AMATH516 - hw1

XIANGYU GAO 1826374

April 2019

1 Chapter 1

Exercise 1.1

From definition:

$$\mathcal{A}(X) := (\langle A_1, X\rangle, \langle A_2, X\rangle, \ldots, \langle A_l, X\rangle), Y := (y_1, y_2, \ldots, y_l)$$

$$\langle A(X), Y\rangle = \sum_{i=1}^{l} \langle A_i, X\rangle y_i = \sum_{i=1}^{l} \langle y_i A_i, X\rangle = \left\langle \sum_{i=1}^{l} y_i A_i, X\right\rangle$$

For any linear mapping $\mathcal{A} : \mathbf{E} \to \mathbf{Y}$, there exists a unique linear mapping $\mathcal{A}^* : \mathbf{Y} \to \mathbf{E}$, called the adjoint, satisfying

$$\langle \mathcal{A}x, y\rangle = \langle x, \mathcal{A}^* y\rangle \quad \text{for all points} \quad x \in \mathbf{E}, y \in \mathbf{Y}$$

Therefore, $\mathcal{A}^* Y = \sum_{i=1}^{l} y_i A_i = y_1 A_1 + y_2 A_2 + \ldots + y_l A_l$

Exercise 1.2

An inner-product on $\mathbf{E}$ is an assignment $\langle \cdot, \cdot\rangle : \mathbf{E} \times \mathbf{E} \to \mathbf{R}$ satisfying three properties for all $x, y, z \in \mathbf{E}$ and scalars $a, b \in \mathbf{R}$:

(1) Symmetry: $\langle v, w\rangle_{\mathcal{A}} := \langle \mathcal{A}v, w\rangle = \langle v, \mathcal{A}^* w\rangle = \langle \mathcal{A}w, v\rangle = \langle \mathcal{A}w, v\rangle$, because A is symmetric $A = A^*$

(2)Bilinearity: $\langle av + bw, z\rangle_{\mathcal{A}} := \langle \mathcal{A}(av + bw), z\rangle = a\langle \mathcal{A}v, z\rangle + b\langle \mathcal{A}w, z\rangle = a\langle v, z\rangle_{\mathcal{A}} + b\langle w, z\rangle_{\mathcal{A}}$

(3) Positive definiteness: $\langle v, v\rangle_{\mathcal{A}} = \langle \mathcal{A}v, v\rangle > 0 \quad$ for all $0 \neq v \in \mathbf{E}$, because A is positive definite operator

The inner product in the Euclidean space $\mathbf{E}$ always induces a norm $\|x\| := \sqrt{\langle x, x\rangle}$. Thus, $\|v\|_{\mathcal{A}} = \sqrt{\langle v, v\rangle_{\mathcal{A}}} = \sqrt{\langle \mathcal{A}v, v\rangle}$

The dual norm

$$
\begin{aligned}
\|v\|_{\mathcal{A}}^{*} &= \max\left\{\langle v, x\rangle : \|x\|_{\mathcal{A}} \le 1\right\} \\
&= \max\{\langle v, x\rangle : \sqrt{\langle \mathcal{A}x, x\rangle} \le 1\} \\
&= \max\left\{\left\langle v, \mathcal{A}^{-\frac{1}{2}}x\right\rangle : \sqrt{\left\langle \mathcal{A}^{\frac{1}{2}}x, \mathcal{A}^{-\frac{1}{2}}x\right\rangle} \le 1\right\} \\
&= \max\left\{\left\langle \mathcal{A}^{-\frac{1}{2}}v, x\right\rangle : \sqrt{\left\langle \mathcal{A}^{\frac{1}{2}}x, \mathcal{A}^{-\frac{1}{2}}x\right\rangle} \le 1\right\} \\
&= \max\left\{\left\langle \mathcal{A}^{-\frac{1}{2}}v, x\right\rangle : \sqrt{\langle x, x\rangle} \le 1\right\} \\
&= \|\mathcal{A}^{-\frac{1}{2}}v\|\|x\| \; since the Cauchy–Schwarz inequality \\
&= \|\mathcal{A}^{-\frac{1}{2}}v\| \\
&= \left\langle \mathcal{A}^{-\frac{1}{2}}v, \mathcal{A}^{-\frac{1}{2}}v\right\rangle \\
&= \sqrt{\langle \mathcal{A}^{-1}v, v\rangle} \\
&= \|v\|_{\mathcal{A}^{-1}}
\end{aligned}
$$

Exercise 1.3

The induced matrix norm implies, $\|\mathcal{A}\|_{a,b} := \max_{x:\|x\|_a \le 1} \|\mathcal{A}x\|_b$.

Thus, $\|\mathcal{A}\|_1 := \max_{x:\|x\|_1 \le 1} \|\mathcal{A}x\|_1 = \|\mathcal{A}x\|_1, with\, \|x\|_1 == 1$.

For $x \in R^n$ and $\|x\|_1 = 1$,

$\|Ax\|_1 = \left\|\sum_{j=1}^{n} x_j a_j\right\|_1 \le \sum_{j=1}^{n} |x_j| \, \|a_j\|_1 \le \max_{1\le j\le n} \|a_j\|_1 \, \|x\|_1 = max_{1\le j\le n} \|a_j\|_1$

Similarly, $\|\mathcal{A}\|_\infty := \max_{x:\|x\|_\infty \le 1} \|\mathcal{A}x\|_\infty = \|\mathcal{A}x\|_\infty, with\, \|x\|_\infty = 1$

For $x \in R^n$ and $\|x\|_\infty = 1$,

$\|Ax\|_\infty = \max_{1\le i\le m} |a_i x| \le \max_{1\le i\le m} \left\|a_i^T\right\|_1 \|x\|_\infty$

Exercise 1.6

(1) $f(x) = \frac{1}{2}\langle\frac{1}{2}(\mathcal{A} + \mathcal{A}^*)x, x\rangle + \langle v, x\rangle + c = \frac{1}{4}\langle \mathcal{A}x + \mathcal{A}^*x, x\rangle + \langle v, x\rangle + c = \frac{1}{4}\langle \mathcal{A}x, x\rangle + \frac{1}{4}\langle \mathcal{A}^*x, x\rangle + \langle v, x\rangle + c$

Because of the definition of adjoint: $\langle \mathcal{A}x, x\rangle = \langle x, \mathcal{A}^*x\rangle$

Thus $f(x) = \frac{1}{4}\langle \mathcal{A}x, x\rangle + \frac{1}{4}\langle \mathcal{A}^*x, x\rangle + \langle v, x\rangle + c = \frac{1}{2}\langle \mathcal{A}x, x\rangle + \langle v, x\rangle + c$

(2)$f(x) = \frac{1}{2}\langle \mathcal{A}x, x\rangle + \langle v, x\rangle + c = \frac{1}{2}(\mathcal{A}x)^T x + v^T x + c = \frac{1}{2}x^T \mathcal{A}^T x + v^T x + c$

$\nabla f(x) = \frac{\partial f(x)}{\partial x} = \frac{1}{2}(\mathcal{A}+\mathcal{A}^T)x+v = \mathcal{A}x+v$, because $\mathcal{A}$ is self adjoint $\mathcal{A} = \mathcal{A}^*$

$\nabla^2 f(x) = \frac{\partial \nabla f(x)}{\partial x} = \mathcal{A}$

(3)From part 1, when $\mathcal{A}$ is replaced by the self-adjoint operator $(\mathcal{A} + \mathcal{A}^*)/2$, the function values $f(x)$ remain unchanged.

Thus let $f(x) = \frac{1}{2}\langle \frac{1}{2}(\mathcal{A} + \mathcal{A}^*)x, x\rangle + \langle v, x\rangle + c$.

From part 2, $\nabla f(x) = \frac{1}{2}(\mathcal{A} + \mathcal{A}^*)x + v$, $\nabla^2 f(x) = \frac{1}{2}(\mathcal{A} + \mathcal{A}^*)$

Exercise 1.7

$f(x) = \frac{1}{2}\|F(x)\|^2$, where $F : \mathbf{E} \to \mathbf{Y}$ is a C^1 -smooth mapping.

$f(x) = \frac{1}{2}\langle F(x), F(x)\rangle$, from the derivative chain rule,

$\nabla f(x) = \frac{\partial f(x)}{\partial x} = \frac{\partial f}{\partial F}\frac{\partial F}{\partial x} = \nabla F(x)^* F(x)$

Exercise 1.8

Consider a function $f : U \to \mathbf{R}$ and a linear mapping $\mathcal{A} : \mathbf{Y} \to \mathbf{E}$ and define the composition $h(x) = f(\mathcal{A}x)$

(1) Denote by J_g the Jacobian matrix of any function g .Applying the chain rule leads to $J_q(x) = J_f(\phi(x))J_\phi(x)$

Since $J_\phi(x) = A$, and $\nabla g = (J_g)^T$ for any scalar function g, this boils down to $(\nabla h(x))^T = (\nabla f(Ax + b))^T A$

Finally, we find: $\nabla h(x) = A^T \nabla f(Ax + b)$

(2) Similarly, if f is twice differentiable at $\mathcal{A}x$, then

$(\nabla^2 h(x))^T = A^T(\nabla^2 f(Ax + b))^T A$

$\nabla^2 h(x) = A^T \nabla^2 f(Ax + b)A$

Exercise 1.9

H is differentiable at x, so we have $H(x+l) = H(x) + \langle \nabla H(x), l\rangle + o(\|l\|)$.$G$ is differentiable at $H(X)$, we have $G(H(x) + h) = G(H(x)) + \langle \nabla G(H(x)), h\rangle + o(\|h\|)$. Therefore

$$\begin{aligned}
\nabla F(x) &= \lim_{l \to 0} \frac{F(x + l) - F(x)}{l} \\
&= \lim_{l \to 0} \frac{G(H(x + l)) - G(H(x))}{l} \\
&= \lim_{l \to 0} \frac{G(H(x) + \langle \nabla H(x), l\rangle + o(\|l|)) - G(H(x))}{l}
\end{aligned}$$

Let $h = \langle \nabla H(x), l \rangle + o(\|l\|)$, we have

$$
\begin{aligned}
\nabla F(x) &= \lim_{l \to 0} \frac{G(H(x) + \langle \nabla H(x), l \rangle + o(\|l\|)) - G(H(x))}{l} \\
&= \lim_{l \to 0} \frac{\langle \nabla G(H(x)), h \rangle + o(\|h\|)}{l} \\
&= \nabla G(H(x)) \nabla H(x), \text{ since } h = \langle \nabla H(x), l \rangle + o(\|l\|)
\end{aligned}
$$

Exercise 1.10

Consider the two functions $f : \mathbf{R}^n_{++} + \to \mathbf{R}$ and $F : \mathbf{S}^n_{++} \to \mathbf{R}$ given by $f(x) = -\sum_{i=1}^n \log x_i$ and $F(X) = -\ln \det(X)$

(1) For $x \in \mathbf{R}^n_{++}$

$$
\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial - \sum_{i=1}^n \log x_i}{\partial x_1} \\ \frac{\partial - \sum_{i=1}^n \log x_i}{\partial x_2} \\ \vdots \\ \frac{\partial - \sum_{i=1}^n \log x_i}{\partial x_n} \end{pmatrix} = \begin{pmatrix} -\frac{1}{x_1} \\ -\frac{1}{x_2} \\ \vdots \\ -\frac{1}{x_i} \end{pmatrix}
$$

$$
\nabla^2 f(x) = \begin{pmatrix} \frac{1}{x_1^2} \\ \frac{1}{x_2^2} \\ \vdots \\ \frac{1}{x_i^2} \end{pmatrix}
$$

(2)

$$
\begin{aligned}
&F(X + tV) - F(X) + t \left\langle X^{-1}, V \right\rangle \\
&= -\ln \det(X + tV) + \ln \det(X) + t \left\langle X^{-1}, V \right\rangle \\
&= -\ln \det(X + tV) + \ln \det(X) + t \left\langle X^{-1}, V \right\rangle \\
&= -\ln \det \left(I + tX^{-1/2} V X^{-1/2} \right) + t \left(X^{-1/2} V X^{-1/2} \right) \\
&= -\ln \left(1 + \operatorname{tr} \left(tX^{-1/2} V X^{-1/2} \right) + o \left(\left\| tX^{-1/2} V X^{-1/2} \right\| \right) \right) + t \left(X^{-1/2} V X^{-1/2} \right) \\
&= -\left(t \operatorname{tr} \left(X^{-1/2} V X^{-1/2} \right) + o \left(\left\| t^2 \right\| \right) \right) + t \left(X^{-1/2} V X^{-1/2} \right) \\
&= o(t)
\end{aligned}
$$

Hence, $\nabla F(X) = -X^{-1}$

To compute the Hessian, observe $(X+V)^{-1} = X^{-1/2} \left(I + X^{-1/2} V X^{-1/2} \right)^{-1} X^{-1/2}$ and $(I+A)^{-1} = I - A + A^2 - A^3 + \ldots = I - A + O\left(\|A\|^2_{0p} \right)$ whenever $\|A\|_{op} < 1$.

We have

$$
\begin{aligned}
\nabla F(X + tV) - \nabla F(X) &= -(X + tV)^{-1} + (X)^{-1} \\
&= -X^{-1/2} \left(I + X^{-1/2} tV X^{-1/2} \right)^{-1} X^{-1/2} + (X)^{-1} \\
&= -X^{-1/2} \left(I - X^{-1/2} tV X^{-1/2} + O(t) \right) X^{-1/2} + (X)^{-1} \\
&= X^{-1} tV X^{-1} + O(t)
\end{aligned}
$$

So, $\nabla^2 F(X)[V] = X^{-1} V X^{-1}$

(3) $\left\langle \nabla^2 F(X)[V], V \right\rangle = tr((\nabla^2 F(X)[V])^T V) = tr(X^{-1} V X^{-1} V^T)$

$= tr((X^{-\frac{1}{2}})^T X^{-\frac{1}{2}} V (X^{-\frac{1}{2}})^T X^{-\frac{1}{2}} V^T)$

$= tr((X^{-\frac{1}{2}} V X^{-\frac{1}{2}})^T X^{-\frac{1}{2}} V X^{-\frac{1}{2}})$

$= \left\| X^{-\frac{1}{2}} V X^{-\frac{1}{2}} \right\|_F^2$, for any $X > 0$ and $V \in \mathcal{S}^n$

$\left\langle \nabla^2 F(X)[V], V \right\rangle = \left\| X^{-\frac{1}{2}} V X^{-\frac{1}{2}} \right\|_F^2 > 0$, for any$X > 0, V \neq 0$. So, $\nabla^2 F(X) : \mathbf{S}^n \to \mathbf{S}^n$ is positive definite.

Exercise 1.11

Consider a function $f : U \to \mathbf{R}$ and two points $x, y \in U$ Define the univariate function $\varphi : [0,1] \to \mathbf{R}$ given by $\varphi(t) = f(x + t(y - x))$ and let $x_t := x + t(y - x)$ for any t.

(1) if f is C^1 -smooth, then according to the chain rule,

$\varphi'(t) = (\nabla f(x_t))^T (y - x) = \langle \nabla f(x_t), y - x \rangle$, holds for any $t \in (0,1)$

(2) if f is C^2 -smooth, then according to the chain rule,

$\varphi''(t) = \frac{\partial \varphi'(t)}{\partial t} = ((\nabla^2 f(x_t))^T (y - x))^T (y - x) = (y - x)^T \nabla^2 f(x_t)(y - x)$

$= \left\langle \nabla^2 f(x_t)(y - x), y - x \right\rangle$, holds for any $t \in (0,1)$

Exercise 1.15

Consider a C^1 -smooth mapping $F : U \to \mathbf{R}^m$ and two points $x, y \in U$.

Define the univariate function $\varphi(t) = F(x + t(y - x))$. The fundamental theorem of calculus yields the relation

$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)dt$, where $\varphi(1) = F(y)$,$\varphi(0) = F(x)$.

According to the result of Exercise 1.11, $\varphi'(t) = \langle \nabla F(x_t), y - x \rangle$

Thus, $F(y) - F(x) = \int_0^1 \nabla F(x + t(y - x))(y - x)dt$

$F(y) = F(x) + \int_0^1 (\nabla F(x + t(y - x)) - \nabla F(x) + \nabla F(x))(y - x)dt$

$= F(x) + \nabla F(x)(y - x) + \int_0^1 (\nabla F(x + t(y - x)) - \nabla F(x))(y - x)dt$

Exercise 1.16

A C^1 -smooth mapping $F : U \to Y$ is L -Lipschitz continuous satisfies:

$|F(y) - F(x)| \leq L\|y - x\|$, for all y and x.

Take $y = x + \nabla F(x)$, $L\|\nabla F(x)\| = L\|y - x\| \geq |F(y) - F(x)|$

For a convex function F and two points x, y in its domain, we have $F(y) \geq l(x, y) = F(x) + \langle y - x, \nabla F(x)\rangle$, then $|F(y) - F(x)| \geq |\langle \nabla F(x), \nabla F(x)\rangle|$

So, $L\|\nabla F(x)\| \geq |\langle \nabla F(x), \nabla F(x)\rangle| = \|\nabla F(x)\|^2$, which gives $\|\nabla F(x)\| \leq L$.

If f is defined on a convex set and x is an interior point, the same argument but with $y = x + \eta\nabla f(x)$ for some small $\eta > 0$ gives us the same bound.

2 Chapter 2

Exercise 2.4*

We first assume $f(x)$ is convex and show its epigraph must be convex:
Let (x_1, y_1) and (x_2, y_2) be in the epigraph of $f(x)$.
Let $(\tilde{x}, \tilde{y}) := \lambda(x_1, y_1) + (1 - \lambda)(x_2, y_2)$, with $\lambda \in [0, 1]$. Then we have,

$$\begin{aligned}
\tilde{y} &= \lambda y_1 + (1 - \lambda)y_2 \\
&\geq \lambda f(x_1) + (1 - \lambda) f(x_2) \quad \text{since points are in epigraph} \\
&\geq f(\lambda x_1 + (1 - \lambda)x_2) \quad \text{since } f(x) \text{ is convex} \\
&= f(\tilde{x})
\end{aligned}$$

Hence, the point$(\tilde{x}, \tilde{y})$ is also in epigraph, and epigraph is convex.

Now we assume the epigraph is convex and show the function must be convex:
Let $x_1, x_2 \in C$, let $\lambda \in [0, 1]$. Then the points $(x_1 f(x_2))$ and $(x_2 f(x_2))$ are in the epigraph of $f(x)$. Since the epigraph is convex, the point $\lambda(x_1, f(x_1)) + (1 - \lambda)(x_2, f(x_2))$ is also in epigraph.

So $f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1)) + (1 - \lambda) f(x_2)$, f is convex.

Exercise 2.5*

If f is convex, according tol last exercise, the epi $f := \{(x, r) \in U \times \mathbf{R} : f(x) \leq r\}$ is convex. Pick arbitrary $(x_1, r_1), \ldots, (x_k, r_k) \in$ epi f, and any $\lambda_1, \ldots, \lambda_k \in [0, 1]$ with $\sum_i \lambda_i = 1$. By convexity of the epigraph, it follows that,

$\sum_i \lambda_i (x_i, r_i) = (\sum_i \lambda_i x_i, \sum_i \lambda_i r_i)$ also belongs to the epigraph. Therefore,

$\sum_i \lambda_i f(x_i) \geq f(\sum_i \lambda_i x_i)$.

Let $\Gamma = \text{epi}(f)$. Let $(x_1, r_1), \ldots, (x_n, r_n) \in \Gamma$. For any $\lambda_1, \ldots, \lambda_n \in [0,1]$ with $\sum_i \lambda_i = 1$, the point $(x, r) = \sum_i \lambda_i (x_i, r_i) = (\sum_i \lambda_i x_i, \sum_i \lambda_i r_i)$ has:

$r = \sum_i \lambda_i r_i \geq \sum_i \lambda_i f(x_i) \geq f(\sum_i \lambda_i x_i) = f(x)$

Hence $(x, r) \in \Gamma$, and Γ is convex. According to the results of last exercise, f is convex.

Exercise 2.6*

Consider an arbitrary set T and a family of convex functions $f_t : U \to (-\infty, +\infty]$ for $t \in T$.

Because a function $f : U \to (-\infty, +\infty]$ is convex if and only if the epigraph epi $f := \{(x, r) \in U \times \mathbf{R} : f(x) \leq r\}$ is a convex subset of $\mathbf{E} \times \mathbf{R}$.

epi $f = \cap_{t \in T} epi f_t$ is convex because all epi f_t is convex.

Hence the function $f(x) := \sup_{t \in T} f_t(x)$ is convex since epi f is convex.

Exercise 2.8*

Define the two sets

$\mathbf{R}^n_{++} := \{x \in \mathbf{R}^n : x_i > 0 \text{ for all } i = 1, \ldots, n\}$

$\mathbf{S}^n_{++} := \{X \in \mathbf{S}^n : X > 0\}$

Consider the two functions $f : \mathbf{R}^n_{++} \to \mathbf{R}$ and $F : \mathbf{S}^n_{++} \to \mathbf{R}$ given by $f(x) = -\sum_{i=1}^n \log x_i \quad$ and $\quad F(X) = -\ln \det(X)$

For $x \in \mathbf{R}^n_{++}, \nabla f(x) = -\sum_{i=1}^n \frac{1}{x_i}, \nabla^2 f(x) = \sum_{i=1}^n \frac{1}{x_i^2}$

For any $y > x$, $\nabla f(y) > \nabla f(x)$, satisfying the monotonicity of gradient, hence f is convex.

From exercise 1.10, the operator $\nabla^2 F(X) : \mathbf{S}^n \to \mathbf{S}^n$ is positive definite, since $\left\langle \nabla^2 F(X)[V], V \right\rangle = \left\| X^{-\frac{1}{2}} V X^{-\frac{1}{2}} \right\|_F^2$ for any $X > 0$ and $V \in \mathcal{S}^n$.

Hence, according to the Theo2.12 (d), F is convex.

Exercise 2.9

(1) $0 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{\beta}{2} \|x - y\|^2$

$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$ for all $x, y \in U$ is equivalent to f is convex

Given any β-smooth function $f : U \to \mathbf{R}$, for any points $x, y \in U$ the inequality $|f(y) - l(x; y)| \leq \frac{\beta}{2} \|y - x\|^2$ holds

Hence $f(y) - f(x) - \langle \nabla f(x), y - x \rangle \leq \frac{\beta}{2} \|y - x\|^2$

(2) $f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2\beta} \|\nabla f(x) - \nabla f(y)\|^2 \leq f(y)$

Suppose now 1 holds and define the function $\phi(y) = f(y) - \langle \nabla f(x), y \rangle$.

Since $0 \leq f(y) - f(x) - \langle \nabla f(x), y - x \rangle = \phi(y) - \phi(x)$, $\phi(y)$ is mono increasing function.

Since f is β-smooth, $\nabla f(y) - \nabla f(x) \leq \beta(y - x)$,

$x \leq y - \frac{1}{\beta}(\nabla f(y) - \nabla f(x)) = y - \frac{1}{\beta}\nabla\phi(y)$

So $\phi(x) = \min \phi \leq \phi\left(y - \frac{1}{\beta}\nabla\phi(y)\right) = f(y - \frac{1}{\beta}\nabla\phi(y)) - \langle \nabla f(x), y - \frac{1}{\beta}\nabla\phi(y)\rangle$

$= f(y - \frac{1}{\beta}\nabla\phi(y)) - f(y) + f(y) - \langle \nabla f(x), y\rangle + \langle \nabla f(x), \frac{1}{\beta}\nabla\phi(y)\rangle$

$= \phi(y) + f(y - \frac{1}{\beta}\nabla\phi(y)) - f(y) + \langle \nabla f(x), \frac{1}{\beta}\nabla\phi(y)\rangle$

$= \phi(y) + f(y - \frac{1}{\beta}\nabla\phi(y)) - f(y) - \langle \nabla f(y), -\frac{1}{\beta}\nabla\phi(y)\rangle + \langle \nabla f(y), -\frac{1}{\beta}\nabla\phi(y)\rangle + \langle \nabla f(x), \frac{1}{\beta}\nabla\phi(y)\rangle$

$\leq \phi(y) + \frac{1}{2\beta}\|\nabla\phi(y)\|^2 + \langle \nabla f(x) - \nabla f(y), \frac{1}{\beta}\nabla\phi(y)\rangle$, since(1)

$\leq \phi(y) + \frac{1}{2\beta}\|\nabla\phi(y)\|^2 + \langle \beta(x - y), \frac{1}{\beta}\nabla\phi(y)\rangle$, since the β-smooth

$\leq \phi(y) + \frac{1}{2\beta}\|\nabla\phi(y)\|^2 + \langle -\frac{1}{\beta}\nabla\phi(y), \frac{1}{\beta}\nabla\phi(y)\rangle$, since the definition of $\phi(x)$

$\leq \phi(y) - \frac{1}{2\beta}\|\nabla\phi(y)\|^2$

Hence $f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2\beta}\|\nabla f(x) - \nabla f(y)\|^2 \leq f(y)$

Obviously, f is convex.

(3) $\frac{1}{\beta}\|\nabla f(x) - \nabla f(y)\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y\rangle$

according to (2),$f(x) + \langle \nabla f(x), y - x\rangle + \frac{1}{2\beta}\|\nabla f(x) - \nabla f(y)\|^2 \leq f(y)$

reverse x and y, $f(y) + \langle \nabla f(y), x - y\rangle + \frac{1}{2\beta}\|\nabla f(y) - \nabla f(x)\|^2 \leq f(x)$

add this two inequality, get $\frac{1}{\beta}\|\nabla f(x) - \nabla f(y)\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y\rangle$

Next applying Cauchy-Schwartz to (3)

$\frac{1}{\beta}\|\nabla f(x) - \nabla f(y)\|^2 \leq \langle \nabla f(x) - \nabla f(y), x - y\rangle \leq \|\nabla f(x) - \nabla f(y)\|\|x - y\|$

Hence f is β-smooth, $\|\nabla f(x) - \nabla f(y)\| \leq \beta\|x - y\|$

Also f is convex, since $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq 0$

(4) $0 \leq \langle \nabla f(x) - \nabla f(y), x - y\rangle \leq \beta\|x - y\|^2$

From(3), $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq 0$

From(1), $f(y) - f(x) - \langle \nabla f(x), y - x\rangle \leq \frac{\beta}{2}\|x - y\|^2$

Reverse x and y, $f(x) - f(y) - \langle \nabla f(y), x - y\rangle \leq \frac{\beta}{2}\|y - x\|^2$

Add this two inequalities, get (4).

To deduce (1) from (4), let $\phi(x) = -f(x) + \frac{\beta}{2}\|x\|^2$

The gradient of $\phi(x)$ is, $\nabla\phi(x) = -\nabla f(x) + \beta x$

Similarly, $\nabla\phi(y) = -\nabla f(y) + \beta y$

$(\nabla\phi(x) - \nabla\phi(y))(x - y) = \beta\|x - y\|^2 - \langle \nabla f(x) - \nabla f(y), x - y\rangle \geq 0$, since (4).

Thus, $\nabla\phi(x)$ is monotone and therefore $\phi(x)$ is convex.

From the gradient inequality of $\phi(x)$, $\phi(y) \geq \phi(x) + \langle \nabla\phi(x), y - x\rangle$

since $\nabla\phi(x) = -\nabla f(x) + \beta x$, $\nabla\phi(y) = -\nabla f(y) + \beta y$, we can get $f(y) - f(x) - \langle \nabla f(x), y - x\rangle \leq \frac{\beta}{2}\|x-y\|^2$

Exercise 2.11

Consider a C^1 -smooth convex function $f : \mathbf{E} \to \mathbf{R}$. Fix a linear subspace $\mathcal{L} \subset \mathbf{E}$ and $g : \mathbf{L} \to \mathbf{R}, f_{\mathcal{L}} = f$.

First, assume $x \in \mathcal{L}$ minimizes the restriction $f_{\mathcal{L}} : \mathcal{L} \to \mathbf{R}$.

Based on the Corollary 2.10, we know that $\nabla f_{\mathcal{L}}(x) = 0$.

We have $\nabla f(x) = (\nabla f_{\mathcal{L}}(x), \nabla f_{\mathbf{E}-\mathcal{L}}(x))$. Let $x = (x_{\mathcal{L}}, 0) \in \mathcal{L}$, we have $\nabla f(x)x = 0$. Hence the gradient $\nabla f(x)$ is orthogonal to $\mathcal{L}$.

Second, assume the gradient $\nabla f(x)$ is orthogonal to $\mathcal{L}$. From the reverse of the above statement, we can derive $\nabla f_{\mathcal{L}}(x) = 0$, so x is a critical point of $f_{\mathcal{L}}$. According to Corollary 2.10, $x \in \mathcal{L}$ minimizes the restriction $f_{\mathcal{L}} : \mathcal{L} \to \mathbf{R}$

Exercise 2.13

Suppose $g(x) = f(x) - \frac{\alpha}{2}\|x\|^2$ is convex. The gradient of $g(x)$ is $\nabla g(x) = \nabla f(x) - \alpha x$

According to the gradient inequality, $g(y) \geq g(x) + \langle \nabla g(x), y - x\rangle$

$f(y) - \frac{\alpha}{2}\|y\|^2 \geq f(x) - \frac{\alpha}{2}\|x\|^2 + \langle \nabla f(x) - \alpha x, y - x\rangle$

Hence $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{2}\|y - x\|^2$

Suppose $f : U \to \mathbf{R}$ is α -strongly convex, then

$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\alpha}{2}\|y - x\|^2$ holds for all $x, y \in U$

Revise x and y, we get $f(x) \geq f(y) + \langle \nabla f(y), x - y\rangle + \frac{\alpha}{2}\|x - y\|^2$

Add this two inequalities, we get $\langle \nabla f(x) - \nabla f(y), y - x\rangle + \alpha\|y - x\|^2 \leq 0$

To prove the $g(x)$ is convex, we could prove $\langle \nabla g(y) - \nabla g(x), y - x\rangle \geq 0$ for all $x, y \in U$.

$\langle \nabla g(y) - \nabla g(x), y - x\rangle = \langle \nabla f(y) - \alpha y - (\nabla f(x) - \alpha x), y - x\rangle = \langle \nabla f(y) - \nabla f(x), y - x\rangle - \alpha\|y - x\|^2 \geq 0$, since the inequality we deduced above.

Hence $\nabla g(x)$ is monotone and therefore $g(x)$ is convex.

Exercise 2.18*

Consider a differentiable convex function, $f : \mathbf{E} \to \mathbf{R}$ and let x^* be any of its minimizers.

Consider the gradient descent iterates $x_{k+1} = x_k - \gamma_k \nabla f(x_k)$, for some sequence $\alpha_k \geq 0$

(1) $\|x_{k+1} - x^*\|^2 = \|(x_{k+1} - x_k) + (x_k - x^*)\|^2$

$\frac{1}{2}\|x_{k+1} - x^*\|^2 = \frac{1}{2}\|x_k - x^*\|^2 + \|x_{k+1} - x_k\|\|x_k - x^*\| + \frac{\gamma_k^2}{2}\|\nabla f(x_k)\|^2$

$= \frac{1}{2}\left\|x_k - x^*\right\|^2 + \gamma_k \|\nabla f(x_k)\| \|x_k - x^*\| + \frac{\gamma_k^2}{2}\left\|\nabla f\left(x_k\right)\right\|^2$
$= \frac{1}{2}\left\|x_k - x^*\right\|^2 + \gamma_k \langle \nabla f(x_k), x_k - x^* \rangle + \frac{\gamma_k^2}{2}\left\|\nabla f\left(x_k\right)\right\|^2$
$= \frac{1}{2}\left\|x_k - x^*\right\|^2 + \gamma_k (\langle \nabla f(x_k), x_k - x^* \rangle + f(x_k) - f(x_k)) + \frac{\gamma_k^2}{2}\|\nabla f(x_k\|^2$
$\leq \frac{1}{2}\left\|x_k - x^*\right\|^2 + \gamma_k (f(x^*) - f(x_k)) + \frac{\gamma_k^2}{2}\|\nabla f(x_k\|^2$, since the inequality of gradient

Hence, $\frac{1}{2}\left\|x_{k+1} - x^*\right\|^2 \leq \frac{1}{2}\left\|x_k - x^*\right\|^2 - \gamma_k\left(f\left(x_k\right) - f\left(x^*\right)\right) + \frac{\gamma_k^2}{2}\left\|\nabla f\left(x_k\right)\right\|^2$

(2) $\frac{1}{2}\left\|x_{k+1} - x^*\right\|^2 \leq \frac{1}{2}\left\|x_k - x^*\right\|^2 - \gamma_k\left(f\left(x_k\right) - f\left(x^*\right)\right) + \frac{\gamma_k^2}{2}\left\|\nabla f\left(x_k\right)\right\|^2$

To minimize the right hand of this inequality, let $F(\gamma_k) = \frac{1}{2}\left\|x_k - x^*\right\|^2 - \gamma_k\left(f\left(x_k\right) - f\left(x^*\right)\right) + \frac{\gamma_k^2}{2}\left\|\nabla f\left(x_k\right)\right\|^2$

Take $\nabla F(\gamma_k) = 0$, get $-(f(x_k) - f(x^*)) + \gamma_k \|\nabla f\left(x_k\right)\|^2 = 0$

Hence, $\gamma_k = \frac{f(x_k) - f^*}{\|\nabla f(x_k)\|^2}$

Substitute γ_k to the initial inequality, thereby yielding the guarantee
$\left\|x_{k+1} - x^*\right\|^2 \leq \left\|x_k - x^*\right\|^2 - \left(\frac{f(x_k) - f^*}{\|\nabla f(x_k)\|}\right)^2$

(3) since f is β -smooth, we have

$$\begin{aligned}
f\left(x_{k+1}\right) &\leq f\left(x_k\right) - \left\langle \nabla f\left(x_k\right), \alpha_k \nabla f\left(x_k\right)\right\rangle + \frac{\beta}{2}\left\|\alpha_k \nabla f\left(x_k\right)\right\|^2 \\
&= f\left(x_k\right) - \left(f\left(x_k\right) - f^*\right) + \frac{\beta}{2} \frac{\left(f\left(x_k\right) - f^*\right)^2}{\left\|\nabla f\left(x_k\right)\right\|^2} \\
&= f^* + \frac{\beta}{2} \frac{\left(f\left(x_k\right) - f^*\right)^2}{\left\|\nabla f\left(x_k\right)\right\|^2} \\
&\leq f^* + \frac{\beta}{2}\left(\left\|x_k - x^*\right\|^2 - \left\|x_{k+1} - x^*\right\|^2\right)
\end{aligned}$$

Sum up from 0 to $k-1$, we have
$\sum_{i=0}^{k-1} f\left(x_{i+1}\right) - k f^* \leq \frac{\beta}{2}\left(\left\|x_0 - x^*\right\|^2 - \left\|x_k - x^*\right\|^2\right)$

Setting $\lim_{k \to \infty} x_k = x^*$ and considering f is convex, we have
$f\left(\frac{1}{k} \sum_{i=0}^{k-1} x_i\right) - f^* \leq \frac{1}{k} \sum_{i=0}^{k-1} f\left(x_{i+1}\right) - f^* \leq \frac{\beta\left\|x_0 - x^*\right\|^2}{2k}$

If f is in addition α -strongly convex, we have

$$\begin{aligned}
\left\|x_{k+1} - x^*\right\|^2 &= \left\| x_k - x^* - \alpha_k \nabla f\left(x_k\right)\right\|^2 \\
&= \left\|x_k - x^*\right\|^2 + 2\alpha_k\left\langle \nabla f\left(x_k\right), x^* - x_k\right\rangle + \alpha_k^2\left\|\nabla f\left(x_k\right)\right\|^2 \\
&\leq \left\|x_k - x^*\right\|^2 + 2\alpha_k\left(f^* - f\left(x_k\right) - \frac{\alpha}{2}\left\|x_k - x^*\right\|^2\right) + \alpha_k^2\left\|\nabla f\left(x_k\right)\right\|^2 \\
&= \left(1 - \alpha\alpha_k\right)\left\|x_k - x^*\right\|^2 + \alpha_k\left(2f^* - 2f\left(x_k\right) + \alpha_k\left\|\nabla f\left(x_k\right)\right\|^2\right)
\end{aligned}$$

$$= (1-\alpha\alpha_k)\|x_k - x^*\|^2 + \left(\frac{f(x_k)-f^*}{\|\nabla f(x_k)\|}\right)^2$$
$$\leq (2-\alpha\alpha_k)\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2$$

So we have

$$\|x_{k+1} - x^*\|^2 \leq \left(1 - \frac{\alpha}{2}\alpha_k\right)\|x_k - x^*\|^2$$
$$= \left(1 - \frac{\alpha}{2}\frac{f(x_k)-f^*}{\|\nabla f(x_k)\|^2}\right)\|x_k - x^*\|^2$$

Based on strong convex, we know that $f(x_k) - f^* \geq +\frac{\alpha}{2}\|x_k - x^*\|^2$

Based on β -smooth, we know that $\|\nabla f(x) - \nabla f(y)\|_2 \leq \beta\|x-y\|_2$

That is, $\|\nabla f(x_k)\|^2 \leq \beta^2 \|x_k - x^*\|^2$, since $\nabla f(x^*) = 0$

Put it in the above equation, wo have $\|x_{k+1} - x^*\|^2 \leq \left(1 - \frac{\alpha^2}{\beta^2}\right)\|x_k - x^*\|^2$

Exercise 2.22 (1)

$f_k(z_1, z_2, \ldots, z_n) = \frac{\beta}{4}\left(\frac{1}{2}\left(z_1^2 + \sum_{i=1}^{k-1}(z_i - z_{i+1})^2 + z_k^2\right) - z_1\right)$

Take the derivative

$$\frac{\partial f_k}{\partial z_1} = \frac{\beta}{4}\left(\frac{1}{2}(2z_1 + 2(z_1 - z_2)) - 1\right)$$
$$=\frac{\beta}{4}(z_1 - z_2 - 1)$$
$$\frac{\partial f_k}{\partial z_{i, 2\leq i\leq k-1}} = \frac{\beta}{4}\left(\frac{1}{2}(2(z_i - z_{i+1}) - 2(z_{i-1} - z_i))\right))$$
$$= \frac{\beta}{4}(2z_i - z_{i-1} - z_{i+1})$$

For $k \leq i$

$$\frac{\partial f_k}{\partial z_k} = \frac{\beta}{4}\left(\frac{1}{2}(-2(z_{k-1} - z_k) + 2z_k)\right)$$
$$= \frac{\beta}{4}(2z_k - z_{k-1})$$

$\frac{\partial f_k}{\partial z_{i,i>k}} = 0$

Let $x_1 = z_1 = 1 - \frac{1}{k+1}$, we can deduct that

$\overline{x}_{k,i} = \begin{cases} 1 - \frac{i}{k+1}, & \text{if } i = 1, \ldots, k \\ 0 & \text{if } i = k+1, \ldots, n \end{cases}$

Substitute the $\overline{x}_k$ value to the equation, we have

$f_k(\overline{x}_{k,1}, \overline{x}_{k,2}, \ldots, \overline{x}_{k,n})$

$= \frac{\beta}{4}\left(\frac{1}{2}\left(\overline{x}_{k,1}^2 + \sum_{i=1}^{k-1}(\overline{x}_{k,i} - \overline{x}_{k,i+1})^2 + \overline{x}_{k,k}^2\right) - \overline{x}_{k,1}\right)^2 + \left(1 - \frac{k}{k+1}\right)^2) - \left(1 - \frac{1}{k+1}\right))$

$= \frac{\beta}{8}\left(-1 + \frac{1}{k+1}\right)$

(2)

$$\|\overline{x}_k\|^2 = \sum_{i=1}^{k}\left(1 - \frac{i}{k+1}\right)^2$$
$$= \frac{1}{(k+1)^2}\sum_{i=1}^{k}\left((k+1)^2 - 2i(k+1) + i^2\right)$$
$$= \frac{1}{(k+1)^2}\left(k(k+1)^2 - 2(k+1)\frac{k(k+1)}{2} + \sum_{i=1}^{k}\left(i^2\right)\right)$$
$$\leq \frac{1}{3}(k+1)$$

(3)

Let $x = (x_1, x_2, \ldots, x_i, 0, 0 \ldots 0) \in \mathbf{R}^i \times \{0\}^{n-i}$. Substitute to the function, we have:

$f_j\left(x_1, x_2, \ldots, x_n\right)$

$= \frac{\beta}{4}\left(\frac{1}{2}\left(x_1^2 + \sum_{k=1}^{j-1}\left(x_k - x_{k+1}\right)^2 + x_j^2\right) - x_1\right)$

$= \frac{\beta}{4}\left(\frac{1}{2}\left(x_1^2 + \sum_{k=1}^{i-1}\left(x_k - x_{k+1}\right)^2 + 0 + x_i^2\right) - x_1\right)$

$= \frac{\beta}{4}\left(\frac{1}{2}\left(x_1^2 + \sum_{k=1}^{i-1}\left(x_k - x_{k+1}\right)^2 + x_i^2\right) - x_1\right)$

$= f_i\left(x_1, x_2, \ldots, x_n\right)$

Based on the gradient result at question (1), we know that the gradient $\nabla f_k(x)$ lies in $\mathbf{R}^i \times \{0\}^{n-(i)} \in \mathbf{R}^{i+1} \times \{0\}^{n-(i+1)}$

Exercise 2.23*

For any point $x \in \mathbf{R}^n$,
The equality $f(x) - f^* = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ -$(\frac{1}{2}\langle Ax^*, x^*\rangle - \langle b, x^*\rangle)$
$= \frac{1}{2}(\langle Ax, x\rangle - \langle Ax, x^*\rangle + \langle Ax, x^*\rangle - \langle Ax^*, x^*\rangle)$ $-\langle b, x - x^*\rangle$
$= \frac{1}{2}(\langle Ax, x - x^*\rangle + \langle Ax^*, x - x^*\rangle)$ $-\langle b, x - x^*\rangle$
$= \frac{1}{2}(\langle A(x + x^*), x - x^*\rangle)$ $-\langle b, x - x^*\rangle$
$= \frac{1}{2}(\langle A(x + x^*) - 2b, x - x^*\rangle)$
$= \frac{1}{2}(\langle A(x - x^*), x - x^*\rangle)$, since $Ax^* = b$
$= \frac{1}{2}\left\|x - x^*\right\|_A^2$

Exercise 2.24*

$t_k = \operatorname{argmin}_t f\left(x_k + tv_k\right)$.

From optimal conditions, let $\nabla f(x_k + tv_k) = 0$

Then, $A(x_k + tv_k) - b)v_k = 0$, we get $t = \frac{bv_k - Ax_k v_k}{Av_k^2} = \frac{\langle r_k, v_k \rangle}{\|v_k\|_A^2}$

Exercise 2.30

Assuming $a_k \le \frac{2}{k+2}$, $k > 0$, then if we can prove $a_{k+1} \le \frac{2}{k+2}$, it will show that there is a bound $a_k \le \frac{2}{k+2}$, for each $k \ge 0$

Since $a_{k+1} = \frac{\sqrt{a_k^4 + 4a_k^2} - a_k^2}{2} \le \frac{a_k^2 + 2a_k - a_k^2}{2} \le a_k \le \frac{2}{k+2}$

Not sure how to use tighter bound to prove $a_{k+1} \le \frac{2}{k+3}$

Exercise 2.42*

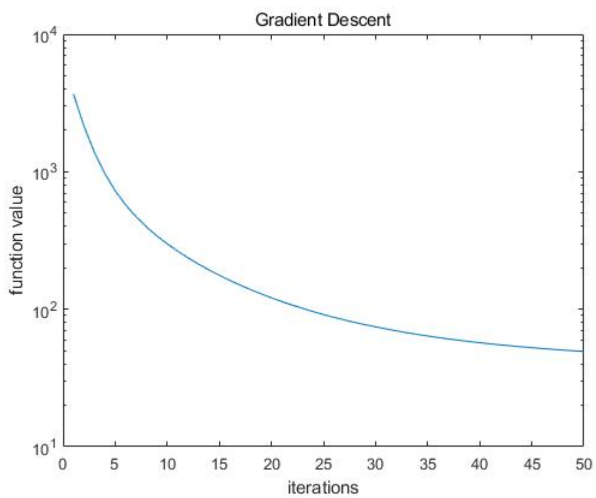

Figure 1:

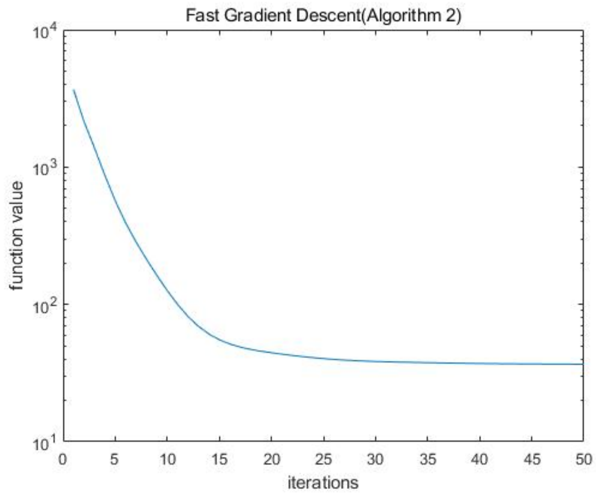

Figure 2:

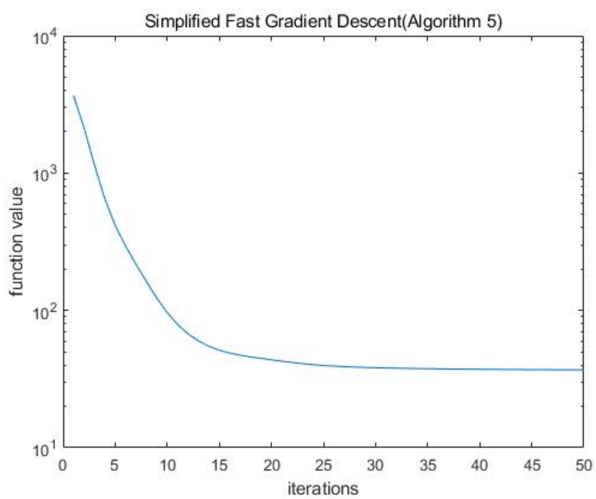

Figure 3:

Exercise 2.43

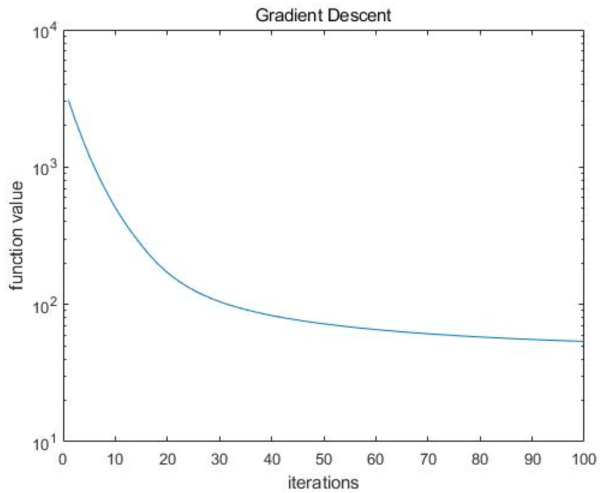

Figure 4:

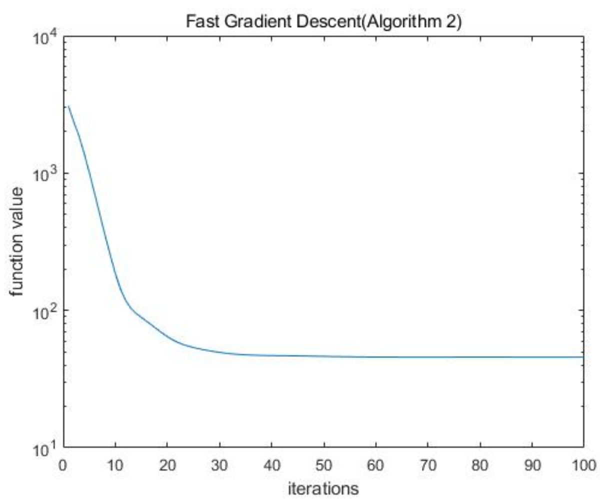

Figure 5:

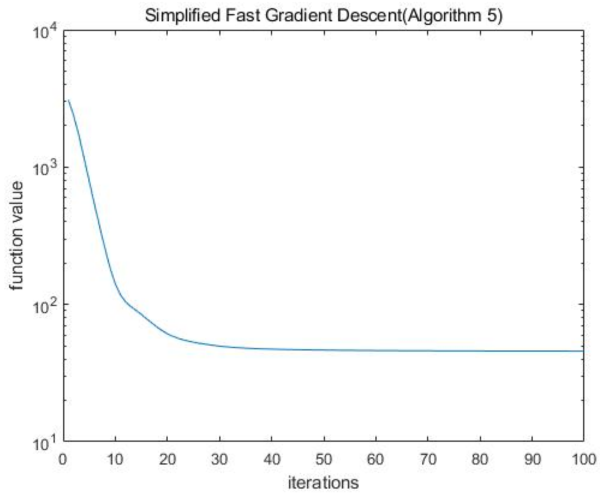

Figure 6:

Exercise 2.44*

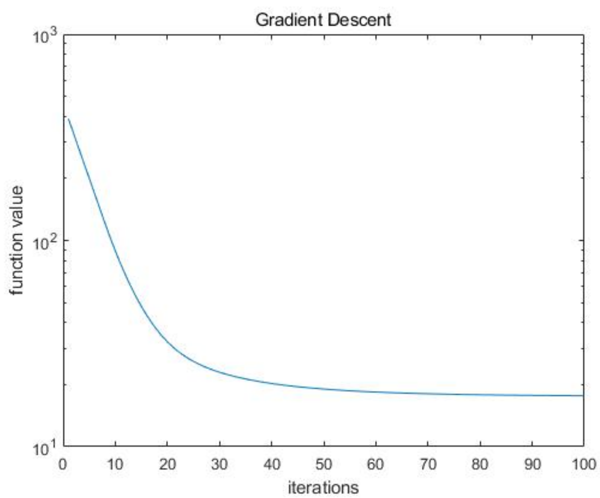

Figure 7:

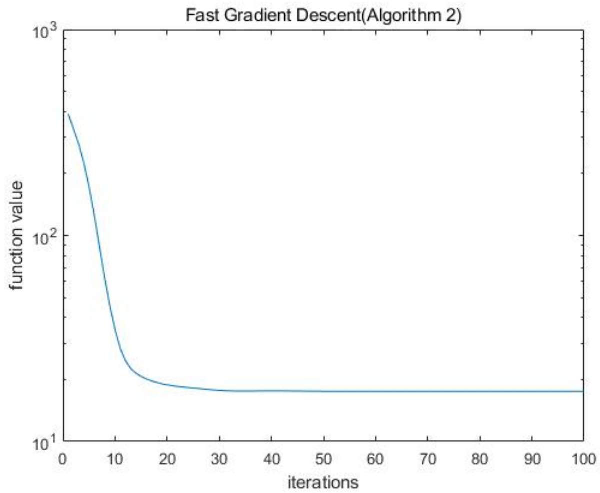

Figure 8:

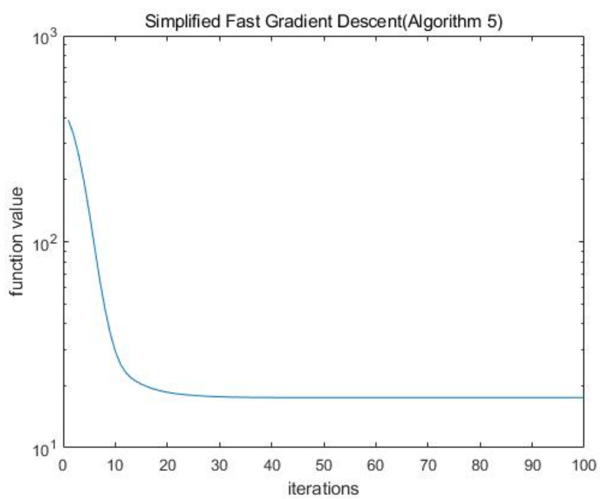

Figure 9: