Technical University of Crete

School of Electrical and Computer Engineering

Course: Optimization 2023-24

Dimitris Angelopoulos 2020030038 Exercise 1

Problem 1

In this problem we will approximate the function $f : \mathbb{R}_+ \to \mathbb{R}$ with $f(x) = \frac{1}{1+x}$ using the first and second order Taylor approximations around a point $x_0 \in \mathbb{R}_+$.

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

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

(a) The analytic expressions of the first and second derrivative of f are :

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

(b) Now we draw the approximations for different x_0 as shown in Fig. 1.

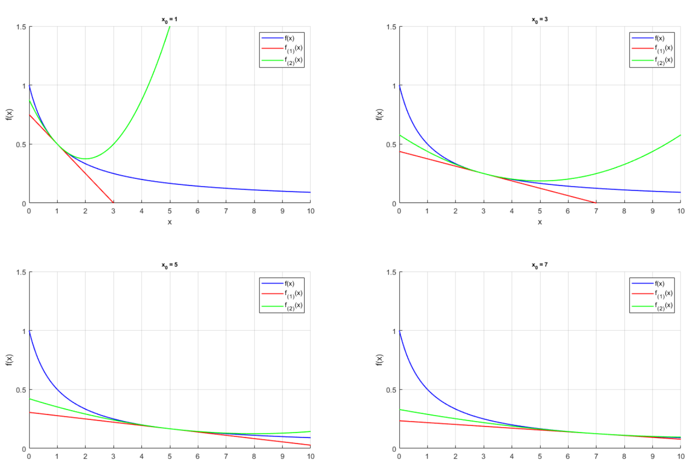

Fig. 1

Problem 2

In this problem we consider the function $f : \mathbb{R}^2_+ \to \mathbb{R}$, with $f(x_1, x_2) = \frac{1}{1+x_1+x_2}$.

(a) We compute and plot f using the function mesh of MATLAB and we get the plot in Fig. 2.

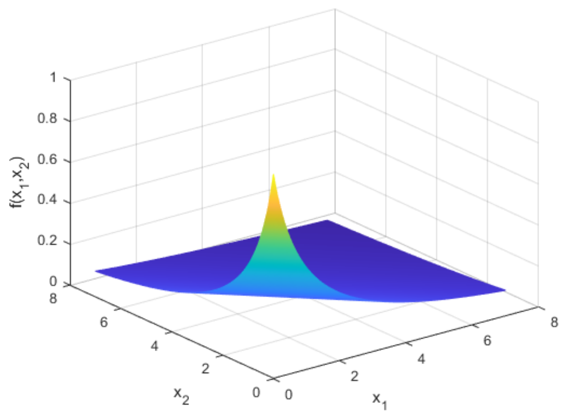

Fig. 2

(b) Now we will plot the same function using contour (Fig. 3).

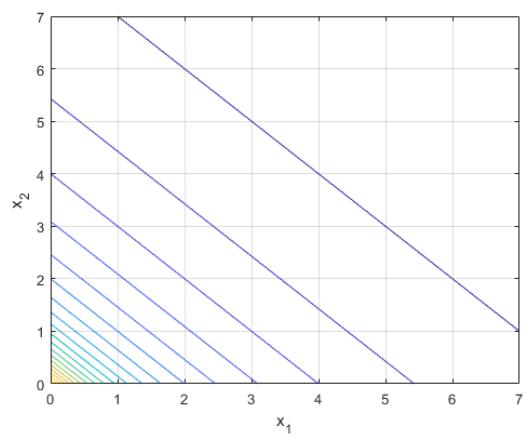

Fig. 3

We observe that as we approach the maximum value of f the level sets get more dense. Each level set represents the line defined by setting $f(x_1, x_2)$ equal to a constant c. As the constant c gets larger we approach the maximum value of f.

(c) The first and second order Taylor approximations in point the point $\mathbf{x}_0 = (x_{0,1}, x_{0,2})$ are:

$f_{(1)}(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0)$ and

$f_{(2)}(\mathbf{x}) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \nabla^2 f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$

Where :

$$\mathbf{x}_0 = \begin{bmatrix} 1.8 \\ 1.8 \end{bmatrix}, \nabla f(\mathbf{x}) = \begin{bmatrix} \frac{-1}{(1+x_1+x_2)^2} \\ \frac{-1}{(1+x_1+x_2)^2} \end{bmatrix} \text{ and } \nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \\ \frac{2}{(1+x_1+x_2)^3} & \frac{2}{(1+x_1+x_2)^3} \end{bmatrix}$$

(d)-(e) Now that we have computed the gradient and the hessian we can plot the function approximations to a common mesh. The results are shown in Fig. 4.

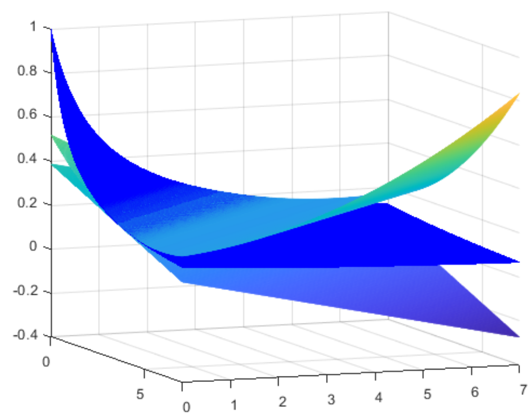

Fig. 4

The blue plain corresponds to the function f. The bottom one corresponds to the first order approximation and the top(quadratic) one to the second order approximation.

Problem 3

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

(a) To prove $\mathbb{S}_{\mathbf{a},b}$ is convex we will consider two random points $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}_{\mathbf{a},b}$. We will show that $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathbb{S}_{\mathbf{a},b}$ with $0 \leq \theta \leq 1$.

Since $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{S}_{\mathbf{a},b}$ we know :

$$\begin{cases} \mathrm{a}^T\mathbf{x}_1 \leq b \\ \mathrm{a}^T\mathbf{x}_2 \leq b \end{cases}$$

Multiplying each inequality with θ and $(1-\theta)$ respectively, for $0 < \theta < 1$, we get :

$$\begin{cases} \theta\mathbf{a}^T\mathbf{x}_1 \leq \theta b \\ (1 \text{ - } \theta)\mathbf{a}^T\mathbf{x}_2 \leq (1-\theta)b \end{cases}$$

Adding them will give us :

$$\mathbf{a}^T(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \leq b$$

For $\theta = 0$ the linear combination gives us $\mathbf{x}_2 \in \mathbb{S}_{\mathbf{a},b}$, similarly for $\theta = 1 \Rightarrow \mathbf{x}_1 \in \mathbb{S}_{\mathbf{a},b}$

That means that $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathbb{S}_{\mathbf{a},b}$ for every $0 \leq \theta \leq 1$ and thus the set is convex.

(b) Now we will prove that $\mathbb{S}_{\mathbf{a},b}$ is not affine.

We assume that the set is affine. For $n = 2$, $\mathbf{a} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $b = 5$ $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^2$ and $\theta \in \mathbb{R}$ such that $\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2 \in \mathbb{S}_{\mathbf{a},b}$.

Choosing $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ results in $\theta \leq 3$, which means that the hypothesis does not hold for every θ, i.e $\mathbb{S}_{\mathbf{a},b}$ is not affine by contradiction.

Problem 4

In this problem we will find the point $\mathbf{x}_*$ that is co-linear with $\mathbf{a}$ and lies to the hyperplane $\mathbb{H}_{\mathbf{a},b} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b\}$.

We know that $\mathbf{x}_* \in \mathbb{H}_{\mathbf{a},b} \Rightarrow \mathbf{a}^T\mathbf{x}_* = b$ (1).

Since $\mathbf{x}_*, \mathbf{a}$ are co-linear, we get $\mathbf{x}_* = \lambda\mathbf{a}$ (2).

Using the equations (1),(2) we get :

$$\lambda\mathbf{a}^T\mathbf{a} = b \Rightarrow$$

$$\lambda = \tfrac{b}{\|\mathbf{a}\|_2^2}$$

The point that is co-linear with $\mathbf{a}$ and lies to $\mathbb{H}_{\mathbf{a},b}$ is

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

Problem 5

We will study whether the following functions are convex or not.

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

$f''(x) = \frac{2}{(1+x)^3} > 0$, as $x \geq 0$

Thus f is convex.

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

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

Since the reduced determinants of the Hessian are $det_1 = \frac{2}{(1+x_1+x_2)^3} > 0$ and $det_2 = 0$. That means that f is convex.

(c) $f : \mathbb{R}_{++} \to \mathbb{R}$ with $f(x) = x^\alpha$

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

i) For $\alpha \geq 1$ or $\alpha \leq 0$ we can observe that $f''(x) \geq 0$ which means that f is convex.
ii) For $0 \leq \alpha \leq 1$ $f''(x) \leq 0$ thus f is not convex.

We can verify this by plotting $f(x)$ for different values of α as shown in Fig 5.

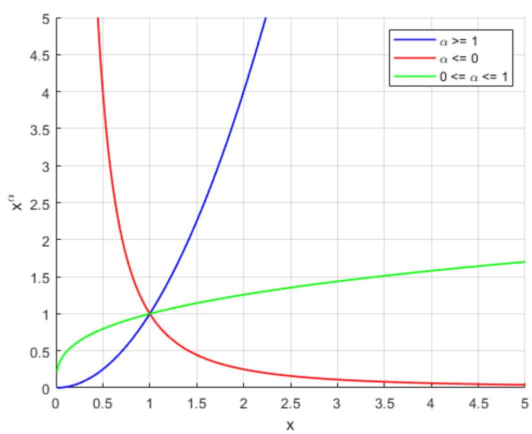

Fig. 5

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

For $n = 2$ we plot the functions and we get Fig. 6:

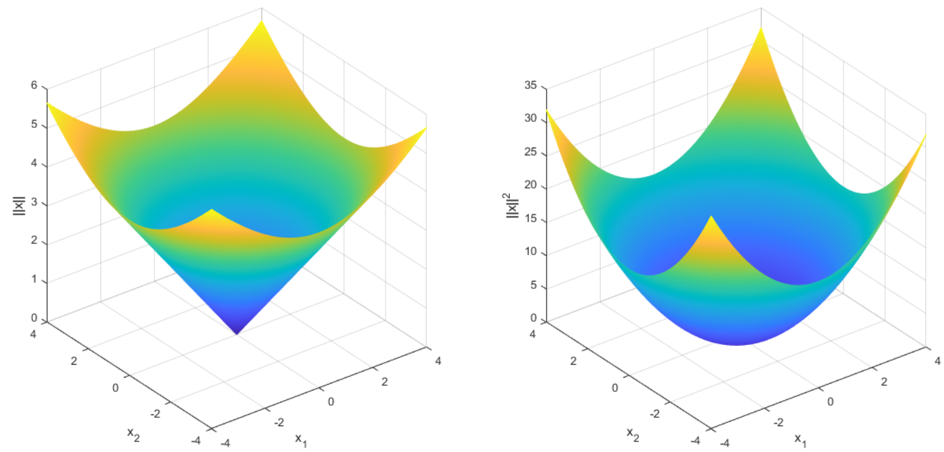

Fig. 6

Indeed both functions are convex. To prove convexity for f_1 we will show that for each $\mathbf{x}, \mathbf{y} \in \mathbf{dom} f_1$ and $0 \leq \theta \leq 1$, it holds that

$$f_1(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f_1(\mathbf{x}) + (1-\theta) f_1(\mathbf{y}) (5.1)$$

From the triangular inequality we get:

$$||\theta\mathbf{x} + (1-\theta)\mathbf{y}|| \leq ||\theta\mathbf{x}|| + ||(1-\theta)\mathbf{y}||,$$
$$||\theta\mathbf{x}|| + ||(1-\theta)\mathbf{y}|| = \theta||\mathbf{x}|| + (1-\theta)||\mathbf{y}||$$

That means that (5.1) holds and thus f_1 is convex.

Now we will prove convexity for f_2 using the second order conditions. f_2 can be re-written in the following form :

$$f_2(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$$

The Gradient and Hessian are :

$$\nabla f_2(\mathbf{x}) = 2\mathbf{x}$$

and

$$\nabla^2 f_2(\mathbf{x}) = 2\mathbb{I}_{nxn} \succ \mathbf{O}$$

Since the hessian is a positive definite matrix f_2 is convex.

Problem 6

In this problem we will study convex quadratic functions of a more general form. Let $\mathbf{P} \in \mathbb{S}^2_{++}$, where $\mathbb{S}^2_{++}$ is the set of all (2x2) positive definite matrices, $\mathbf{q} \in \mathbb{R}^2$ and $r \in \mathbb{R}$. Consider the quadratic function $f : \mathbb{R}^2 \to \mathbb{R}$ with :

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r$$

Since we know f is convex it can be minimized by setting its gradient to $\mathbf{0}$.

$$\nabla f(\mathbf{x}) = \mathbf{P}\mathbf{x} + \mathbf{q}$$

$\mathbf{P}$ is a (2x2) positive definite matrix and thus it can be inverted. The global minimum is the point:

$$\mathbf{x}_* = -\mathbf{P}^{-1}\mathbf{q}$$

Now we will plot the function choosing $\mathbf{P}, \mathbf{q}, r$ randomly in a way that yield f convex and we get the following figures.

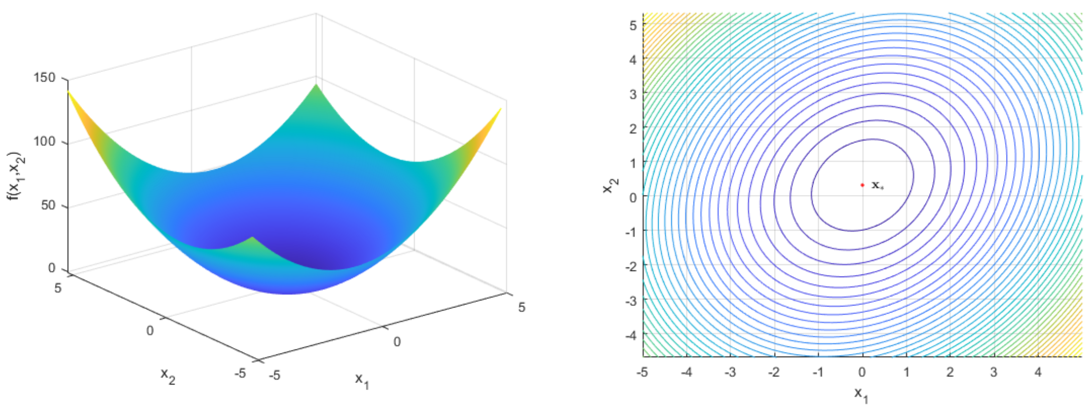

Fig. 7

Indeed it can be verified that the minimum exists. $\mathbf{q}, r$ are chosen to be a white Gaussian random vector and a white Gaussian random variable respectively and that is the reason that the minimum is close to $\mathbf{0}$ for every implementation of f ran in MATLAB.