Command to get executed:

```
>> gd_run_script()
```

1.1 $X1$, $b1$

- Range of γ that leads to convergence: $(0, 2)$
- Range of γ that leads to divergence: $(2, +\infty)$
- Explanation: if $\gamma = 2$, the program indicates that

$$\forall k,\ f(x^{k+1}) = f(x^k)$$

Since the above equation is constantly true (independent of the minima), we can conclude that gradient descent with $\gamma = 2$ goes to the opposite side of that quadratic curve. Intuitively, the program will diverge if we set larger ($\gamma > 2$) and converge if we set smaller ($\gamma < 2$).

- Two illustrative examples: $\gamma = 0.5$ and $\gamma = 3.0$

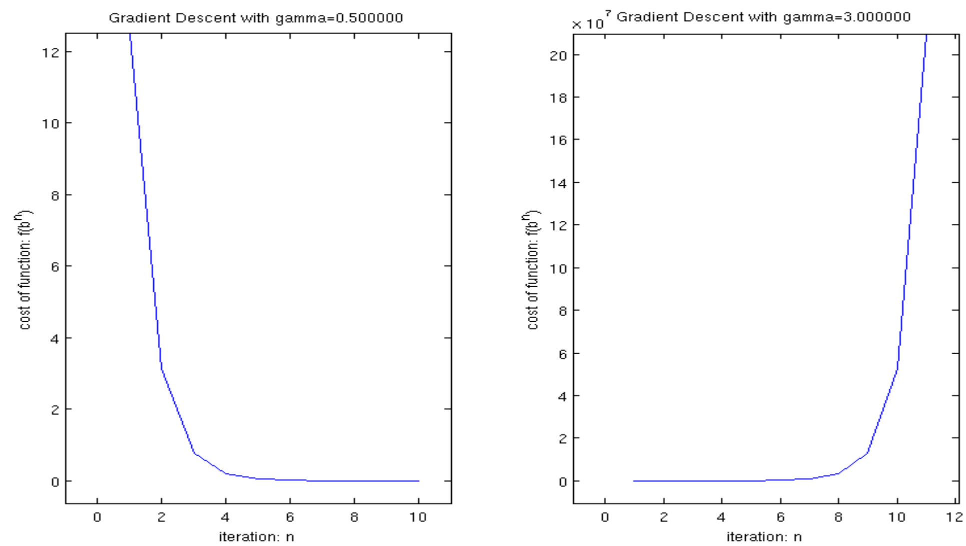

Figure 1: Illustration for gradient descent on $X1$, staring with $b1$ by $\gamma = 0.5$ and 3.0

1.2 $X2$, $b2$

- Range of γ that leads to convergence: $(0, 2)$
- Range of γ that leads to divergence: $(2, +\infty)$
- Explanation: if $\gamma = 2$, the program indicates that

$$\forall k,\ f(x^{k+1}) = f(x^k)$$

  Since the above equation is constantly true (independent of the minima), we can conclude that gradient descent with $\gamma = 2$ goes to the opposite side of that quadratic curve. Intuitively, the program will diverge if we set larger ($\gamma > 2$) and converge if we set smaller ($\gamma < 2$).

- Two illustrative examples: $\gamma = 1.5$ and $\gamma = 3.0$

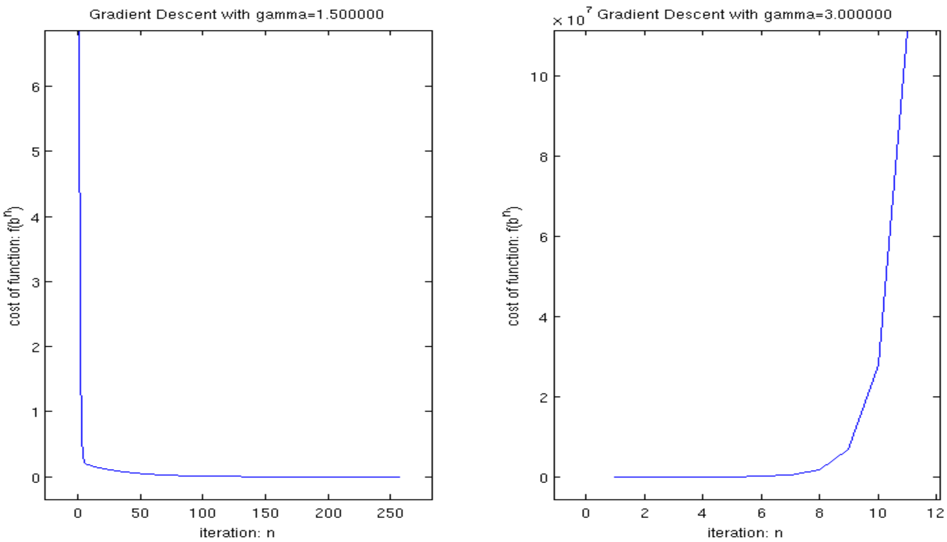

Figure 2: Illustration for gradient descent on $X2$, starting with $b2$ by $\gamma = 1.5$ and 3.0

1.3 $X3$, $b3$

- Range of γ that leads to convergence: $(0, 0.02)$
- Range of γ that leads to divergence: $(0.02, +\infty)$
- Explanation: if $\gamma = 0.02$, the program indicates that

$$\forall k,\ f(x^{k+1}) = f(x^k)$$

Since the above equation is constantly true (independent of the minima), we can conclude that gradient descent with $\gamma = 0.02$ goes to the opposite side of that quadratic curve. Intuitively, the program will diverge if we set larger ($\gamma > 0.02$) and converge if we set smaller ($\gamma < 0.02$).

- Two illustrative examples: $\gamma = 0.005$ and $\gamma = 0.05$

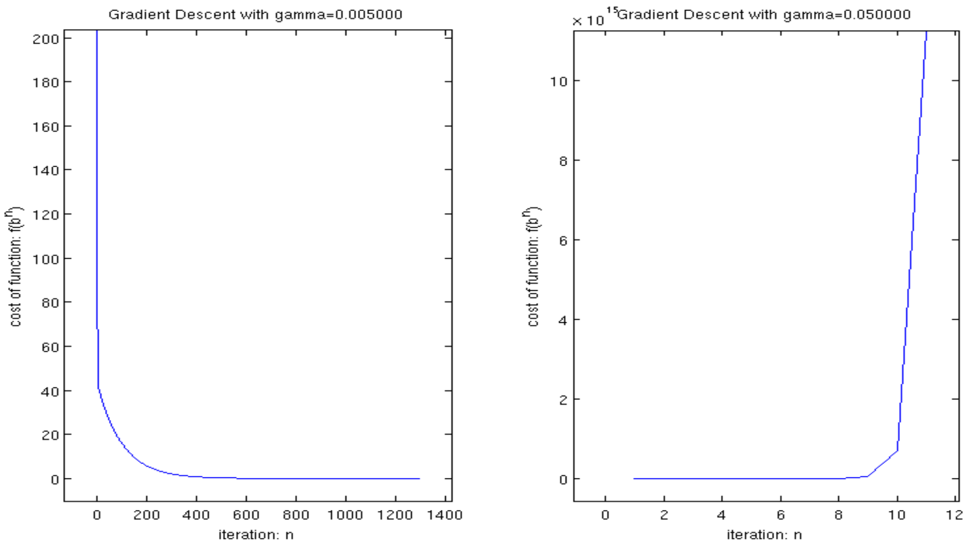

Figure 3: Illustration for gradient descent on $X3$ staring with $b3$ by $\gamma = 0.005$ and 0.05

2 $\gamma = 1$ for the second matrix

Command to get executed:

```
>> gamma = 1;
>> [b2_opt, iters, all_costs] = gd (X2, b2, gamma);
```

Plotting: figure for $\gamma = 1$

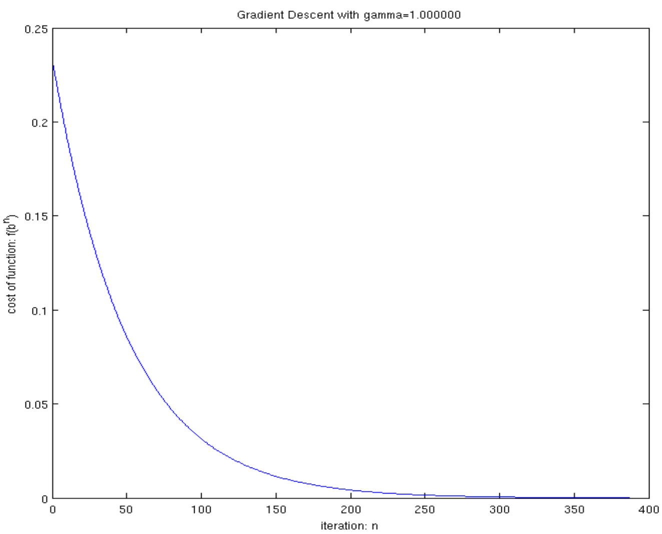

Figure 4: Plotting figure for gradient descent with $\gamma = 1$ on the second matrix

Explanation: Through the smooth plotted curve, we guess that the gradient descent method got linear convergence when $\gamma = 1$ on X_2. Hence, we trace convergence rate conv_rate $= f(x^k)/f(x^{k-1})$ as follows:

```
Iter: 2, Cost: 2.254428e-01, Conv_Rate: 0.980100
Iter: 3, Cost: 2.209565e-01, Conv_Rate: 0.980100
Iter: 4, Cost: 2.165594e-01, Conv_Rate: 0.980100
Iter: 5, Cost: 2.122499e-01, Conv_Rate: 0.980100
Iter: 6, Cost: 2.080261e-01, Conv_Rate: 0.980100
Iter: 7, Cost: 2.038864e-01, Conv_Rate: 0.980100
...
...
Iter: 381, Cost: 1.107934e-04, Conv_Rate: 0.980100
Iter: 382, Cost: 1.085886e-04, Conv_Rate: 0.980100
Iter: 383, Cost: 1.064277e-04, Conv_Rate: 0.980100
Iter: 384, Cost: 1.043098e-04, Conv_Rate: 0.980100
Iter: 385, Cost: 1.022340e-04, Conv_Rate: 0.980100
Iter: 386, Cost: 1.001996e-04, Conv_Rate: 0.980100
Iter: 387, Cost: 9.820558e-05, Conv_Rate: 0.980100
Converged to zeros!
```

In terms of above dumps and the fact that $f(x^*) = 0$, we can conclude that when $\gamma = 1$

$$f(x^{k+1}) - f(x^*) = 0.9801 \cdot \big(f(x^k) - f(x^*)\big)$$

which supports our previous guess that

Gradient Descent with $\gamma = 1$ on second matrix leads to **linear convergence**.

Part II
Written Problems

1 Othorognal Subspace

(a) Show that if U is a subspace, then so is $U^\perp$

Proof. Since U is a subspace, then we have U satisfying all three properties shown below:

- $\mathbf{0} \in U$
- $\forall \mathbf{u}_1, \mathbf{u}_2 \in U, \mathbf{u}_1 + \mathbf{u}_2 \in U$
- $\forall \mathbf{u} \in U, \alpha \in \mathbb{R}, \alpha\mathbf{u} \in U$

Now we show that $U^\perp$ is also a subspace by indicating $U^\perp$ satisfies all three properties as above U does.

- Since $\forall \mathbf{u} \in U, \langle \mathbf{0}, \mathbf{u}\rangle = 0$ and $\mathbf{0} \in V$ $(\mathbf{0} \in U \subseteq V)$, then it turned out that $\mathbf{0} \in U^\perp$.
- Let $\mathbf{u}$ be arbitrary vector s.t. $\mathbf{u} \in U$, and $\mathbf{x}_1$, $\mathbf{x}_2$ to be distinct vector s.t. $\mathbf{x}_1 \in U^\perp$ and $\mathbf{x}_2 \in U^\perp$. By definition of $U^\perp$, we have $\langle \mathbf{x}_1, \mathbf{u}\rangle = 0$ and $\langle \mathbf{x}_2, \mathbf{u}\rangle = 0$. Then it is obvious that $\langle \mathbf{x}_1 + \mathbf{x}_2, \mathbf{u}\rangle = 0$. That is $\mathbf{x}_1 + \mathbf{x}_2 \in U^\perp$. Therefore, $\forall \mathbf{x}_1, \mathbf{x}_2 \in U^\perp, \mathbf{x}_1 + \mathbf{x}_2 \in U^\perp$ was proved.
- Let $\mathbf{x}$ be arbitrary vector s.t. $\mathbf{x} \in U^\perp$, $\mathbf{u}$ be arbitrary vector s.t. $\mathbf{u} \in U$ and arbitrary $\alpha \in \mathbb{R}$. By definition of $U^\perp$, we have $\langle \mathbf{x}, \mathbf{u}\rangle = 0$. Since inner product is linear operator, it is obvious that $\langle \alpha\mathbf{x}, \mathbf{u}\rangle = 0$. That is $\alpha\mathbf{x} \in U^\perp$. Therefore, $\forall \mathbf{x} \in U^\perp, \alpha \in \mathbb{R}, \alpha\mathbf{x} \in U^\perp$ was proved.

Since $U^\perp$ contains $\mathbf{0}$, and is closed under addition and scalar multiplication, it turned out that $U^\perp$ is a subspace. Therefore, the statement that if U is a subspace, then so is $U^\perp$ was proved. □

(b) Show that $(U^\perp)^\perp = U$

Proof. By contradiction. Assume that $(U^\perp)^\perp \neq U$ and then show the contradiction. By definition of $U^\perp$, we have $U^\perp = \{\forall \mathbf{v} \in V, \langle \mathbf{v}, \mathbf{u}\rangle = 0, \forall \mathbf{u} \in U\}$ and $(U^\perp)^\perp = \{\forall \mathbf{x} \in V, \langle \mathbf{x}, \mathbf{v}\rangle = 0, \forall \mathbf{v} \in U^\perp\}$. Since $(U^\perp)^\perp \neq U$, then we can say that $\exists \mathbf{x} \notin U, \langle \mathbf{x}, \mathbf{v}\rangle = 0, \forall \mathbf{v} \in U^\perp$. That is to say, $\exists \mathbf{x} \in V$ *but* $\notin U, \langle \mathbf{x}, \mathbf{v}\rangle = 0, \forall \mathbf{v}\ s.t. \langle \mathbf{v}, \mathbf{u}\rangle = 0, \forall \mathbf{u} \in U$. However, such $\mathbf{x}$ does not exist. Hence, we reject initial assumption and conclude that $(U^\perp)^\perp = U$. □

(c) Show that if $U, W \subseteq V$ are subspaces of V, then $U \subseteq W \Leftrightarrow U^\perp \supseteq W^\perp$

Proof of $U \subseteq W \Rightarrow U^\perp \supseteq W^\perp$. $U^\perp = \{\forall \mathbf{v} \in V, \langle \mathbf{v}, \mathbf{u}\rangle = 0, \forall \mathbf{u} \in U\}$ and $W^\perp = \{\forall \mathbf{v} \in V, \langle \mathbf{v}, \mathbf{w}\rangle = 0, \forall \mathbf{w} \in W\} = \{\forall \mathbf{v} \in V, \langle \mathbf{v}, \mathbf{w}\rangle = 0, \forall \mathbf{w} \in W$ and $\langle \mathbf{v}, \mathbf{u}\rangle = 0, \forall \mathbf{u} \in U\}$ (This is valid because $\forall U \subseteq W$ and then $\mathbf{u} \in W$). Now since membership of $W^\perp$ requires one more condition, then it is obvious that $\mathbf{v} \in W^\perp \Rightarrow \mathbf{v} \in U^\perp$, and $\mathbf{v} \in U^\perp \not\Rightarrow \mathbf{v} \in W^\perp$ hold for arbitrary $\mathbf{v}$. Therefore, we can conclude that $U^\perp \supseteq W^\perp$. □

Proof of $U^\perp \supseteq W^\perp \Rightarrow U \subseteq W$. By definition, we have $U^\perp = \{\mathbf{v} \in V | \langle \mathbf{v}, \mathbf{u}\rangle = 0, \forall \mathbf{u} \in U\}$ and $W^\perp = \{\mathbf{v} \in V | \langle \mathbf{v}, \mathbf{w}\rangle = 0, \forall \mathbf{w} \in W\}$. Since $U^\perp \supseteq W^\perp$, then we have $U^\perp \cap W^\perp = W^\perp$. Then $U^\perp \cap W^\perp = \{\mathbf{v} \in V | \langle \mathbf{v}, \mathbf{u}\rangle = 0, \forall \mathbf{u} \in U$ and $\langle \mathbf{v}, \mathbf{w}\rangle = 0, \forall \mathbf{w} \in W\} = \{\mathbf{v} \in V | \langle \mathbf{v}, \mathbf{x}\rangle = 0, \forall \mathbf{x} \in W \cup U\} = W^\perp$ Then we have $W \cup U = W$, which naturally derives $U \subseteq W$. □

(d) Show that $X^\perp$ makes sense, $X^\perp$ is a subspace and $(X^\perp)^\perp \supseteq X$

Explanation of $X^\perp$ makes sense. $X^\perp$ also makes sense. This is because we can view all vectors in set X as bases of the minimal subspace that is superset of X. Therefore, although the $X^\perp$ is defined through the set X, it makes sense as othogonal complement over the minimal subspace that contains X. □

Proof of $X^\perp$ is a subspace. We prove $X^\perp$ is a subspace by showing $X^\perp$ satisfies three properties as $U^\perp$ above does.

- Since $\forall \mathbf{x} \in X, \langle \mathbf{0}, \mathbf{x}\rangle = 0$ and $\mathbf{0} \in V$, then it turned out that $\mathbf{0} \in X^\perp$.
- Let $\mathbf{x}$ be arbitrary vector s.t. $\mathbf{x} \in U$, and $\mathbf{x}_1^\perp$, $\mathbf{x}_2^\perp$ to be distinct vector s.t. $\mathbf{x}_1^\perp \in X^\perp$ and $\mathbf{x}_2^\perp \in X^\perp$. By definition of $X^\perp$, we have $\langle \mathbf{x}_1^\perp, \mathbf{x}\rangle = 0$ and $\langle \mathbf{x}_2^\perp, \mathbf{x}\rangle = 0$. Then it is obvious that $\langle \mathbf{x}_1^\perp + \mathbf{x}_2^\perp, \mathbf{x}\rangle = 0$. That is, $\mathbf{x}_1^\perp + \mathbf{x}_2^\perp \in X^\perp$. Therefore, $\forall \mathbf{x}_1^\perp, \mathbf{x}_2^\perp \in X^\perp, \mathbf{x}_1^\perp + \mathbf{x}_2^\perp \in X^\perp$ was proved. (That is, membership of $X^\perp$ is closed under vector addition).
- Let $\mathbf{x}^\perp$ be arbitrary vector s.t. $\mathbf{x} \in X^\perp$, $\mathbf{x}$ be arbitrary vector s.t. $\mathbf{x} \in X$ and arbitrary $\alpha \in \mathbb{R}$. By definition of $X^\perp$, we have $\langle \mathbf{x}^\perp, \mathbf{x}\rangle = 0$. Since inner product is linear operator, it is obvious that $\langle \alpha\mathbf{x}^\perp, \mathbf{x}\rangle = 0$. That is $\alpha\mathbf{x}^\perp \in X^\perp$. Therefore, $\forall \mathbf{x}^\perp \in X^\perp, \alpha \in \mathbb{R}, \alpha\mathbf{x}^\perp \in X^\perp$ was proved. (That is, membership of $X^\perp$ is closed under scalar multiplication.)

Since we have shown that $X^\perp$ satisfies all three above properties, then we conclude that $X^\perp$ is a subspace. □

Proof of $(X^\perp)^\perp \supseteq X$. Through the definition of $X^\perp$, it is easy to derive the property for set X

$$\forall x \in X,\ \langle x, x^\perp\rangle = 0,\ \forall x^\perp \in X^\perp \tag{1}$$

And similarly, we can derive the property for subspace $(X^\perp)^\perp$

$$\forall x \in (X^\perp)^\perp,\ \langle x, x^\perp\rangle = 0,\ \forall x^\perp \in X^\perp \tag{2}$$

It is obvious that, all elements in set X and subspace $(X^\perp)^\perp$ are perpendicular to any vectors in $X^\perp$. Since X is a set with finite element, and $(X^\perp)^\perp$ is a subspace with infinite elements, it is natural to derive that

$$(X^\perp)^\perp \supseteq X \tag{3}$$

□

(e) Show that any $v \in V$ can be written uniquely as $v = u + u^{\perp}$

Proof. Assume that the representation of arbitrary vector $v \in V$ is not unique. Let u_1, u_2 be distinct vector in space U and $u_1^{\perp}$,$u_2^{\perp}$ be distinct vector in space $U^{\perp}$ such that $v = u_1 + u_1^{\perp}$ and $v = u_2 + u_2^{\perp}$. Then

$$0 = v - v = u_1 + u_1^{\perp} - u_2 - u_2^{\perp} \tag{4}$$

Then we have

$$0 = (u_1 + u_1^{\perp} - u_2 - u_2^{\perp})^T (u_1 + u_1^{\perp} - u_2 - u_2^{\perp}) \tag{5}$$
$$0 = u_1^T (u_1 - u_2) - u_2^T (u_1 - u_2) + u_1^{\perp} (u_1^{\perp} - u_2^{\perp}) - u_2^{\perp} (u_1^{\perp} - u_2^{\perp}) \tag{6}$$
$$0 = (u_1 - u_2)^T (u_1 - u_2) + (u_1^{\perp} - u_2^{\perp})^T (u_1^{\perp} - u_2^{\perp}) \tag{7}$$
$$0 = ||u_1 - u_2||_2^2 + ||u_1^{\perp} - u_2^{\perp}||_2^2 \tag{8}$$

Since

$$||u_1 - u_2||_2^2 \geq 0 \text{ and } ||u_1^{\perp} - u_2^{\perp}||_2^2 \geq 0 \tag{9}$$

In terms of (7), we have is that

$$||u_1 - u_2||_2 = 0 \text{ and } ||u_1^{\perp} - u_2^{\perp}||_2 = 0 \tag{10}$$

Hence, we have

$$u_1 = u_2 \text{ and } u_1^{\perp} = u_2^{\perp} \tag{11}$$

which indicates uniqueness of representation of arbitrary vector $v \in V$. □

2 Boyd and Vandenberghe, Ex. 2.10

(a) Show that if $A \in \mathbb{S}_+^n$ then the set C is convex

Proof. Assume that $A \in \mathbb{S}_+^n$. Let $\mathbf{x}_1$, $\mathbf{x}_2$ to be arbitrary vector such that $\mathbf{x}_1 \in C$, $\mathbf{x}_2 \in C$ and $\mathbf{x}_1 \neq \mathbf{x}_2$. Then we show arbitrary linear combination $\mathbf{x}_* = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2,\ \lambda \in [0,1]$ also belongs to the set C. According to the definition of set C, we have

$$\mathbf{x}_1^T A\mathbf{x}_1 + b^T\mathbf{x}_1 + c \leq 0 \tag{12}$$
$$\mathbf{x}_2^T A\mathbf{x}_2 + b^T\mathbf{x}_2 + c \leq 0 \tag{13}$$

By positive semidefiniteness of A, we have

$$\mathbf{x}^T A\mathbf{x} \geq 0,\ \forall\mathbf{x} \tag{14}$$

That is

$$\mathbf{x}_1^T A\mathbf{x}_1 \geq 0 \tag{15}$$
$$\mathbf{x}_2^T A\mathbf{x}_2 \geq 0 \tag{16}$$
$$(\mathbf{x}_2 - \mathbf{x}_1)^T A(\mathbf{x}_2 - \mathbf{x}_1) \geq 0 \tag{17}$$

Besides,

$$\lambda \geq 0 \tag{18}$$
$$1 - \lambda \geq 0 \tag{19}$$
$$\lambda - 1 \leq 0 \tag{20}$$

Then we investigate the property of $\mathbf{x}_*^T A\mathbf{x}_* + b^T\mathbf{x}_* + c$.

$$\begin{aligned}
&\mathbf{x}_*^T A\mathbf{x}_* + b^T\mathbf{x}_* + c && (21)\\
&= (\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)^T A(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + b^T(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + c && (22)\\
&= \lambda^2\mathbf{x}_1^T A\mathbf{x}_1 + (1-\lambda)^2\mathbf{x}_2^T A\mathbf{x}_2 + \lambda(1-\lambda)(\mathbf{x}_1^T A\mathbf{x}_2 + \mathbf{x}_2^T A\mathbf{x}_1) + \lambda b^T\mathbf{x}_1 + (1-\lambda)b^T\mathbf{x}_2 + c && (23)\\
&= \lambda^2\mathbf{x}_1^T A\mathbf{x}_1 - \lambda\mathbf{x}_1^T A\mathbf{x}_1 + (1-\lambda)^2\mathbf{x}_2^T A\mathbf{x}_2 - (1-\lambda)\mathbf{x}_2^T A\mathbf{x}_2\\
&+ \lambda(\mathbf{x}_1^T A\mathbf{x}_1 + b^T\mathbf{x}_1 + c) + (1-\lambda)(\mathbf{x}_2^T A\mathbf{x}_2 + b^T\mathbf{x}_2 + c) + \lambda(1-\lambda)(\mathbf{x}_1^T A\mathbf{x}_2 + \mathbf{x}_2^T A\mathbf{x}_1) && (24)\\
&\leq \lambda(\lambda-1)(\mathbf{x}_1^T A\mathbf{x}_1 + \mathbf{x}_2^T A\mathbf{x}_2) + \lambda(1-\lambda)(\mathbf{x}_1^T A\mathbf{x}_2 + \mathbf{x}_2^T A\mathbf{x}_1) && (25)\\
&= \lambda(\lambda-1)((\mathbf{x}_2 - \mathbf{x}_1)^T A(\mathbf{x}_2 - \mathbf{x}_1)) && (26)\\
&\leq 0 && (27)
\end{aligned}$$

Then, it is obvious that

$$\mathbf{x}_* \in C \tag{28}$$

Since $\mathbf{x}_*$ is arbitrary convex combination of $\mathbf{x}_1$ and $\mathbf{x}_2$, then

$$C \text{ is convex set.} \tag{29}$$

Hence, it is proved that

$$A \in \mathbb{S}_+^n \Rightarrow C \text{ is convex.} \tag{30}$$

□

(b) Show that C_1 is convex if there exists $\lambda \in \mathbb{R}$ such that $(A + \lambda g g^T) \in \mathbb{S}^n_+$

Proof. Assume that there exists $\lambda \in \mathbb{R}$ such that $(A + \lambda g g^T) \in \mathbb{S}^n_+$. Let λ^* be the λ that satisfies $(A + \lambda g g^T) \in \mathbb{S}^n_+$. Now we show that

$$\begin{aligned} C_1 &= C \cap \{\mathbf{x} \in \mathbb{R}^n : g^T\mathbf{x} + h = 0\} && (31)\\ &= \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^T A\mathbf{x} + b^T\mathbf{x} + c \le 0 \text{ and } g^T\mathbf{x} + h = 0\} && (32)\end{aligned}$$

is convex.
Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be arbitrary vector that $\mathbf{x}_1 \in C_1$, $\mathbf{x}_2 \in C_1$ and $\mathbf{x}_1 \neq \mathbf{x}_2$. Let $\mathbf{x}_*$ to be the arbitrary convex combination of $\mathbf{x}_1$ and $\mathbf{x}_2$ as $\mathbf{x}_* = \lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2$, and next we show that $\mathbf{x}_* \in C_1$.
By the definition of C_1 (See. (31)), we have

$$\begin{aligned} \mathbf{x}_1^T A\mathbf{x}_1 + b^T\mathbf{x}_1 + c \le 0 \text{ and } g^T\mathbf{x}_1 + h = 0 && (33)\\ \mathbf{x}_2^T A\mathbf{x}_2 + b^T\mathbf{x}_2 + c \le 0 \text{ and } g^T\mathbf{x}_2 + h = 0 && (34)\end{aligned}$$

Then we have

$$\begin{aligned} g^T\mathbf{x}_2 + h - (g^T\mathbf{x}_1 + h) = 0 && (35)\\ g^T(\mathbf{x}_2 - \mathbf{x}_1) = 0 && (36)\end{aligned}$$

Derivation for $g^T\mathbf{x}_* + h = 0$ is as follows:

$$\begin{aligned} & g^T\mathbf{x}_* + h && (37)\\ &= g^T(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + h && (38)\\ &= \lambda g^T\mathbf{x}_1 + (1-\lambda)g^T\mathbf{x}_2 + h && (39)\\ &= \lambda(g^T\mathbf{x}_1 + h) + (1-\lambda)(g^T\mathbf{x}_2 + h) && (40)\\ &= 0 && (41)\end{aligned}$$

Derivation for $\mathbf{x}_*^T A\mathbf{x}_* + b^T\mathbf{x}_* + c \le 0$ is as follows:

$$\begin{aligned} & \mathbf{x}_*^T A\mathbf{x}_* + b^T\mathbf{x}_* + c && (42)\\ &= (\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2)^T A(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + b^T(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) + c && (43)\\ &= \lambda^2\mathbf{x}_1^T A\mathbf{x}_1 + (1-\lambda)^2\mathbf{x}_2^T A\mathbf{x}_2 + \lambda(1-\lambda)(\mathbf{x}_1^T A\mathbf{x}_2 + \mathbf{x}_2^T A\mathbf{x}_1) + \lambda b^T\mathbf{x}_1 + (1-\lambda)b^T\mathbf{x}_2 + c && (44)\\ &= \lambda^2\mathbf{x}_1^T A\mathbf{x}_1 - \lambda\mathbf{x}_1^T A\mathbf{x}_1 + (1-\lambda)^2\mathbf{x}_2^T A\mathbf{x}_2 - (1-\lambda)\mathbf{x}_2^T A\mathbf{x}_2 \\ &+ \lambda(\mathbf{x}_1^T A\mathbf{x}_1 + b^T\mathbf{x}_1 + c) + (1-\lambda)(\mathbf{x}_2^T A\mathbf{x}_2 + b^T\mathbf{x}_2 + c) + \lambda(1-\lambda)(\mathbf{x}_1^T A\mathbf{x}_2 + \mathbf{x}_2^T A\mathbf{x}_1) && (45)\\ &\le \lambda(\lambda-1)(\mathbf{x}_1^T A\mathbf{x}_1 + \mathbf{x}_2^T A\mathbf{x}_2) + \lambda(1-\lambda)(\mathbf{x}_1^T A\mathbf{x}_2 + \mathbf{x}_2^T A\mathbf{x}_1) && (46)\\ &= \lambda(\lambda-1)((\mathbf{x}_2 - \mathbf{x}_1)^T A(\mathbf{x}_2 - \mathbf{x}_1)) && (47)\\ &= \lambda(\lambda-1)\Big((\mathbf{x}_2 - \mathbf{x}_1)^T(A + \lambda^* g g^T)(\mathbf{x}_2 - \mathbf{x}_1) - (\mathbf{x}_2 - \mathbf{x}_1)^T(\lambda^* g g^T)(\mathbf{x}_2 - \mathbf{x}_1)\Big) && (48)\\ &\le \lambda(1-\lambda)(\mathbf{x}_2 - \mathbf{x}_1)^T(\lambda^* g g^T)(\mathbf{x}_2 - \mathbf{x}_1) && (49)\\ &= 0 && (50)\end{aligned}$$

Then we have

$$\mathbf{x}_*^T A\mathbf{x}_* + b^T\mathbf{x}_* + c \le 0 \text{ and } g^T\mathbf{x}_* + h = 0 \quad (51)$$

That is to say, for arbitrary convex combination $\mathbf{x}_* \in C_1$. Then

$$C_1 \text{ is convex set.} \quad (52)$$

Hence, it is proved that

$$\exists\lambda \in \mathbb{R},\ (A + \lambda g g^T) \in \mathbb{S}^n_+ \Rightarrow C_1 \text{ is convex.} \quad (53)$$

□

3 Boyd and Vandenberghe, Ex. 2.21

Proof. Let $(a_1, b_1) \in S$, $(a_2, b_2) \in S$ and $(a_1, b_1) \neq (a_2, b_2)$, then show that arbitrary convex combination $(a_*, b_*) = (\lambda a_1 + (1-\lambda)a_2, \lambda b_1 + (1-\lambda)b_2) \in S$.
By definition of set S, we have

$$a_1^T x \leq b_1 \ \forall x \in C, \text{ and } a_1^T x \geq b_1 \ \forall x \in D \tag{54}$$
$$a_2^T x \leq b_2 \ \forall x \in C, \text{ and } a_2^T x \geq b_2 \ \forall x \in D \tag{55}$$

Then we turn to discuss $a_*^T x_c$ as follows: let x_c be arbitrary vector in set C,

$$a_*^T x_c = (\lambda a_1 + (1-\lambda)a_2)^T x_c \tag{56}$$
$$= \lambda a_1^T x_c + (1-\lambda) a_2^T x_c \tag{57}$$
$$\leq \lambda b_1 + (1-\lambda) b_2 \tag{58}$$
$$= b_* \tag{59}$$

and now let x_d be arbitrary vector in set D,

$$a_*^T x_d = (\lambda a_1 + (1-\lambda)a_2)^T x_d \tag{60}$$
$$= \lambda a_1^T x_d + (1-\lambda) a_2^T x_d \tag{61}$$
$$\geq \lambda b_1 + (1-\lambda) b_2 \tag{62}$$
$$= b_* \tag{63}$$

Now since we have $a_*^T x_c \leq b_* \ \forall x_c \in C$ and $a_*^T x_d \geq b_* \ \forall x_d \in D$, then we can conclude that

$$(a_*, b_*) \in S \tag{64}$$

Hence, it is proved that

$$S \text{ is convex.} \tag{65}$$

□

7 $\{x : ||x - v_1|| \leq ||x - v_2||\} = \{x : c^T x \leq d\}$

Proof.

$$||x - v_1|| \leq ||x - v_2|| \tag{66}$$
$$\Leftrightarrow \quad ||x - v_1||^2 \leq ||x - v_2||^2 \tag{67}$$
$$\Leftrightarrow \quad (x - v_1)^T (x - v_1) \leq (x - v_2)^T (x - v_2) \tag{68}$$
$$\Leftrightarrow \quad x^T x - x^T v_1 - v_1^T x + v_1^T v_1 \leq x^T x - x^T v_2 - v_2^T x + v_2^T v_2 \tag{69}$$
$$\Leftrightarrow \quad -v_1^T x - v_1^T x + v_1^T v_1 \leq -v_2^T x - v_2^T x + v_2^T v_2 \tag{70}$$
$$\Leftrightarrow \quad 2(v_2 - v_1)^T x \leq v_2^T v_2 - v_1^T v_1 \tag{71}$$

Let $c = 2(v_2 - v_1)$ and $d = v_2^T v_2 - v_1^T v_1 = ||v_2||^2 - ||v_1||^2$, then

$$c^T x \leq d \tag{72}$$

where $c = 2(v_2 - v_1)$ and $d = ||v_2||^2 - ||v_1||^2$.
Hence, we can conclude that

$$\{x : ||x - v_1|| \leq ||x - v_2||\} = \{x : c^T x \leq d\} \tag{73}$$

□

8 Exists C such that $CA = B$

Proof. By Contradiction. Assume that

$$\nexists C,\ CA = B \tag{74}$$

Then let the A_j denotes the j-th column vector of A and the B_j denotes the j-th column vector of B,

$$\nexists C,\ \forall j,\ CA_j = B_j \tag{75}$$

This naturally derives

$$\nexists C,\ A_j = C^{-1}B_j \tag{76}$$

Let x be the one that $Ax = 0$, that is $\sum_j A_j x_j = 0$.

$$\nexists C,\ \sum_j C^{-1}B_j x_j = 0 \tag{77}$$

$$\Leftrightarrow \nexists C,\ C^{-1}\sum_j B_j x_j = 0 \tag{78}$$

Since we have

$$Ax = 0 \Rightarrow Bx = 0 \tag{79}$$

$$\tag{80}$$

Then we have

$$Bx = 0 \tag{81}$$

$$\Leftrightarrow \sum_j B_j x_j = 0 \tag{82}$$

In terms of (77) and (81), we have

$$\nexists C,\ C^{-1} \cdot 0 = 0 \tag{83}$$

which contradicts the common sense that any arbitrary invertible matrix C would satisfies $C^{-1} \cdot 0 = 0$. Hence, the initial assumption should be rejected and then it is proved that

$$\exists C \text{ such that } CA = B \tag{84}$$

$\square$

A Codes Printout

(a) Gradient Descent Routine

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% HW2: Gradient Descent
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [b, iter, all_costs] = Gradient_Descent(X, b_init, gamma)

eps = 10e-5
b = b_init;
last_cost = 0.5 * b' * X * b;
iter = 1;
all_costs = [];
while true,
    %% compute essential numerics and do gradient descent
    gradient = X * b;
    b = b - gamma * gradient;
    cost = 0.5 * b' * X * b;
    rate = (last_cost / cost);
    all_costs = [all_costs cost];
    %% output numeric information of this iteration
    iter_str = sprintf('Iter: %d, Cost: %e, Conv_Rate: %f',iter,cost,rate);
    disp(iter_str);
    %% quadratic optimization converges to zero
    if cost < eps,
        disp('Converged to zeros!')
        break
    end
    %% qudratic optimization diverges
    if cost >= last_cost && iter > 10,
        disp('Problem diverges!')
        break
    end
    %% prepare for next iteration
    last_cost = cost;
    iter = iter + 1;
end

%% uncomment following code for plotting individual gradient descent run
%plot f(b^(n)) with regard to n
%plot (1:iter, all_costs)
%title (sprintf ('Gradient Descent with gamma=%f', gamma))
%xlabel ('iteration: n')
%ylabel ('cost of function: f(b^n)')

end
```

(b) Running Script

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% Running scripts for applying gradient descent
%%% on three given dataset
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% for X1, b1
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
gamma1_one = 0.5; gamma2_two = 3;
[b1_opt_one, iter1_one, costs1_one] = Gradient_Descent(X1, b1, gamma1_one);
[b1_opt_two, iter1_two, costs1_two] = Gradient_Descent(X1, b1, gamma2_two);
subplot (1, 2, 1)
plot (1:iter1_one, costs1_one)
axis ([-0.1*iter1_one 1.1*iter1_one -0.05*max(costs1_one) max(costs1_one)])
title (sprintf ('Gradient Descent with gamma=%f', gamma1_one))
xlabel ('iteration: n')
ylabel ('cost of function: f(b^n)')
subplot (1, 2, 2)
plot (1:iter1_two, costs1_two)
axis ([-0.1*iter1_two 1.1*iter1_two -0.05*max(costs1_two) max(costs1_two)])
title (sprintf ('Gradient Descent with gamma=%f', gamma2_two))
xlabel ('iteration: n')
ylabel ('cost of function: f(b^n)')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% for X2, b2
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
gamma2_one = 1.5; gamma2_two = 3;
[b2_opt_one, iter2_one, costs2_one] = Gradient_Descent(X2, b2, gamma2_one);
[b2_opt_two, iter2_two, costs2_two] = Gradient_Descent(X2, b2, gamma2_two);
figure()
subplot (1, 2, 1)
plot (1:iter2_one, costs2_one)
axis ([-0.1*iter2_one 1.1*iter2_one -0.05*max(costs2_one) max(costs2_one)])
title (sprintf ('Gradient Descent with gamma=%f', gamma2_one))
xlabel ('iteration: n')
ylabel ('cost of function: f(b^n)')
subplot (1, 2, 2)
plot (1:iter2_two, costs2_two)
axis ([-0.1*iter2_two 1.1*iter2_two -0.05*max(costs2_two) max(costs2_two)])
title (sprintf ('Gradient Descent with gamma=%f', gamma2_two))
xlabel ('iteration: n')
ylabel ('cost of function: f(b^n)')

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% for X3, b3
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
gamma3_one = 0.005; gamma3_two = 0.05;
[b3_opt_one, iter3_one, costs3_one] = Gradient_Descent(X3, b3, gamma3_one);
[b3_opt_two, iter3_two, costs3_two] = Gradient_Descent(X3, b3, gamma3_two);
figure()
subplot (1, 2, 1)
plot (1:iter3_one, costs3_one)
axis ([-0.1*iter3_one 1.1*iter3_one -0.05*max(costs3_one) max(costs3_one)])
title (sprintf ('Gradient Descent with gamma=%f', gamma3_one))
xlabel ('iteration: n')
ylabel ('cost of function: f(b^n)')
subplot (1, 2, 2)
plot (1:iter3_two, costs3_two)
axis ([-0.1*iter3_two 1.1*iter3_two -0.05*max(costs3_two) max(costs3_two)])
title (sprintf ('Gradient Descent with gamma=%f', gamma3_two))
xlabel ('iteration: n')
ylabel ('cost of function: f(b^n)')
```