Optimization

1st Problem Set

Kalamarakis Theodoros: 2018030022

3/11/2022

Problem 1

(0) Let $f : \mathbb{R}_+ \to \mathbb{R}$, with $f(x) = \frac{1}{1+x}$. Let $x_0 \in \mathbb{R}_+$, and define the first- and second-order Taylor approximations of f at x_0 as

$$f_{(1)}(x) = f(x_0) + f'(x_0)(x - x_0),$$

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

(a) Find the analytic expressions for functions f' and f'';

Solution:

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

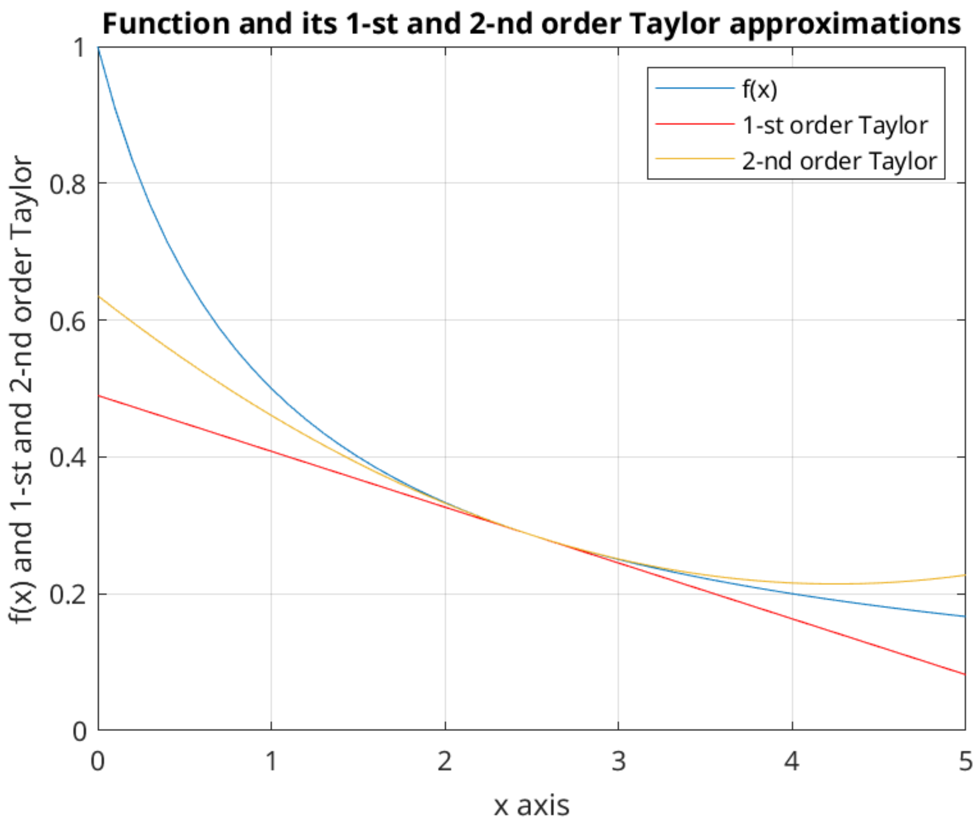
Hence the first and second order Taylor approximations will be:

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

$$f_{(2)}(x) = \frac{1}{(1 + x_0)} - \frac{1}{(1 + x_0)^2}(x - x_0) + \frac{1}{(1 + x_0)^3}(x - x_0)^2.$$

(b) (0) Draw in a common plot $f(x)$, $f_{(1)}(x)$ and $f_{(2)}(x)$ and, in order to understand the behavior of the approximations, consider various values of x_0.

Figure 1: common plot for $f(x)$, $f_{(1)}(x)$ and $f_{(2)}(x)$.

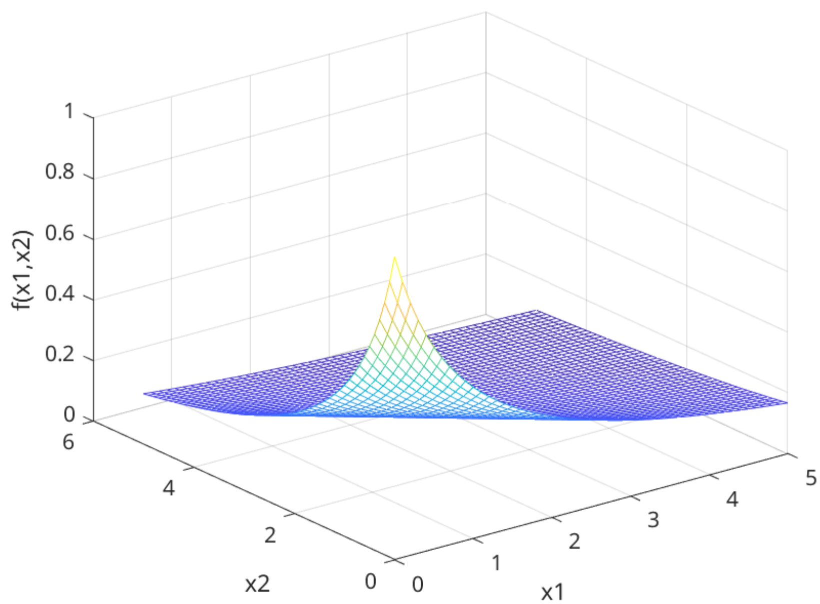


.

Problem 2

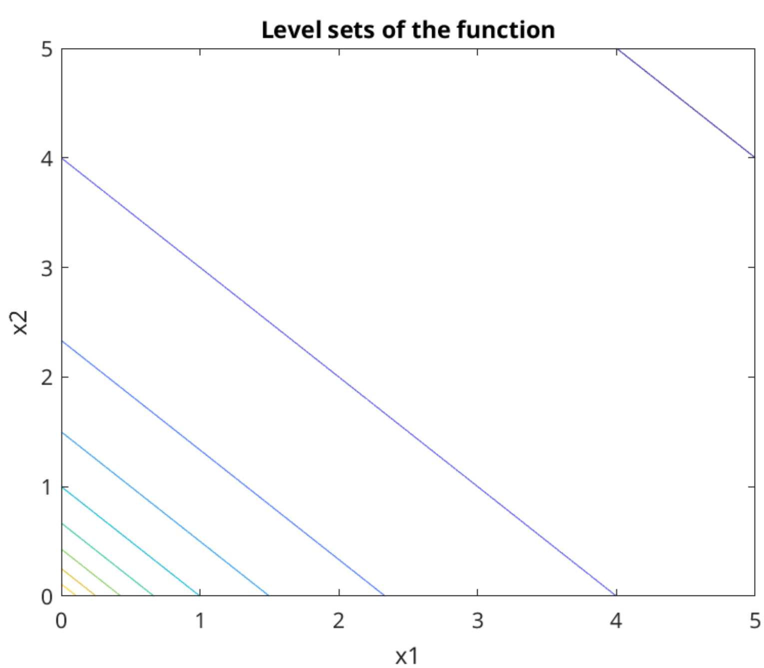
(0) Let $f : \mathbb{R}^2_+ \to \mathbb{R}$, with $f(x_1, x_2) = \frac{1}{1+x_1+x_2}$.

(a) (0) Compute and plot, via mesh, f for $x_1,\ x_2 \in [0, x_*]$, with $x_* > 0$.



.

(b) (0) Plot the level sets of f , via contour. What do you observe? Can you explain the phenomenon?



.

(c) (0) Compute the first- and second-order Taylor approximations of f at point

$\mathbf{x_0} = (x_{0,1}, x_{0,2})$.

Solution:

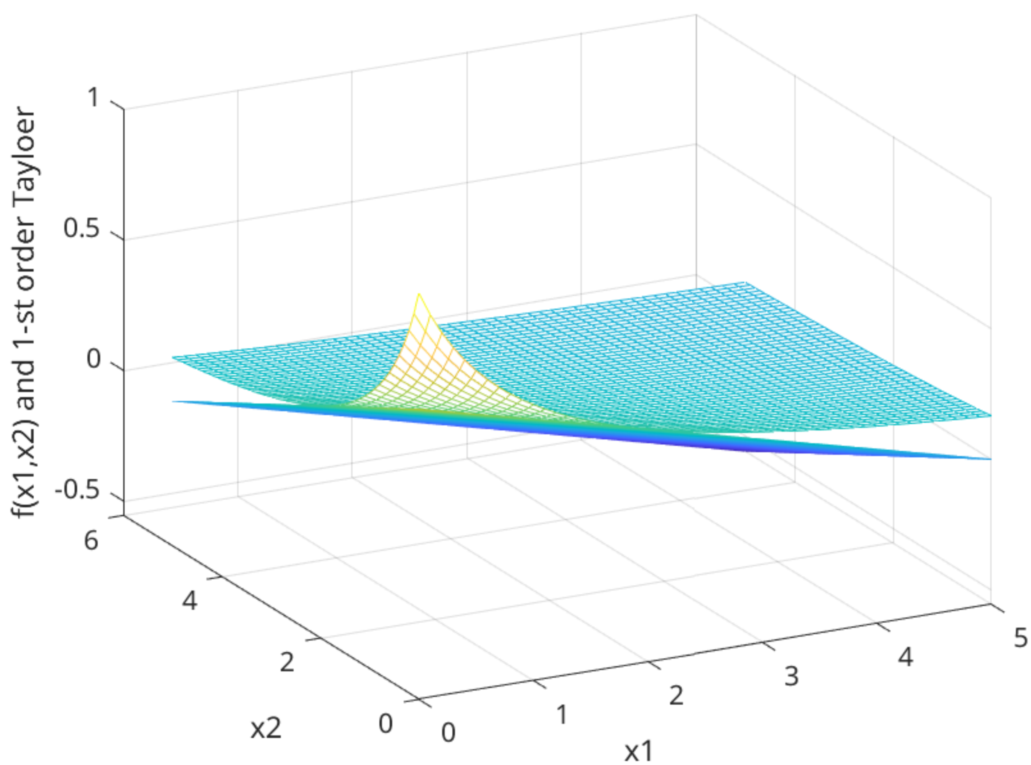
$$Df = \left(-\frac{1}{(1+x_1+x_2)^2}\ ,\ -\frac{1}{(1+x_1+x_2)^2}\right) \text{ and } D^2 f = \begin{pmatrix} \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \\ \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \end{pmatrix}$$

The first and second order Taylor aproxiamtions are computed as:

$$f_{(1)}(\mathbf{x}) = f(\mathbf{x_0}) + Df(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0}),$$

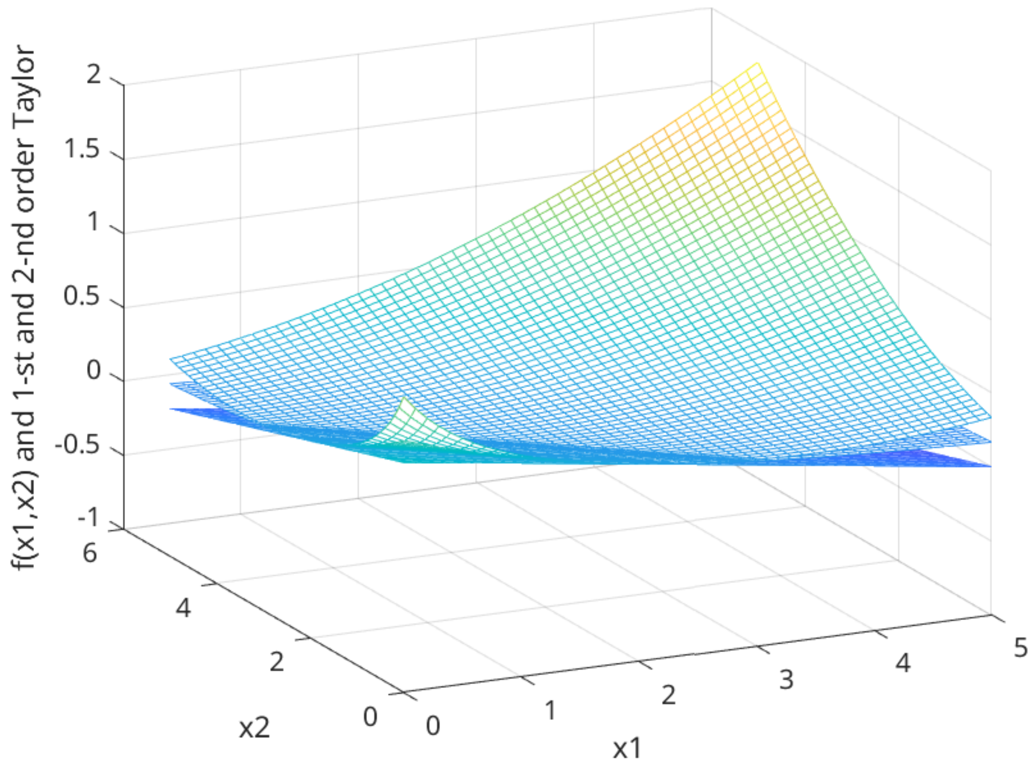
$$f_{(2)}(\mathbf{x}) = f(\mathbf{x_0}) + Df(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0}) + \frac{1}{2}(\mathbf{x} - \mathbf{x_0})^T D^2 f(\mathbf{x_0})(\mathbf{x} - \mathbf{x_0}).$$

(d) (0) Draw on a common plot f and its first-order Taylor approximation.



.

(e) (0) Draw on a common plot f and its second-order Taylor approximation.



.

Problem 3

(20) Let $\mathbb{S}_{\mathbf{a},b} = \{\mathbf{x} \in \mathbb{R}_n | \mathbf{a}^T\mathbf{x} \le b\}$

(a) (5) Prove that $\mathbb{S}_{\mathbf{a},b}$ **is convex.**

Solution:

Let $\mathbf{x_1}, \mathbf{x_2} \in \mathbb{S}_{\mathbf{a},b}$

Then

$$\left\{\begin{array}{c} \mathbf{a}^T\mathbf{x_1} \le b \\ \\ \mathbf{a}^T\mathbf{x_2} \le b \end{array}\right\} \xLeftrightarrow[0<\theta_2<1]{0<\theta_1<1} \left\{\begin{array}{c} \mathbf{a}^T\theta_1\mathbf{x_1} \le \theta_1 b \\ \\ \mathbf{a}^T\theta_2\mathbf{x_2} \le \theta_2 b \end{array}\right\} \Leftrightarrow \mathbf{a}^T(\theta_1\mathbf{x_1} + \theta_2\mathbf{x_2}) \le (\theta_1 + \theta_2)b \xLeftrightarrow{\theta_1+\theta_2=1}$$

$$\xLeftrightarrow{\theta_1+\theta_2=1} \mathbf{a}^T(\theta_1\mathbf{x_1} + \theta_2\mathbf{x_2}) \leq b$$

Hence $\theta_1\mathbf{x_1} + \theta_2\mathbf{x_2} \in \mathbb{S}_{\mathbf{a},b}$ for $\theta_1 + \theta_2 = 1$

(b) (15) **Prove that $\mathbb{S}_{\mathbf{a},b}$ is not affine (a counterexample is sufficient).**

Solution:

Let $\mathbf{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $b = 2$

then the vectors $\mathbf{x_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{x_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ lie on the hyperplane $\mathbb{S}_{\mathbf{a},b}$ because:

$$\mathbf{a}^T\mathbf{x_1} = [\,2 \quad 1\,] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2 + 0 \leq 2 = b$$

$$\mathbf{a}^T\mathbf{x_2} = [\,2 \quad 1\,] \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0 + 1 \leq 2 = b$$

However for a constant $\theta = 2$ we have:

$$\mathbf{a}^T(\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}) = [\,2 \quad 1\,](2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 1 \begin{bmatrix} 0 \\ 1 \end{bmatrix}) = [\,2 \quad 1\,] \begin{bmatrix} 2 \\ -1 \end{bmatrix} = 4 - 1 = 3 \geq b$$

which proves that $\mathbb{S}_{\mathbf{a},b}$ is **not** affine

---

Problem 4

(10) **Find the point $\mathbf{x}_*$ that is co-linear with a and lies on the hyperplane** $\mathbb{H}_{a,b} := \{\mathbf{x} \in \mathbb{R}_n \mid \mathbf{a}^T\mathbf{x} = b\}$.

Solution:

$\mathbf{x}_*$ is co-linear with $\mathbf{a}$ means that $\mathbf{x}_* = k\mathbf{a}$ (1)

if $\mathbf{x}_*$ lies on the hyperplane $\mathbb{H}_{\mathbf{a},b}$ then

$$\mathbf{a}^T\mathbf{x}_* = b \overset{(1)}{\Longleftrightarrow} \mathbf{a}^T k\mathbf{a} = b \Leftrightarrow k\|\mathbf{a}\| = b \Leftrightarrow k = \frac{b}{\|\mathbf{a}\|} \ (2)$$

Hence

$$\overset{(1)}{\underset{(2)}{\Longrightarrow}} \mathbf{x}_* = b\frac{\mathbf{a}}{\|\mathbf{a}\|}$$

Problem 5

(20) Check whether the following functions are convex or not (using, for example, the second derivative rule).

(a) (5) $f : \mathbb{R}_+ \to \mathbb{R}$**, with** $f(x) = \frac{1}{1+x}$

Solution:

$\frac{\partial f(x)}{\partial x} = -\frac{1}{(1+x)^2} \Rightarrow \frac{\partial^2 f(x)}{\partial^2 x} = \frac{2}{(1+x)^3} > 0$ since $x > 0$

(b) (5) $f : \mathbb{R}^2_+ \to \mathbb{R}$**, with** $f(x1, x2) = \frac{1}{1+x_1+x_2}$

Solution:

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

let $\mathbf{v} = [v_1, v_2]^T \in \mathbb{R}^2$

Then

$$\mathbf{v}^T \nabla^2 f \mathbf{v} \Rightarrow \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \\ \\ \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \end{bmatrix} \begin{bmatrix} v_1 \\ \\ v_2 \end{bmatrix} =$$

$$= \begin{bmatrix} v_1 \frac{2}{(1+x_1+x_2)^3} + v_2 \frac{2}{(1+x_1+x_2)^3} & , v_1 \frac{2}{(1+x_1+x_2)^3} + v_2 \frac{2}{(1+x_1+x_2)^3} \end{bmatrix} \begin{bmatrix} v_1 \\ \\ v_2 \end{bmatrix} =$$

$$= v_1^2 \frac{2}{(1+x_1+x_2)^3} + v_1 v_2 \frac{2}{(1+x_1+x_2)^3} + v_1 v_2 \frac{2}{(1+x_1+x_2)^3} + v_2^2 \frac{2}{(1+x_1+x_2)^3} =$$

$$= \frac{2}{(1+x_1+x_2)^3}(v_1^2 + 2v_1v_2 + v_2^2) = \frac{2}{(1+x_1+x_2)^3}(v_1+v_2)^2 > 0$$

since $x_1, x_2 > 0$

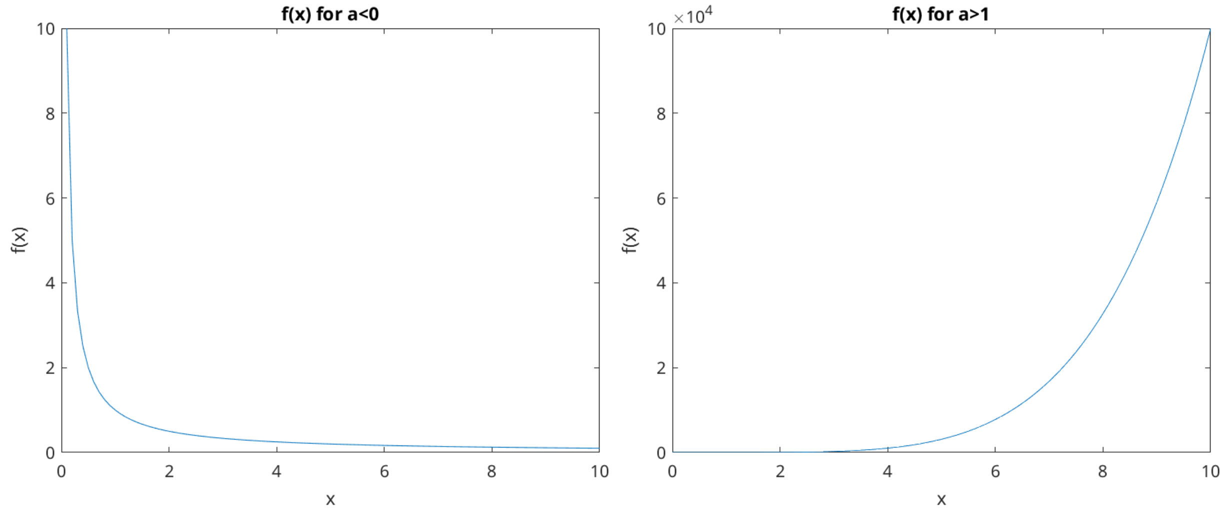
Hence $\nabla^2 f \succeq \mathbf{O}$ which means that f is convex

(c) (5) $f : \mathbb{R}_{++} \to \mathbb{R}$, 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 $a \leq 0$;

$\frac{\partial^2 f(x)}{\partial^2 x} = a(a-1)x^{a-2} \geq 0$ since $a \geq 1$ or $a \leq 0$

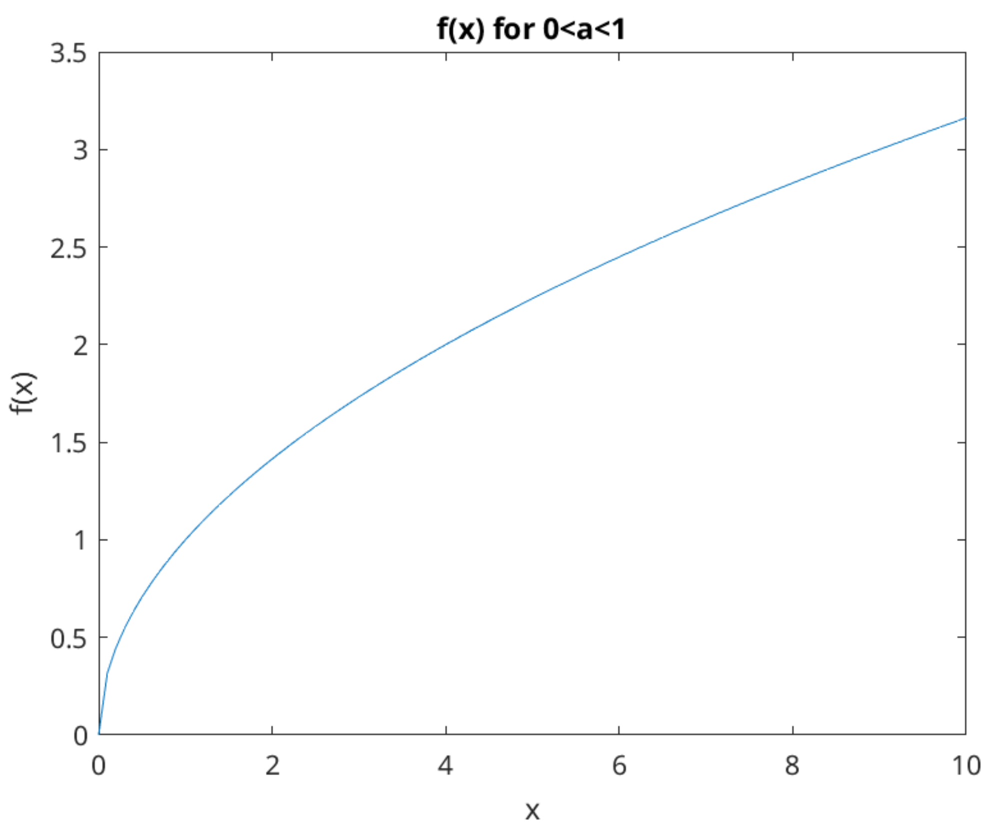
Figure 2: plots for function f for $a \geq 1$ and $a \leq 0$.



ii. $0 \leq a \leq 1$.

$\frac{\partial^2 f(x)}{\partial^2 x} = a(a-1)x^{a-2} \leq 0$ since $0 \leq a \leq 1$

Figure 3: plots for function f for $0 \leq a \leq 1$.



(d) (5) $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$, with $f_1(\mathbf{x}) = \|\mathbf{x}\|_2$ and $f_2(\mathbf{x}) = \|\mathbf{x}\|_2^2$ (plot the functions for n = 2).

Solution:

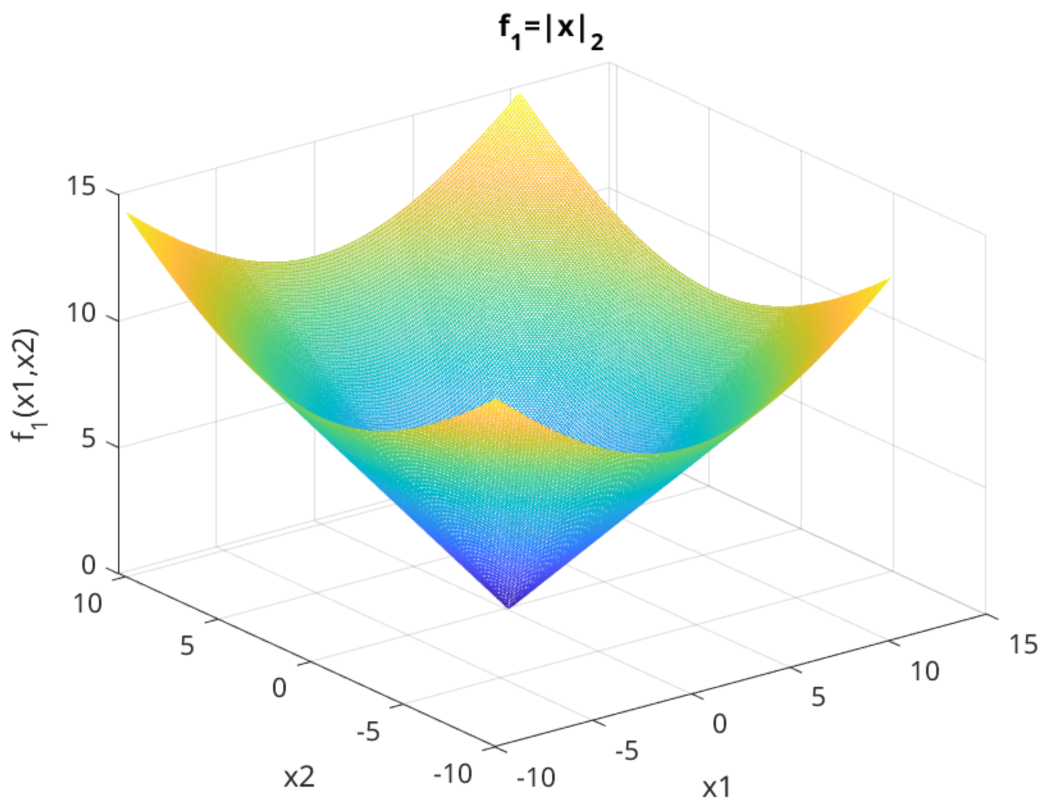
For $f_1(\mathbf{x}) = \|\mathbf{x}\|_2$ using the triangle inequality for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and a $0 \leq \theta \leq 1$ we have

$$\|\theta\mathbf{x} + (1-\theta)\mathbf{y}\|_2 \leq \|\theta\mathbf{x}\|_2 + \|(1-\theta)\mathbf{y}\|_2 = \theta\|\mathbf{x}\|_2 + (1-\theta)\|\mathbf{y}\|_2 \; (since \; 0 \leq \theta \leq 1) \Leftrightarrow$$

$$\Leftrightarrow f_1(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f_1(\mathbf{x}) + (1-\theta) f_1(\mathbf{y})$$

The inequality above indicates that f_1 is strictly convex

Figure 4: plot for function f_1.



For $f_2(\mathbf{x}) = \|\mathbf{x}\|_2^2$:

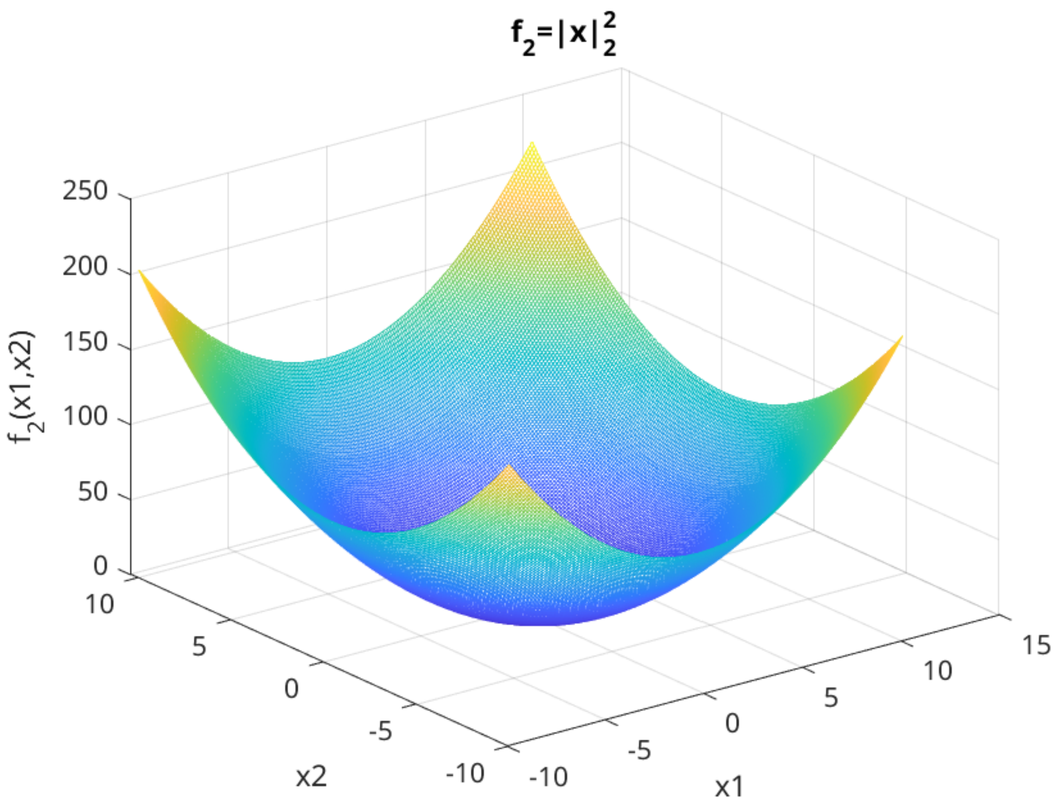
$$f_2(\mathbf{x}) = \|\mathbf{x}\|_2^2 = x_1^2 + x_2^2 + x_3^2 + ... + x_n^2$$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \\ \cdots \\ \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ \\ 2x_2 \\ \\ \cdots \\ \\ 2x_n \end{bmatrix} = 2\mathbf{x}$$

Hence

$$\nabla^2 f(\mathbf{x}) = 2\mathbf{I} \succeq \mathbf{O}$$

Figure 5: plot for function f_2.



Problem 6

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

(a) (10) Assume that the columns of A are linearly independent and prove that f is strictly convex (Prove that the Hessian of $f(\mathbf{x})$ is positive definite).

Solution:

Let $\mathbf{z} = \mathbf{Ax} - \mathbf{b}$ and $g(\mathbf{x}) = \|\mathbf{x}\|_2^2$ then $f(\mathbf{x}) = g(\mathbf{z})$

Hence

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial g(\mathbf{z})}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial(\|\mathbf{z}\|_2^2)}{\partial \mathbf{z}} \frac{\partial(\mathbf{Ax} - \mathbf{b})}{\partial \mathbf{x}} = 2\mathbf{z}^T\mathbf{A}(*) \Rightarrow$$

$$\Rightarrow \frac{\partial^2 f(\mathbf{x})}{\partial^2 \mathbf{x}} = \frac{\partial(2\mathbf{z}^T\mathbf{A})}{\partial \mathbf{x}} = \frac{2(\mathbf{Ax} - \mathbf{b})^T\mathbf{A}}{\partial \mathbf{x}} = \frac{2(\mathbf{x}^T\mathbf{A}^T\mathbf{A} - \mathbf{b}^T\mathbf{A})}{\partial \mathbf{x}} = \mathbf{A}^T\mathbf{A} = \nabla^2 f$$

$(*) \frac{\partial(\|\mathbf{x}\|_2^2)}{\partial \mathbf{x}} = \nabla(\|\mathbf{x}\|_2^2)^T = 2\mathbf{x}^T$ as proven in Problem 5 (d)

For any vector $\mathbf{v} \in \mathbb{R}^n$ we have:

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

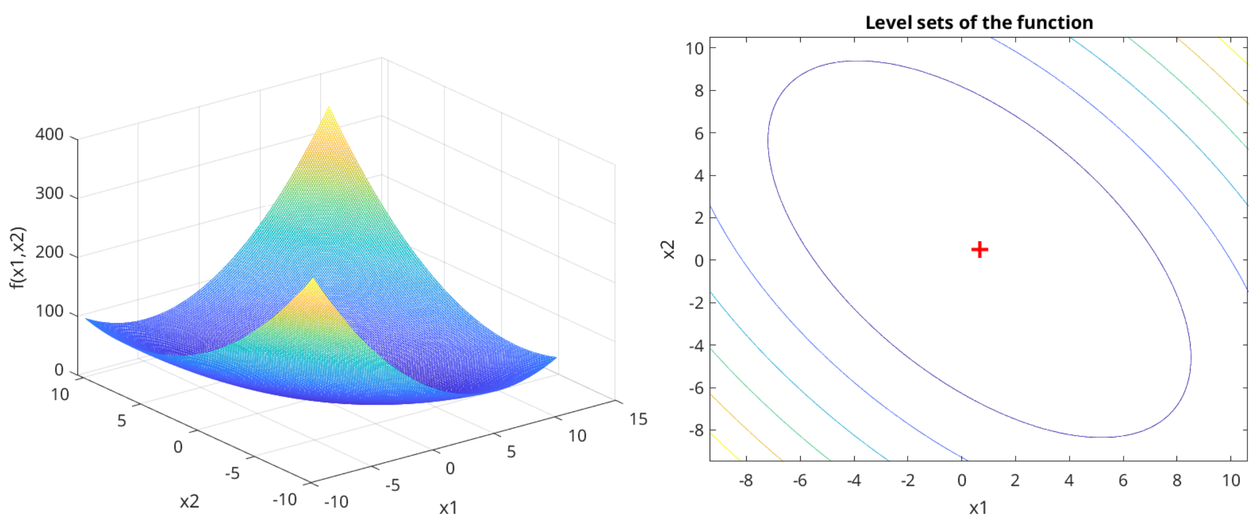
But since the columns of $\mathbf{A}$ are linearly independent we know that there is no vector $\mathbf{v}$ so that $\mathbf{Av} = \mathbf{0}$

Hence

$$(\mathbf{Av})^T(\mathbf{Av}) > 0 \Leftrightarrow \nabla^2 f \succ \mathbf{O}$$

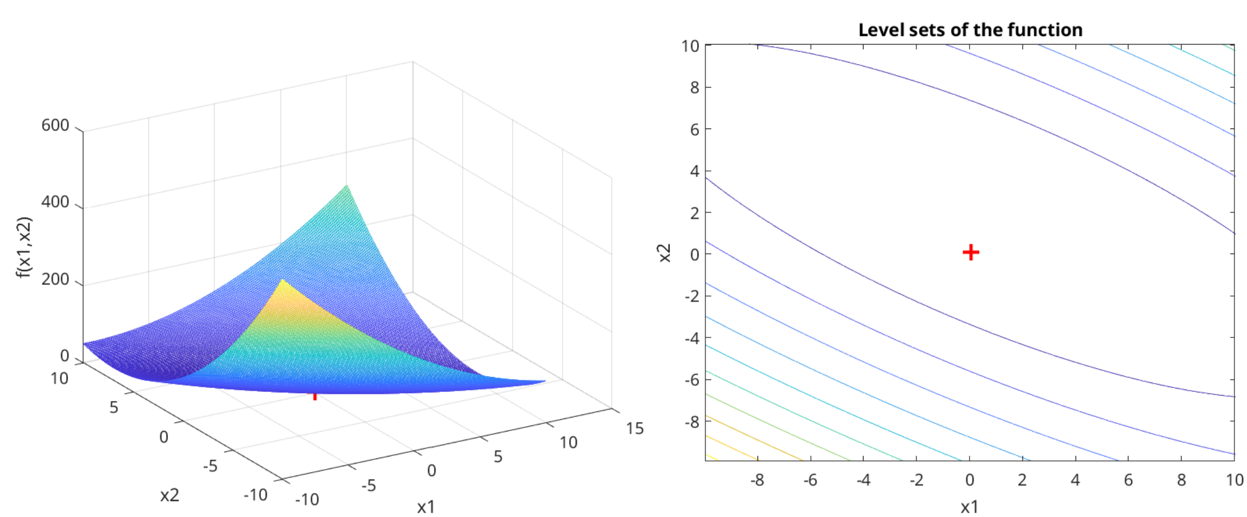
(b) (10) Plot f for $m = 3$ and $n = 2$. In order to generate the data, generate a random (3 $\times$ 2) matrix A, a random (2 $\times$ 1) vector x, and compute b = Ax. Then, plot, via mesh, function f in a square around the true value x - use also the contour statement. What do you observe? Repeat the above procedure by assuming that $b = \mathbf{Ax} + \mathbf{e}$, where e is a "small noise" vector. What do you observe?

Figure 6: mesh and contour plots for function f without noise. The red cross marks the true value $\mathbf{x}$



We observe that function $f(\mathbf{x})$ is indeed convex and the The value of the function attains a minimum at the point $\mathbf{x}$ (where $\mathbf{b} = \mathbf{A}\mathbf{x}$)

Figure 7: mesh and contour plots for function f with noise. The red cross marks the true value $\mathbf{x}$



We observe that function is still convex regardless the noise, yet the minimun of $f(\mathbf{x})$ has been shifted away from the point $\mathbf{x}$ (where $\mathbf{b} = \mathbf{A}\mathbf{x}$)