Fisrt Pset
Course : Optimization

Leonidas Bakopoulos A.M. 2018030036

November 7, 2022

Introduction

In this report, are going to be answeredthe exercises from the first pset of the optimization class.

First exercise

1.b

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

1.c

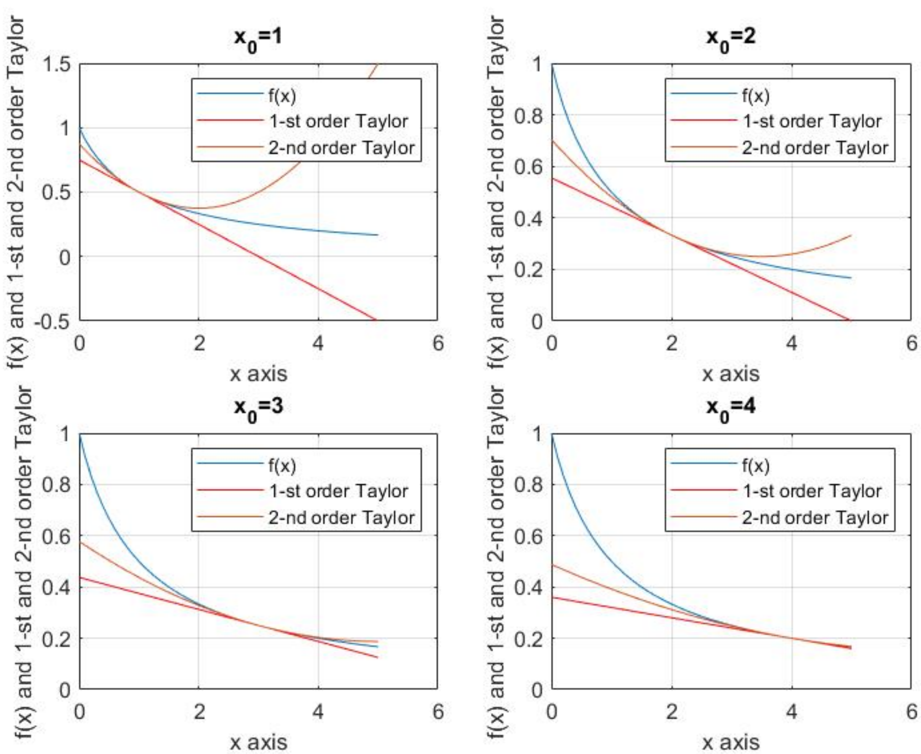


Figure 1:

In the figure above, is shown that the approximations of $f(x)$ are fitted at x_0.Due to their characteristics, (the first order is a line - ax -, the second is a curve - $ax^2 + bx$ -) the second approximation, fits to the $f(x)$ for more values of x than the other.

Second exercise

2.a

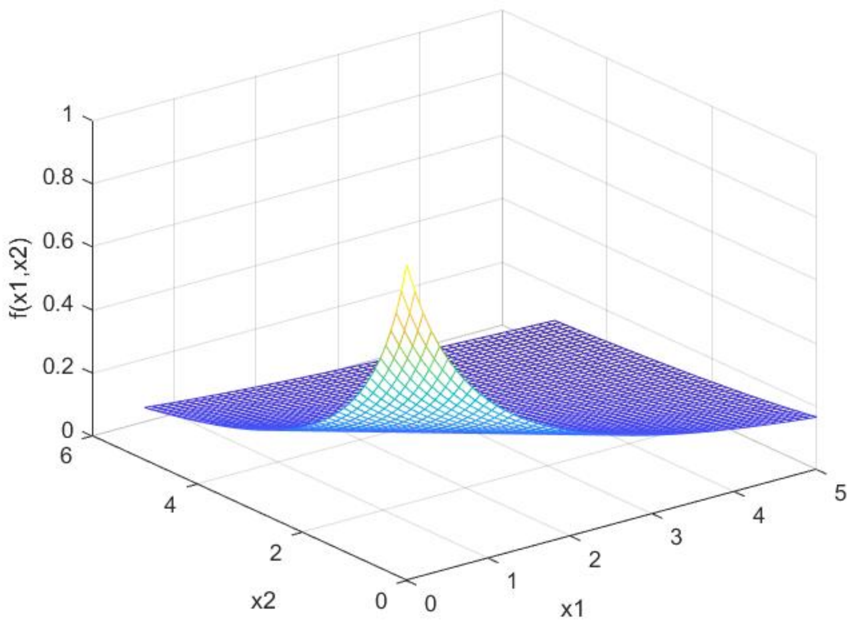


Figure 2: .

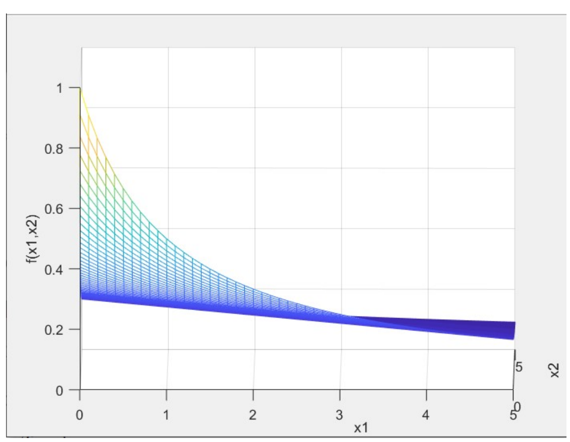
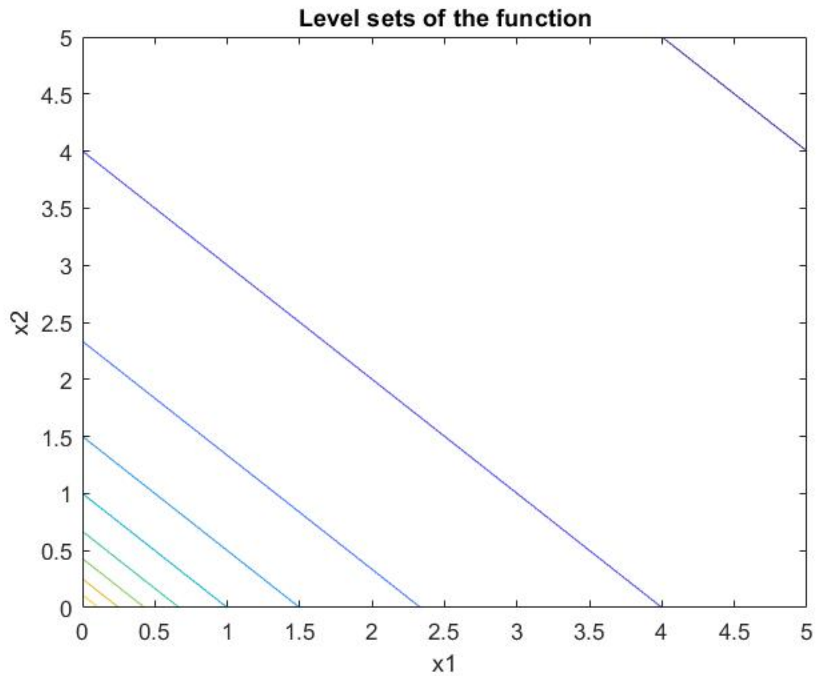


Figure 3: .

2.b



Level sets of the function
x2
x1
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5

Figure 4: .

In the figure above, we can see that f gets its maximum value for $x_1 = x_2 \simeq 0$ and decreases as x increases.

2.d

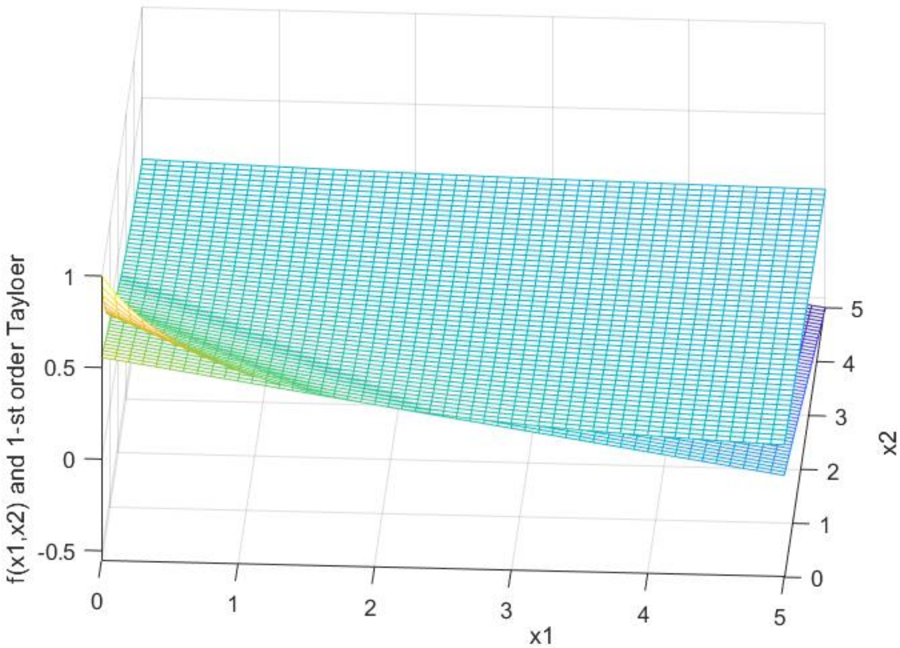


Figure 5: .

We can see, that the first order Taylor approximation, fits in the $f(x)$, for $x_0 = x_1 = 1$.

2.e

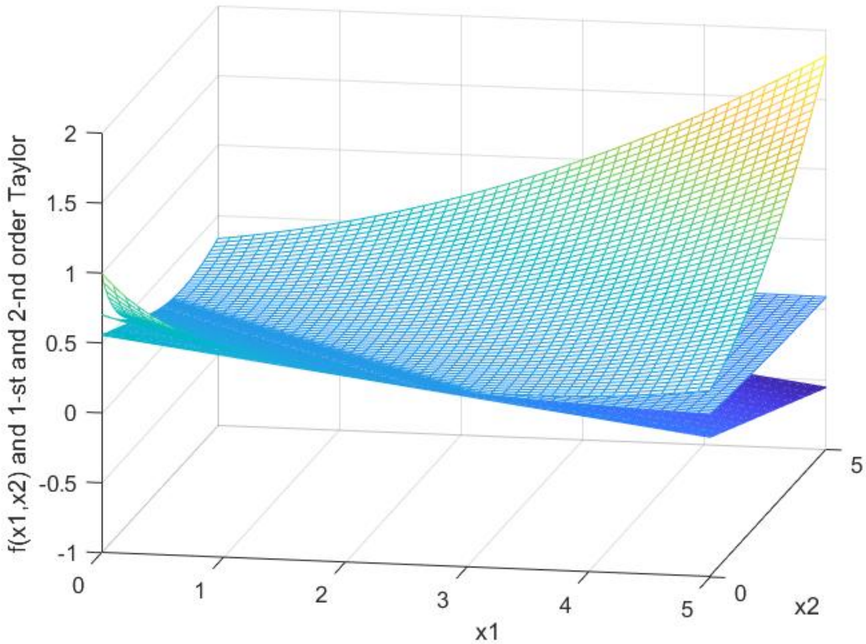


Figure 6: .

We can see, that the second order Taylor approximation, fits in the $f(x)$, for $x_0 = x_1 = 1$. Even though the second order approximation is better than the first(previous figure),sometimes the first order is more applicable (due to the complexity of calculating bigger orders).

Third exercise

3.a

Let x_1, x_2 be any two points in the halfspace $H : a^t x \leq b$. Then, $\forall \theta \varepsilon [0, 1]$, we can see that.
$\alpha^t(\theta x_1 + (1 - \theta)x_2) =$
$= \theta\alpha^t x_1 + (1 - \theta)\alpha^t x_2) \leq$
$\leq \theta b + (1 - \theta)b = b$

So $H : a^T x \leq b$ is convex.

3.b graphic proof

From the definition of affine (given book chapter 2.1.2) it is known that: "A set$C \subseteq R^n$ is affine if the line through any two distinct points in C lies in C". In the figure below, is shown that the two red points belong to the hyperplane $a^\top x \leq b$ and the line that connects them does not lie in the hyperplane. So $a^T x \leq b$ is not a hyperplane but it is convex.

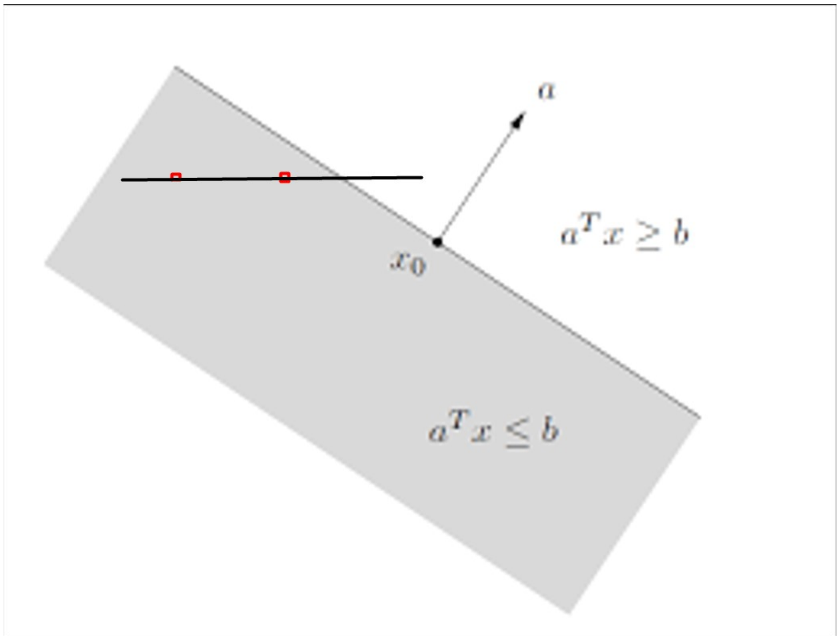


Figure 7: .

3.b mathematical proof

Let's suppose that $a = [1, 1]^\top, x = [1, 1], y = [0.5, 1]^\top$ and $b = 3$.
First of all, in order to find out if the points belong in the plane we must calculate:
$a^\top x = [1, 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2 < b$
$a^\top y = [1, 1] \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} = 1.5 < b$
So both x,y belong to the plane.
Now using those points, we are going to prove that the plane is not affine.
$a^\top \left\{ \theta \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (1 - \theta) \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \right\} = 2\theta + (1 - \theta)1.5 \leq 3 \Rightarrow \theta \leq 3$

But, in order for H to be affine, θ must belong in all of R, so H is not affine. (Note: we can see that for $\theta \epsilon [0, 1]$, the equation is satisfied, so this is another indication of convexity)

Fourth exercise

In order to lie in the hyperplane, the x_* must satisfy the $a^\top x = b$, given a and b.
Also in order to be co-linear with the a, must satisfy the $x_* = \lambda a$

so $a^\top x_* = b \Rightarrow a^\top \lambda \cdot a = b \Rightarrow \lambda a^\top a = b \Rightarrow \lambda = \frac{b}{\|a\|_2^2}$
and $x_* = \frac{b}{\|a\|_2^2} a$

Fifth exercise

5.a

$f(x) = \frac{1}{x+1} = (x + 1)^{-1}$
$\frac{\partial f}{\partial x} = -(x + 1)^{-2}$
$\frac{\partial^2 f}{\partial x^2} = 2(x + 1)^{-3} > 0$ for $x > -1$ so $f(x)$ is convex only for $x > -1$. But the domain of f is R_+ so f is convex in its domain.

5.b

$f(x_1, x_2) = \frac{1}{x_1+x_2+1}$

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

$\frac{\partial f}{\partial x_1} = \frac{\partial((x_1+x_2+1)^{-1})}{\partial x_1} = -(x_1+x_2+1)^{-2}$

$\frac{\partial^2 f}{\partial x_1^2} = 2(x_1+x_2+1)^{-3}$ $\frac{\partial^2 f}{\partial x_1 \partial x_2} = 2(x_1+x_2+1)^{-3}$ $\frac{\partial^2 f}{\partial x_2^2} = 2(x_1+x_2+1)^{-3}$ so:

$H_f = \begin{bmatrix} 2(x_1+x_2+1)^{-3} & 2(x_1+x_2+1)^{-3} \\ 2(x_1+x_2+1)^{-3} & 2(x_1+x_2+1)^{-3} \end{bmatrix}$

In order to find the eigenvalues of H_f :

$\begin{vmatrix} 2(x_1+x_2+1)^{-3} - \lambda & 2(x_1+x_2+1)^{-3} \\ 2(x_1+x_2+1)^{-3} & 2(x_1+x_2+1)^{-3} - \lambda \end{vmatrix} = 0 \Rightarrow$

$\Rightarrow 4((x_1+x_2+1)^{-3} - \lambda)^2 - 4((x_1+x_2+1)^{-3})^2 = 0$

$((x_1+x_2+1)^{-3} - \lambda)^2 = ((x_1+x_2+1)^{-3})^2 \Rightarrow$

$\Rightarrow (x_1+x_2+1)^{-3} - \lambda = \pm(x_1+x_2+1)^{-3}$so
$\lambda = 0$ or $\lambda = 2(x_1+x_2+1)^{-3}$
so $f(\mathbf{x})$ is convex (but not strictly convex), because H_f is (semi)positive in the R_+

5.c

$f(x) = x^a, x > 0$
$f'(x) = ax^{a-1}$
$f''(x) = a(a-1)x^{a-2}$
if $a \geq 1$ then$f''(x) \geq 0$ for $x > 0$ so f is a convex
if $a \leq 0$ then$f''(x) \geq 0$ for $x > 0$ so f is a convex
if $0 \leq a \leq 1$ then$f''(x) \leq 0$ for $x > 0$ s f is a concave

As previously mentioned, both orange and blue are convex and red is concave (as it can be seen in the figure below)

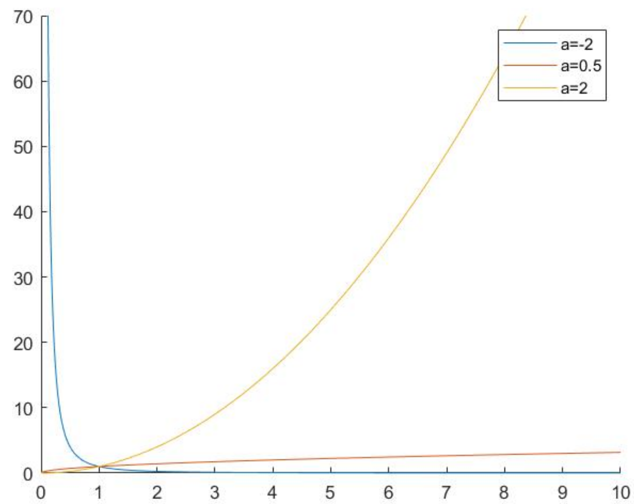


Figure 8: .

5.d

There are many types of proofs that we can follow in order to answer the question, but I am going to select the more "mathematical" way and i am not going to calculate the second derivative. First of all we are going to prove that every norm is convex(and this is also a hint for the 6^{th} exercise). By the definition of norm we know that

$\forall \nu \epsilon \mathbf{V} : \|u\| \geq 0$ and $\|u\| = 0$ iff $u = 0$.
$\forall \nu \epsilon \mathbf{V},\ \lambda \epsilon R : |\lambda|\, \|u\| = \|\lambda u\|$.
$\forall \nu, w \epsilon \mathbf{V} : \|\nu + w\| \leq \|\nu\| + \|w\|$ (Triangle inequality). Now using the triangle inequality we are going to prove that the norm is convex.
$\|\lambda v + (1-\lambda)w\| \leq \|\lambda v\| + \|(1-\lambda)w\| = \lambda\|v\| + (1-\lambda)\|w\|$
so we prove that:
$\forall v, w \in V, \lambda \in [0,1] : f(\lambda v + (1-\lambda)w) \leq \lambda f(v) + (1-\lambda)f(w)$ so f (aka norm) is convex.

Now for the $norm^2$ we are going to prove that the composition of a convex function f(x) and a convex non-decreasing function g(x) is a convex. First of all the range of f(x) (norm in our case) belongs in R. As a result, g(x) is defined in the R_* and in this sub space of R, g(x) is non-decreasing.

Proof:

$$\begin{aligned}(g \circ f)(\lambda x + (1-\lambda)y) &= g(f(\lambda x + (1-\lambda)y)) \\ &\leq g(\lambda f(x) + (1-\lambda)f(y)) \\ &\leq \lambda g(f(x)) + (1-\lambda)g(f(y)) \\ &= \lambda(g \circ f)(x) + (1-\lambda)(g \circ f)(y)\end{aligned}$$

Conclusion: As it was proved(using maths and matlab), both $f_1(x) = \|x\|_2$ and $f_2(x) = \|x\|_2^2$ are convex.

```
[x,y]=meshgrid(x_1,x_2);
for i=1:size
    for j=1:size
        tmp=[x(i,j); y(i,j)];
        f_1(i,j)=norm(tmp);
        f_2(i,j)=norm(tmp)^2;
    end
end
```

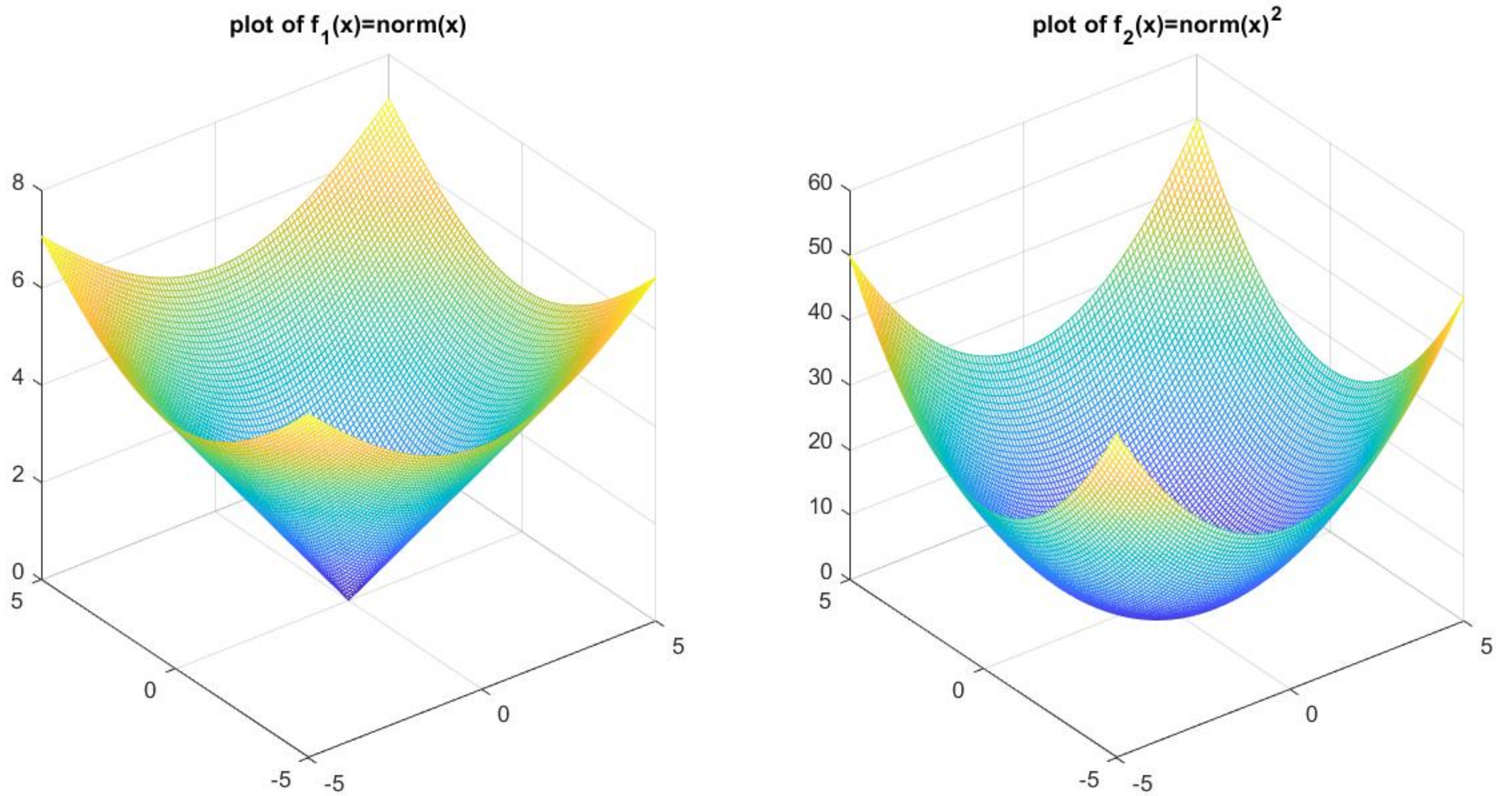

plot of f_1(x)=norm(x)
8
6
4
2
0
5
0
-5
-5
0
5
plot of f_2(x)=norm(x)^2
60
50
40
30
20
10
0
5
0
-5
-5
0
5

Figure 9: .

Sixth exercise

6.a

First of all we are going to calculate the Hessian matrix using the "matrix" representation. So:
$J = \nabla f(\mathbf{x}) = 2A^T(A\mathbf{x} - \mathbf{b}) = 2A^TA\mathbf{x} - 2A^T\mathbf{b}$ and the second derivative is :
$H = \nabla J = \nabla^2 f(\mathbf{x}) = 2A^TA$
Using the information in paragraph 9.6.6 we can see that H is positive definite so $f(\mathbf{x})$ is strictly convex.

6.b

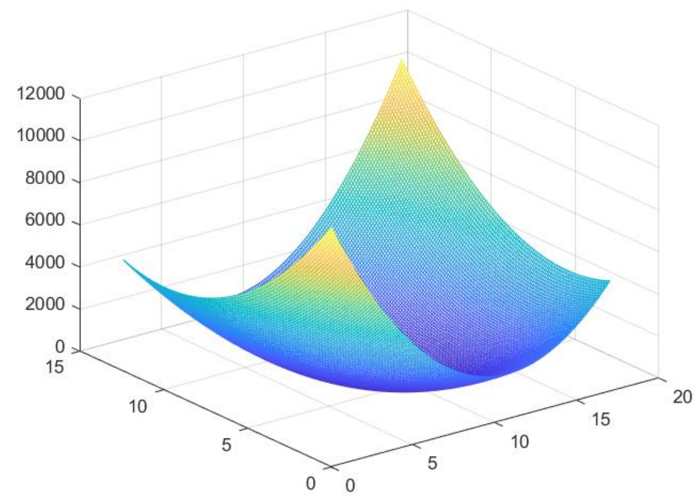


Figure 10: .

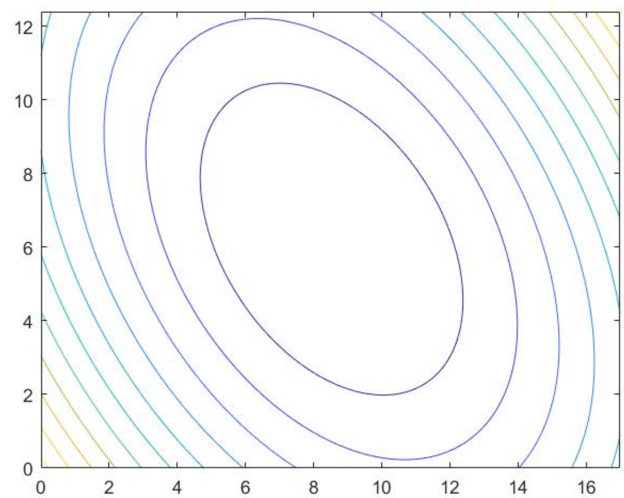


Figure 11: .

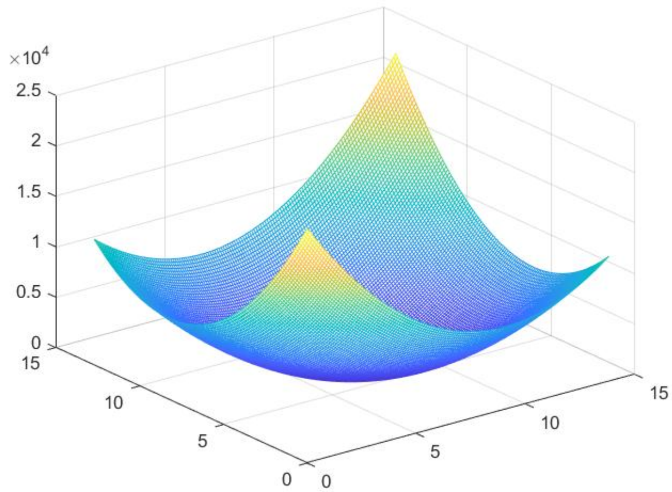


Figure 12: .

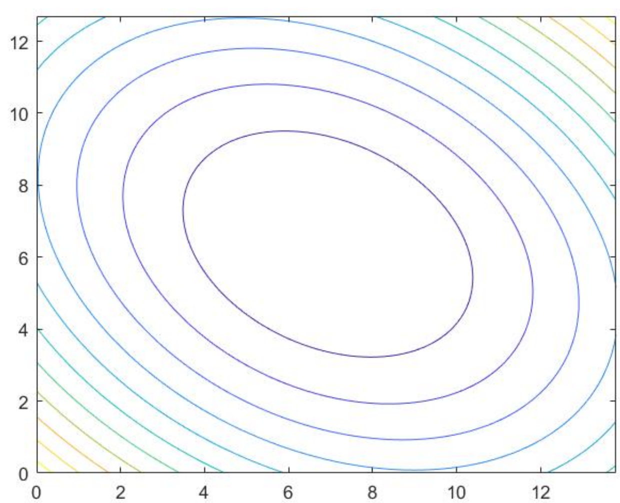


Figure 13: .

As it is shown in the figures above, $f(x) = \|\mathbf{A}x - \mathbf{b}\|_2^2$ is strictly convex. In the above matlab file , you can see that that plots have been drawn with the help of data produced by a random generator and using the values of square X that depend on that random data.

```
A=10*rand(3,2);
X=10*rand(2,1);
...
...
x_1=0:.1:2*X(1);
x_2=0:.1:2*X(2);

[x,y]=meshgrid(x_1,x_2);

....
....
[size_1,size_2]=size(x);

for i=1:size_1
    for j=1:size_2
        tmp=[x(i,j); y(i,j)];
        f(i,j)=norm(A*tmp-b)^2;;
    end
end

figure(1)
mesh(x,y,f);
figure(2)
contour(x,y,f)
```