ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ
TECHNICAL UNIVERSITY OF CRETE

Course "Optimization" Report

Exercise 1

Georgios Frangias 2018030086
`gfrangias@tuc.gr`

Course Professor:
Athanasios Liavas

Technical University of Crete

School of Electrical & Computer Engineering

October 2022

1.

(a)

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

$$f''(x) = \frac{2}{(1+x)^3}$$

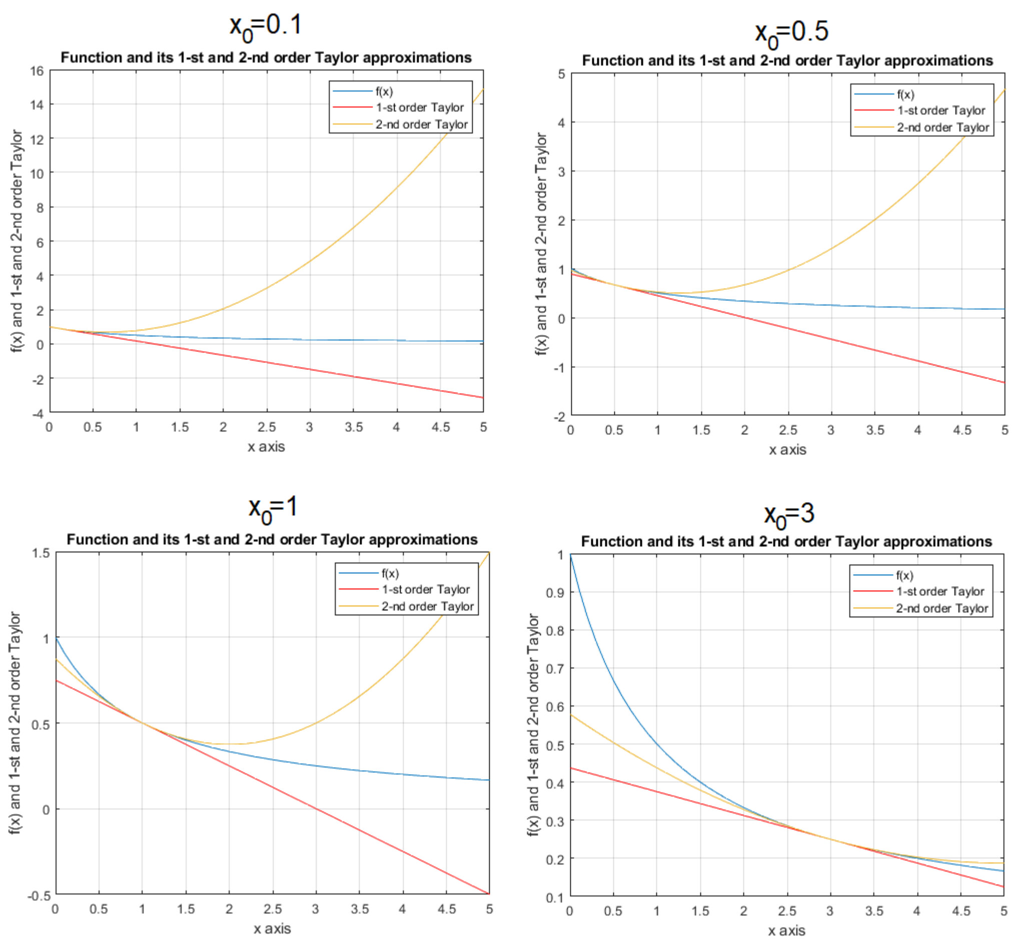
$$f_{(1)}(x) = f(x_0) + f'(x_0)(x - x_0) = \frac{1}{1+x_0} - \frac{1}{(1+x_0)^2}(x - x_0) =$$

$$= \frac{1}{1+x_0}\left(1 + \frac{x_0 - x}{1+x_0}\right)$$

$$f_{(2)}(x) = f_{(1)}(x) + \frac{1}{2}f''(x_0)(x - x_0)^2 = \frac{1}{1+x_0}\left(1 + \frac{x_0 - x}{1+x_0}\right) + \frac{1}{2}\frac{2}{(1+x_0)^3}(x - x_0)^2 =$$

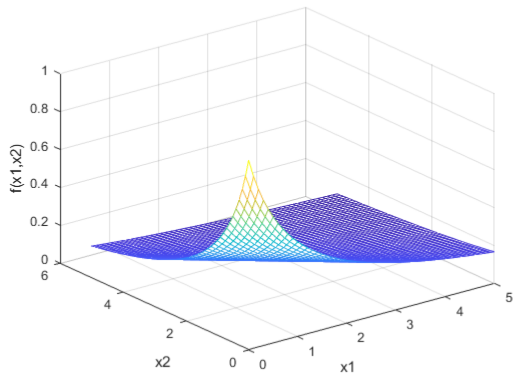
$$= \frac{1}{1+x_0}\left(1 + \frac{x_0 - x}{1+x_0} + \left(\frac{x - x_0}{1+x_0}\right)^2\right)$$

(b)

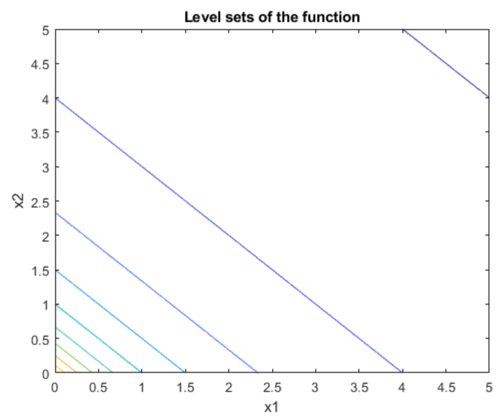


2.

(a)



(b)



It can be observed that as the values of x_1 and x_2 decrease, the value of the function increases. Actually, the increase is exponential as the sub-level lines become denser coming closer to the origin.

(c)

$$\nabla f(x_0) = \begin{bmatrix} \frac{\partial f(x_0)}{\partial x_1} \\ \frac{\partial f(x_0)}{\partial x_2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{(1+x_1+x_2)^2} \\ -\frac{1}{(1+x_1+x_2)^2} \end{bmatrix}$$

$$f_{(1)}(x_0) = f(x_0) + \nabla f(x_0)^T (x - x_0)$$

$$\Rightarrow f_{(1)}(x_{0,1}, x_{0,2}) = \frac{1}{1 + x_{0,1} + x_{0,2}} + \begin{bmatrix} -\frac{1}{(1+x_{0,1}+x_{0,2})^2} & -\frac{1}{(1+x_{0,1}+x_{0,2})^2} \end{bmatrix} \begin{bmatrix} x_1 - x_{0,1} \\ x_2 - x_{0,2} \end{bmatrix}$$

$$\Rightarrow f_{(1)}(x_{0,1}, x_{0,2}) = \frac{1}{1 + x_{0,1} + x_{0,2}} + \frac{x_{0,1} - x_1}{(1 + x_{0,1} + x_{0,2})^2} + \frac{x_{0,2} - x_2}{(1 + x_{0,1} + x_{0,2})^2}$$

$$\Rightarrow f_{(1)}(x_{0,1}, x_{0,2}) = \frac{1}{1 + x_{0,1} + x_{0,2}} \left(1 + \frac{x_{0,1} + x_{0,2} - x_1 - x_2}{1 + x_{0,1} + x_{0,2}} \right)$$

$$D^2 f(x_0) = \begin{bmatrix} \frac{\partial^2 f(x_0)}{\partial x_1^2} & \frac{\partial^2 f(x_0)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x_0)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x_0)}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} \frac{2}{(1+x_{0,1}+x_{0,2})^3} & \frac{2}{(1+x_{0,1}+x_{0,2})^3} \\ \frac{2}{(1+x_{0,1}+x_{0,2})^3} & \frac{2}{(1+x_{0,1}+x_{0,2})^3} \end{bmatrix}$$

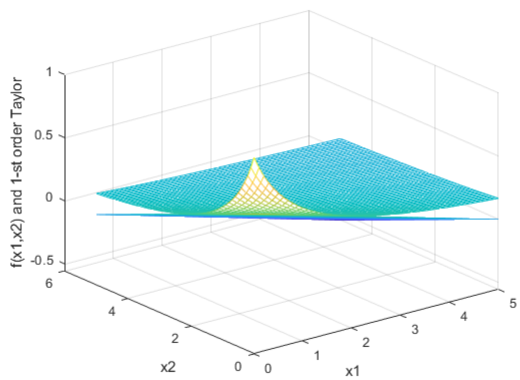
$$f_{(2)}(x_0) = f_{(1)}(x_0) + \frac{1}{2}(x - x_0)^T D^2 f(x_0)(x - x_0) =$$

$$= \frac{1}{1+x_{0,1}+x_{0,2}} \left(1 + \frac{x_{0,1}+x_{0,2}-x_1-x_2}{1+x_{0,1}+x_{0,2}}\right) + \frac{1}{2} \begin{bmatrix} x_1 - x_{0,1} & x_2 - x_{0,2} \end{bmatrix} \begin{bmatrix} \frac{2}{(1+x_{0,1}+x_{0,2})^3} & \frac{2}{(1+x_{0,1}+x_{0,2})^3} \\ \frac{2}{(1+x_{0,1}+x_{0,2})^3} & \frac{2}{(1+x_{0,1}+x_{0,2})^3} \end{bmatrix} \begin{bmatrix} x_1 - x_{0,1} \\ x_2 - x_{0,2} \end{bmatrix} =$$

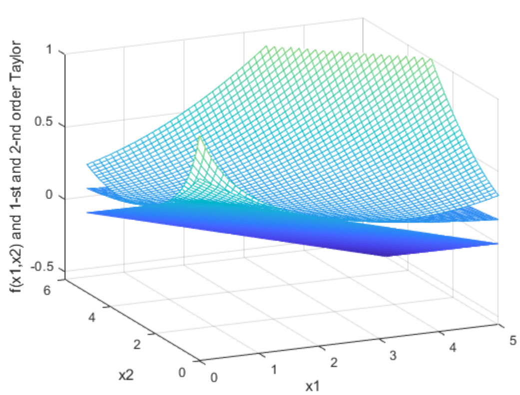
$$= \frac{1}{1+x_{0,1}+x_{0,2}} \left(1 + \frac{x_{0,1}+x_{0,2}-x_1-x_2}{1+x_{0,1}+x_{0,2}}\right) + \begin{bmatrix} \frac{x_1+x_2-x_{0,1}-x_{0,2}}{(1+x_{0,1}+x_{0,2})^3} & \frac{x_1+x_2-x_{0,1}-x_{0,2}}{(1+x_{0,1}+x_{0,2})^3} \end{bmatrix} \begin{bmatrix} x_1 - x_{0,1} \\ x_2 - x_{0,2} \end{bmatrix} =$$

$$= \frac{1}{1+x_{0,1}+x_{0,2}} \left(1 + \frac{x_{0,1}+x_{0,2}-x_1-x_2}{1+x_{0,1}+x_{0,2}} + \frac{x_1+x_2-x_{0,1}-x_{0,2}}{(1+x_{0,1}+x_{0,2})^2}\right)$$

(d)



(e)



3. Let $\mathbb{S}_{a,b} = \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{a}^T\mathbf{x} \leq b\}$

(a) Prove that $\mathbb{S}_{a,b}$ is convex.

Let $\theta \in [0,1]$ and $\mathbf{x_1}, \mathbf{x_2} \in \mathbb{S}_{a,b}$

$$\mathbf{a}^T(\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}) = \theta\mathbf{a}^T\mathbf{x_1} + (1-\theta)\mathbf{a}^T\mathbf{x_2} \overset{\theta \in [0,1]}{\leq} \theta b + (1-\theta)b = b$$

So, $\mathbf{a}^T(\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}) \leq b$ and therefore $\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \in \mathbb{S}_{a,b}$ and by definition $\mathbb{S}_{a,b}$ is convex.

(b) Prove that $\mathbb{S}_{a,b}$ is not affine.

For $\mathbb{S}_{a,b}$ to be affine it needs to be applied that:

$$\forall \mathbf{x_1}, \mathbf{x_2} \in \mathbb{S}_{a,b} \;\&\; \theta \in \mathbb{R} \Rightarrow \theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \in \mathbb{S}_{a,b}$$

Trying to prove that $\mathbb{S}_{a,b}$ is not affine using a counterexample.

For example, let $\theta = 2 \in \mathbb{R}$ and $b = 1$. Then choose the points $\mathbf{x_1}$, $\mathbf{x_2} \in \mathbb{S}_{a,b}$ such that $\begin{cases} \mathbf{a}^T\mathbf{x_1} = 1 = b \\ \mathbf{a}^T\mathbf{x_2} = -2 < b \end{cases}$

$$\Rightarrow \begin{cases} 2\mathbf{a}^T\mathbf{x_1} = 2 \\ -\mathbf{a}^T\mathbf{x_2} = 2 \end{cases} \Rightarrow 2\mathbf{a}^T\mathbf{x_1} - \mathbf{a}^T\mathbf{x_2} = 4 \Rightarrow \mathbf{a}^T(2\mathbf{x_1} + (1-2)\mathbf{x_2}) = 4 > b$$

So, for this example $\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2} \notin \mathbb{S}_{a,b}$ and hence $\mathbb{S}_{a,b}$ is not affined.

4. Find the point $\mathbf{x}_*$ that is collinear with a and lies on the hyperplane $\mathbb{H}_{a,b} := \{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{a}^T\mathbf{x} = b\}$

The points $\mathbf{x}_*$ and $\mathbf{a}$ are collinear and this means that a constant $c \in \mathbb{R}$ exists such that $\mathbf{x}_* = c\mathbf{a}$
Thus, because $\mathbf{x}_*$ lies on $\mathbb{H}_{a,b}$:

$$\mathbf{a}^T\mathbf{x}_* = b \Rightarrow \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} x_{*1} & x_{*2} & \dots & x_{*n} \end{bmatrix} = \frac{1}{c}x_{*1}^2 + \frac{1}{c}x_{*2}^2 + \dots + \frac{1}{c}x_{*n}^2 = b \Rightarrow x_{*1}^2 + x_{*2}^2 + \dots + x_{*n}^2 = cb$$

In this way, it is proven that the point $\mathbf{x}_*$ has two properties:

1. $x_{*1} = ca_1,\ x_{*2} = ca_2,\ \dots,\ x_{*n} = ca_n$
2. $x_{*1}^2 + x_{*2}^2 + \dots + x_{*n}^2 = cb$

From 1. and 2. it can be concluded that:

$$\mathbf{x}_* = \frac{b}{||\mathbf{x}_*||_2^2}\mathbf{a}$$

5. Check whether the following functions are convex or not.

(a) $f : \mathbb{R}_+ \to \mathbb{R}$, with $f(x) = \frac{1}{1+x}$

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

$$f''(x) = \frac{2}{(1+x)^3}$$

The second derivative of f is positive $\forall x \in \mathbb{R}_+$. So based on the second derivative rule f is a convex function.

(b) $f : \mathbb{R}^2_+ \to \mathbb{R}$, with $f(x_1, x_2) = \frac{1}{1+x_1+x_2}$

From exercise 2.(c) it is known that:

$$D^2 f(x) = \begin{bmatrix} \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \\ \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \end{bmatrix}$$

Let a 2x1 non-zero column vector $\begin{bmatrix} a \\ b \end{bmatrix}$. If $\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \\ \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$ is positive for any value of a, b, then the matrix is positive-definite.

$$\begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \\ \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \ldots = \frac{2(a+b)^2}{1+x_1+x_2} > 0 \quad \forall \mathbf{x} \in \mathbb{R}^2_+$$

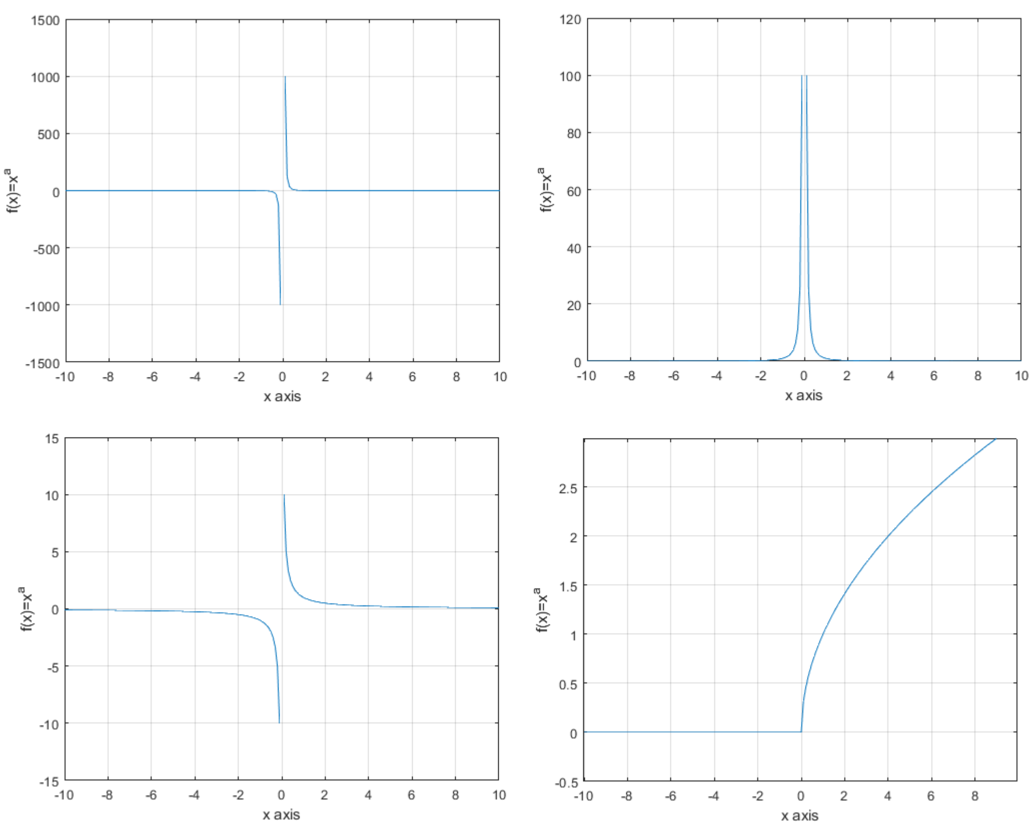
The fact that $D^2 f(x)$ is found to be positive-definite indicates that f is a convex function according to "Convex Optimization" by S.Boyd and L.Vandenberghe (page 71).

(c) $f : \mathbb{R}_{++} \to \mathbb{R}$, with $f(x) = x^a$, for (to get a better feeling, plot function x^a, for various values of a)

i. $a \geq 1$ **and** $a \leq 0$

$$f''(x) = a(a-1)x^{a-2}$$

For $a \leq 0 \Rightarrow a - 1 \leq -1$ and so $a(a-1) \geq 0$. [1]
For $a \geq 1 \Rightarrow a - 1 \geq 0$ and so $a(a-1) \geq 0$. [2]



Above, some typical graphs of x^{a-2} are depicted. Top-right $a = -1$, top-left $a = 0$, bottom-left $a = 1$ and bottom-right $a = 2.5$. It is obvious that for odd values of a the graph in $\mathbb{R}_{++}$ is mirrored based on the line $y = -x$ in $\mathbb{R}_{--}$, and for even values it is mirrored based on the y axis. For a rational the graph has the typical shape of $y = \sqrt{x}$. In any case in $\mathbb{R}_{++}$ the graph is strictly positive. [3]
So from [1], [2], [3] we conclude that f'' is non-negative in $\mathbb{R}_{++}$ and so f is convex.

ii. $0 \leq a \leq 1$

For $0 \leq a \leq 1 \Rightarrow -1 \leq a - 1 \leq 0$ and so $-1 \leq a(a-1) \leq 0$.
Similarly to i. a^{x-2} is positive in $\mathbb{R}_{++}$ and thus f'' is non-positive and f is not convex.

(d) $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$, with $f_1(x) = ||x||_2$ and $f_2(x) = ||x||_2^2$ (plot the functions for $\mathbf{n = 2}$).

Let $\theta \in [0,1]$ and $\mathbf{x_1}, \mathbf{x_2} \in \mathbb{R}^n$. Then

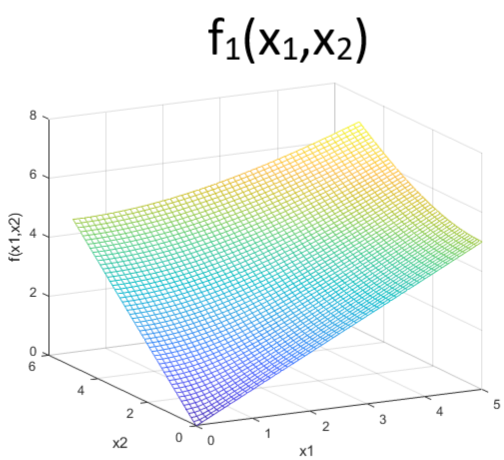
$$f_1(\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}) = ||\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}||_2 \leq ||\theta\mathbf{x_1}||_2 + ||(1-\theta)\mathbf{x_2}||_2 = \theta f(\mathbf{x_1}) + (1-\theta)f(\mathbf{x_2})$$

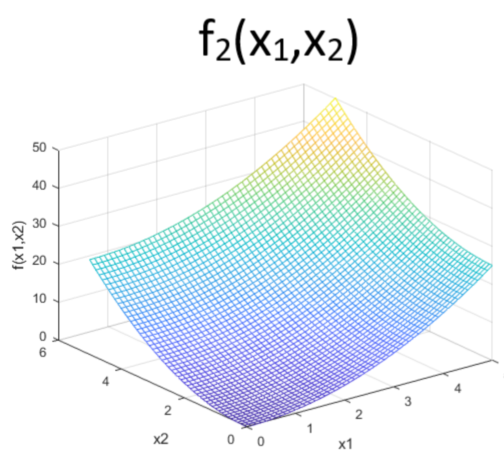
And so f_1 is convex.

$$D^2 f_2(x) = \begin{bmatrix} \frac{\partial^2 f_2(x)}{\partial x_1^2} & \frac{\partial^2 f_2(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f_2(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f_2(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f_2(x)}{\partial x_2^2} & \dots & \frac{\partial^2 f_2(x)}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f_2(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f_2(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f_2(x)}{\partial x_n^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 & \dots & 0 \\ 0 & 2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & & 2 \end{bmatrix} = 2\mathbf{I_n}$$

Let non-zero vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$. Then, $2\mathbf{v}^T\mathbf{I_n}\mathbf{v} = 2(v_1^2 + v_2^2 + \ldots + v_n^2) > 0$

$D^2 f_2$ is clearly positive-definite and therefore f_2 is convex.





6. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, and $f : \mathbb{R}^n \to \mathbb{R}$, with $f(x) = ||\mathbf{A}x - \mathbf{b}||_2^2$.

(a) Assume that the columns of A are linearly independent and prove that f is strictly convex

$$\mathbf{Ax} - \mathbf{b} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n - b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n - b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n - b_m \end{bmatrix}_{(m\times 1)}$$

$$||\mathbf{Ax} - \mathbf{b}||_2^2 = (a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n - b_1)^2 + \ldots + (a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n - b_m)^2$$

$$\nabla(||\mathbf{Ax}-\mathbf{b}||_2^2) = \begin{bmatrix} 2a_{11}(a_{11}x_1 + \ldots + a_{1n}x_n - b_1) + \ldots + 2a_{m1}(a_{m1}x_1 + \ldots + a_{mn}x_n - b_n) \\ \vdots \\ 2a_{1n}(a_{11}x_1 + \ldots + a_{1n}x_n - b_1) + \ldots + 2a_{mn}(a_{m1}x_1 + \ldots + a_{mn}x_n - b_n) \end{bmatrix}_{n\times 1} = 2\mathbf{A}^T(\mathbf{Ax}-\mathbf{b})$$

$$D^2(||\mathbf{Ax} - \mathbf{b}||_2^2) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} =$$

$$= \begin{bmatrix} 2a_{11}^2 + \ldots + 2a_{m1}^2 & \dots & 2a_{11}a1n + \ldots + 2a_{m1}a_{mn} \\ \vdots & \ddots & \vdots \\ 2a_{1n}a11 + \ldots + 2a_{mn}a_{m1} & \dots & 2a_{1n}^2 + \ldots + 2a_{mn}^2 \end{bmatrix}_{(n\times n)} = 2\mathbf{A}^T\mathbf{A}$$

So,

$$\nabla f(x) = 2\mathbf{A}^T(\mathbf{Ax} - \mathbf{b})$$
$$D^2 f(x) = 2\mathbf{A}^T\mathbf{A}$$

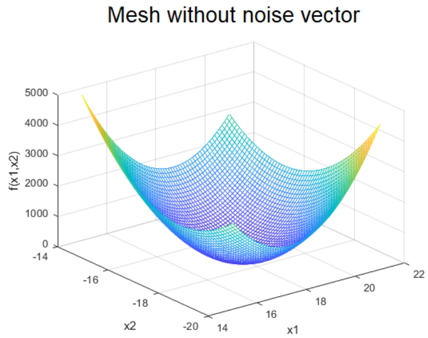
Let $\mathbf{v}$ a non-zero $n \times 1$ vector.

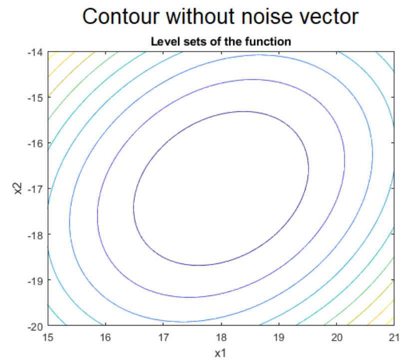
$$2\mathbf{v}^T\mathbf{A}^T\mathbf{Av} = 2(\mathbf{Av})^T\mathbf{Av} \geq 0$$

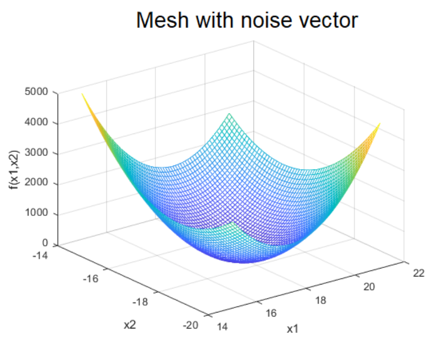
And because the columns of $\mathbf{A}$ are linearly independent $2\mathbf{v}^T\mathbf{A}^T\mathbf{Av} > 0$.
So $D^2 f(x) \succ 0$ and f is strictly convex.

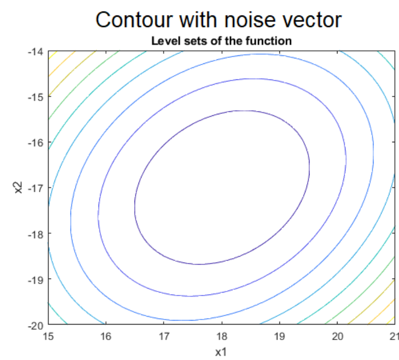
(b) Plot f for $m = 3$ and $n = 2$. In order to generate the data, generate a random (3×2) matrix A , a random (2×1) vector x, and compute b = Ax. Then, plot, via `mesh`, function f in a square around the true value x - use also the `contour` statement. What do you observe? Repeat the above procedure by assuming that b = Ax + e, where e is a "small noise" vector. What do you observe?

Matrix $\mathbf{A}$ and vector $\mathbf{x}$ elements are random integers in the range of $[-20, 20]$. Vector $\mathbf{x}$elements are also in $[-20, 20]$ and vector $\mathbf{e}$ elements are in $[-0.01, 0.01]$.









From the graphs above, it is clear that the minimum of the function is on the coordinates of the random vector $\mathbf{x}$. Without the small noise error this minimum is exactly zero, whereas for the graph with error the minimum is determined by the vector $\mathbf{e}$. Either way, it is obvious that the function is convex and smooth, based on the mesh graphs, as it was proven on 6.(a). The function's value is increasing exponentially, as it was expected. This can be observed on the contour graphs, where the sub-level lines become denser moving away from the center $\mathbf{x}$.