EECS 127/227AT Optimization Models in Engineering
Spring 2019 Homework 1

Release date: 9/05/19.
Due date: 9/12/19, 23:00 (11 pm). Please LaTeX or handwrite your homework solution and submit an electronic version.

Submission Format
Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned) as well as your IPython notebook saved as a PDF.

If you do not attach a PDF "printout" of your IPython notebook, you will not receive credit for problems that involve coding. Make sure that your results and your plots are visible. Assign the IPython printout to the correct problem(s) on Gradescope.

1. Norm and angles

(a) Let $x, y \in \mathbb{R}^n$ be two unit-norm vectors, that is, such that $\|x\|_2 = \|y\|_2 = 1$. Show algebraically that the vectors $x - y$ and $x + y$ are orthogonal. Then, show this graphically by drawing the two vectors on the 2D plane, as well as any other necessary shapes. You may use right angles, circles and straight lines to make your point.

Solution: When x, y are both unit-norm, we have

$$(x-y)^\top(x+y) = x^\top x - y^\top y - y^\top x + x^\top y = x^\top x - y^\top y = 0,$$

as claimed. Let us note that we can express any vector $z \in \text{span}(x, y)$ as $z = \lambda x + \mu y$, for

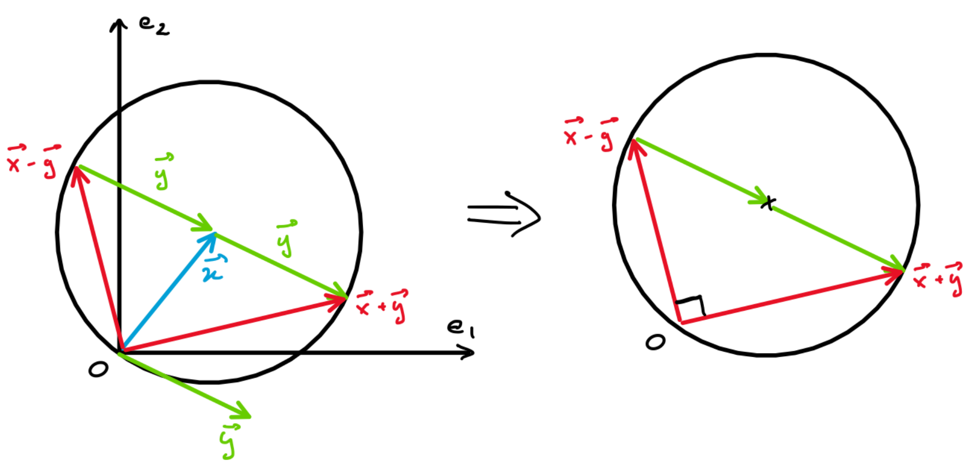

Figure 1: Because $\|x\|_2 = \|y\|_2$, we can put the triangle form by the origin O, the point $x - y$ and $x + y$ in a circle with ($x - y$,$x + y$) as diameter. Therefore $(O, x - y) \perp (O, x + y)$

some $\lambda, \mu \in \mathbb{R}$. We have $z = \alpha u + \beta v$, where

$$\alpha = \frac{\lambda + \mu}{2}, \quad \beta = \frac{\lambda - \mu}{2}.$$

Hence $z \in \mathrm{span}(u, v)$. The converse is also true for similar reasons. Thus, (u, v) is an orthogonal basis for $\mathrm{span}(x, y)$. We finish by normalizing u, v, replacing them with $(u/\|u\|_2, v/\|v\|_2)$. The desired orthogonal basis is thus given by $((x-y)/\|x-y\|_2, (x+y)/\|x+y\|_2)$.

(b) Show that the following inequalities hold for any vector $x \in \mathbb{R}^n$:

$$\frac{1}{\sqrt{n}}\|x\|_2 \le \|x\|_\infty \le \|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2 \le n\|x\|_\infty.$$

Hint: For $\|x\|_1 \le \sqrt{n}\|x\|_2$, how might you express $\|x\|_1$ as the dot product of two vectors? Can you then use the Cauchy-Schwarz inequality to bound this dot product?

Solution: We have

$$\|x\|_2^2 = \sum_{i=1}^n x_i^2 \le n \cdot \max_i \, x_i^2 = n \cdot \|x\|_\infty^2.$$

Also, $\|x\|_\infty \le \sqrt{x_1^2 + \ldots + x_n^2} = \|x\|_2$.
The inequality $\|x\|_2 \le \|x\|_1$ is obtained after squaring both sides, and checking that

$$\sum_{i=1}^n x_i^2 \le \sum_{i=1}^n x_i^2 + \sum_{i \ne j} |x_i x_j| = \left(\sum_{i=1}^n |x_i|\right)^2 = \|x\|_1^2.$$

Finally, the condition $\|x\|_1 \le \sqrt{n}\|x\|_2$ is due to the Cauchy-Schwarz inequality

$$|z^\top y| \le \|y\|_2 \cdot \|z\|_2,$$

applied to the two vectors $y = (1, \ldots, 1)$ and $z = |x| = (|x_1|, \ldots, |x_n|)$.

(c) Show that for any non-zero vector x,

$$\mathrm{card}(x) \ge \frac{\|x\|_1^2}{\|x\|_2^2},$$

where $\mathrm{card}(x)$ is the *cardinality* of the vector x, defined as the number of non-zero elements in x. Find all vectors x for which the lower bound is attained.

Hint: Try using Cauchy-Schwarz like in the previous part.

Solution: Let us apply the Cauchy-Schwarz inequality with $z = |x|$ again, and with y a vector with $y_i = 1$ if $x_i \ne 0$, and $y_i = 0$ otherwise. We have $\|y\|_2 = \sqrt{k}$, with $k = \mathrm{card}(x)$. Hence

$$|z^\top y| = \|x\|_1 \le \|y\|_2 \cdot \|z\|_2 = \sqrt{k} \cdot \|x\|_2,$$

which proves the result. The bound is attained for vectors with k non-zero elements, all with the same magnitude.

2. Gradients and Hessians

The *gradient* of a scalar-valued function $g : \mathbb{R}^n \to \mathbb{R}$, is the column vector of length n, denoted as ∇g, containing the derivatives of components of g with respect to the variables:

$$(\nabla g(x))_i = \frac{\partial g}{\partial x_i}(x),\ i = 1, \dots n.$$

The *Hessian* of a scalar-valued function $g : \mathbb{R}^n \to \mathbb{R}$, is the $n \times n$ matrix, denoted as $\nabla^2 g$, containing the second derivatives of components of g with respect to the variables:

$$(\nabla^2 g(x))_{ij} = \frac{\partial^2 g}{\partial x_i \partial x_j}(x),\quad i = 1, \dots, n,\quad j = 1, \dots, n.$$

For the remainder of the class, we will repeatedly have to take gradients and Hessians of functions we are trying to optimize. This exercise serves as a warm up for future problems.

Compute the gradients and Hessians for the following functions:

(a) $g(x) = y^\top Ax$

Solution: Let $A = [a_1, a_2, \dots, a_n]$. then

$$\begin{aligned} g(x) &= y^\top Ax \\ &= y^\top [a_1,\ a_2,\ \dots\ , a_n]x \\ &= y^\top (a_1 x_1 + a_2 x_2 +\ \dots\ + a_n x_n) \\ &= \sum_{i=1}^n (y^\top a_i) x_i. \end{aligned}$$

Thus

$$\frac{\partial g}{\partial x_j}(x) = y^\top a_j = a_j^\top y$$

so $\nabla g(x) = A^\top y$. Since the gradient does not depend on x, we then have $\nabla^2 g(x) = 0$.

(b) $g(x) = x^\top Ax$

Solution: Let $A = [a_1,\ a_2,\ \dots\ , a_n]$ where a_i is the i-th column of A. Similarly, let $a_i^\top$ be the i-th row of $A^\top$. For notational convenience, let α_i^T denote the i-th row of A. Finally, let a_{ij} denote the (i, j)th entry of A. Then

$$\begin{aligned} g(x) &= x^\top Ax \\ &= x^\top [a_1,\ a_2,\ \dots\ , a_n]x \\ &= x^\top (a_1 x_1 + a_2 x_2 +\ \dots\ + a_n x_n) \\ &= \sum_{i=1}^n (x^\top a_i) x_i. \end{aligned}$$

Then,

$$\begin{aligned} \frac{\partial g}{\partial x_j}(x) &= \frac{\partial}{\partial x_j}\Big[(x^\top a_j)x_j + \sum_{i \neq j} (x^\top a_i) x_i\Big] \\ &= x^\top a_j + a_{jj} x_j + \sum_{i \neq j} a_{ji} x_i \\ &= a_j^\top x + \alpha_j^\top x. \end{aligned}$$

It follows that $\nabla g(x) = (A + A^\top)x$. Note if A is symmetric this reduces to $2Ax$. Based on the definition of the Hessian, it follows that the ith column of the Hessian is the ith column of $A + A^\top$. Thus $\nabla^2 g(x) = A + A^\top$.

(c) $g(x) = \|Ax - b\|_2^2$

Solution: Expanding the norm and using the fact that $c^\top x = x^\top c$ we have that $g(x) = x^\top A^\top Ax - 2b^\top Ax + b^\top b$. Using the previous results and the fact that $A^\top A$ is symmetric, it follows that $\nabla g(x) = 2(A^\top Ax - A^\top b)$ and $\nabla^2 g(x) = 2A^\top A$.

(d) $g(x) = \sin(x_1^2)\log(x_3 - x_2)$ where x_i are scalars and $x_3 - x_2 > 0$.

Solution: Our first order derivatives are

$$\begin{aligned}
\frac{\partial g}{\partial x_1}(x) &= 2x_1 \cos(x_1^2)\log(x_3 - x_2) \\
\frac{\partial g}{\partial x_2}(x) &= -\sin(x_1^2)\frac{1}{x_3 - x_2} \\
\frac{\partial g}{\partial x_3}(x) &= \sin(x_1^2)\frac{1}{x_3 - x_2}.
\end{aligned}$$

The second order derivatives are

$$\begin{aligned}
\frac{\partial^2 g}{\partial x_1^2}(x) &= \log(x_3 - x_2)\Big(2\cos(x_1^2) - 4x_1^2\sin(x_1^2)\Big) \\
\frac{\partial^2 g}{\partial x_2 \partial x_1}(x) = \frac{\partial^2 g}{\partial x_1 \partial x_2}(x) &= -2x_1\cos(x_1^2)\frac{1}{x_3 - x_2} \\
\frac{\partial^2 g}{\partial x_3 \partial x_1}(x) = \frac{\partial^2 g}{\partial x_1 \partial x_3}(x) &= 2x_1\cos(x_1^2)\frac{1}{x_3 - x_2} \\
\frac{\partial^2 g}{\partial x_2^2}(x) &= -\frac{\sin(x_1^2)}{(x_3 - x_2)^2} \\
\frac{\partial^2 g}{\partial x_2 \partial x_3}(x) = \frac{\partial^2 g}{\partial x_3 \partial x_2}(x) &= \frac{\sin(x_1^2)}{(x_3 - x_2)^2} \\
\frac{\partial^2 g}{\partial x_3^2}(x) &= \frac{-\sin(x_1^2)}{(x_3 - x_2)^2}.
\end{aligned}$$

3. Jacobians

The *Jacobian* of a vector-valued function $g : \mathbb{R}^n \to \mathbb{R}^m$ is the $m \times n$ matrix, denoted as Dg, containing the derivatives of components of g with respect to the variables:

$$(Dg)_{ij} = \frac{\partial g_i}{\partial x_j}, \ \ i = 1, \dots, m, \ \ j = 1, \dots, n.$$

In class, we will also be using the Jacobian frequently.

Compute the Jacobians for the following maps:

(a) $g(x) = Ax$

Solution: Note $g_i(x) = \alpha_i^\top x$ where $\alpha_i^\top$ is the i-th row of A. Then $\dfrac{\partial g_i}{\partial x_j} = \alpha_{ij}$ which is simply the (i, j) entry of A. It follows that $Dg(x) = A$.

(b) $g(x) = f(x)x$ where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is once-differentiable

Solution: Again $g_i(x) = f(x)x_i$. Then

$$\frac{\partial g_i}{\partial x_i} = f(x) + (\nabla f(x))_i x_i$$
$$\frac{\partial g_i}{\partial x_j} = 0 + (\nabla f(x))_j x_i.$$

It follows that $Dg(x) = x(\nabla f(x))^\top + f(x)I$.

(c) $g(x) = f(Ax + b)x$ where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is once differentiable and $A \in \mathbb{R}^{n\times n}$.

Solution: Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ a once differentiable function. We note ∇f the gradient of f. By definition $\nabla f = (D_i f)_{i\in\mathbb{N}_n}$.

Begin remark ....

Some notations include that if $\mathbf{z} \in \mathbb{R}^n$ is the argument of f – i.e. $f(\mathbf{z})$ – then $D_i f$ is written as $D_i f = \frac{\partial f}{\partial z_i}$. Evaluated in $\mathbf{x}$ the gradient could be denoted as $\nabla f(\mathbf{x}) = (D_i f(\mathbf{x}))_{i\in\mathbb{N}_n} = (\frac{\partial f}{\partial z_i}(\mathbf{x}))_{i\in\mathbb{N}_n}$. Some notations might include using the same variable for the evaluation point of the gradient and the variable used for the function: $\nabla f(\mathbf{z}) = (D_i f(\mathbf{z}))_{i\in\mathbb{N}_n} = (\frac{\partial f}{\partial z_i}(\mathbf{z}))_{i\in\mathbb{N}_n}$.

Some notation might include $\nabla_x f(Ax + b)\Big|_x$ to consider the gradient of the function $x \mapsto f(Ax + b)$ evaluated in x. We consider these notations as really bad because they might be confusing for the readers (and even the writer!).

End remark ....

Let $h : \mathbb{R}^n \mapsto \mathbb{R}^n$ define such that $h(x) = Ax + b$. h is differentiable as a linear function.

Let $\hat{g} : \mathbb{R}^n \mapsto \mathbb{R}$ define such that $\hat{g}(x) = f(Ax+b) = f(h(x))$. The function $\hat{g}$ is the composition of h and f and it denoted $\hat{g} = f \circ h$.

Let $g : \mathbb{R}^n \mapsto \mathbb{R}^n$ define such that $g(x) = \hat{g}(x)x = f(h(x))x = f(Ax + b)x$.

First, we derive $\nabla\hat{g}$. Let $\alpha_i^\top$ and a_i denote the i-th row and i-th column of A respectively. Finally let a_{ij} denote (i, j)th entry of A. We can now write $h(x) = (h_j(x))_{j\in\mathbb{N}_n}$ with $h_j(x) =$

$\alpha_i^\top x + b_i$. Then by the chain rule we have

$$\begin{aligned} D_i \hat{g} = \frac{\partial \hat{g}}{\partial x_i} &= \frac{\partial (f \circ h)}{\partial x_i} \\ &= \sum_{j=1}^{n} ((D_j f) \circ h) \frac{\partial h_j}{\partial x_i} \\ &= \sum_{j=1}^{n} ((D_j f) \circ h) \, a_{ji} \\ &= a_i^\top (\nabla f) \circ h \end{aligned}$$

It follows that $\nabla \hat{g} = A^\top (\nabla f) \circ h$. This might be written as $\nabla \hat{g}(x) = A^\top (\nabla f)(Ax + b)$.

Begin remark . . . .

With the previously really bad notations, the gradient of $\hat{g}$ might be written as $\nabla_x f(Ax + b)\Big|_x = A^\top \nabla_x f(x)\Big|_{Ax+b}$. This sometimes becomes $\nabla_x f(Ax + b)\Big|_x = A^\top \nabla f\Big|_{Ax+b}$. Which also sometimes becomes $\nabla_x f(Ax + b) = A^\top \nabla f\Big|_{Ax+b}$. Which also sometimes becomes $\nabla_x f(Ax + b) = A^\top \nabla f(Ax + b)$. Which also sometimes becomes $\nabla f(Ax + b) = A^\top \nabla f(Ax + b)$. Which is confusing for the reader who might see a recursive definition of $\nabla f(Ax + b)$. We consider that this is not valid notation and therefore should not be written by the students.

End remark . . . .

Returning to the original problem, we have $g_i(x) = \hat{g}(x) x_i = (f \circ h)(x) x_i = f(Ax + b) x_i$. Then using the derivation in the previous part, it follows that $Dg(x) = x(\nabla \hat{g}(x))^\top + \hat{g}(x) I$.
Unrolling the notations: $Dg(x) = x(A^\top (\nabla f)(Ax + b))^\top + f(Ax + b) I = x((\nabla f)(Ax + b))^\top A + f(Ax + b) I$.
The solution is: $Dg(x) = x((\nabla f)(Ax + b))^\top A + f(Ax + b) I$.

(d)

$$g(x) = \begin{bmatrix} x_1^2 / x_2 \\ \log(x_3) \sin(x_1 / x_3) \end{bmatrix}$$

Solution: We see that $g : \mathbb{R}^3 \to \mathbb{R}^2$. Then $Dg =$

$$\begin{bmatrix} \frac{2x_1}{x_2} & \frac{-x_1^2}{x_2^2} & 0 \\ \frac{\log(x_3)\cos(x_1/x_3)}{x_3} & 0 & \frac{\sin(x_1/x_3)}{x_3} - \frac{\log(x_3)\cos(x_1/x_3) x_1}{x_3^2} \end{bmatrix}.$$

4. Level Sets

Plot/hand-draw the level sets of the following functions.

Also draw the gradient directions in the level-set diagram. Additionally, compute the first and second order Taylor series approximation around the point $(1, 1)$ for each function and comment on how accurately they approximate the true function.

(a) $g(x_1, x_2) = \frac{x_1^2}{4} + \frac{x_2^2}{9}$

Solution:

We first compute the first and second order partial derivatives of g as follows:

$$\frac{\partial g}{\partial x_1}(x_1, x_2) = \frac{x_1}{2}, \quad \frac{\partial g}{\partial x_2}(x_1, x_2) = \frac{2x_2}{9},$$
$$\frac{\partial^2 g}{\partial x_1^2}(x_1, x_2) = \frac{1}{2}, \quad \frac{\partial^2 g}{\partial x_2 x_1}(x_1, x_2) = 0,$$
$$\frac{\partial^2 g}{\partial x_2^2}(x_1, x_2) = \frac{2}{9}, \quad \frac{\partial^2 g}{\partial x_1 x_2}(x_1, x_2) = 0.$$

The gradient of g is then given by,

$$\nabla g(x_1, x_2) = \begin{bmatrix} \frac{\partial g}{\partial x_1}(1, 1) \\ \frac{\partial g}{\partial x_2}(1, 1) \end{bmatrix},$$

and the Hessian matrix is given by,

$$H(x_1, x_2) = \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2}(x_1, x_2) & \frac{\partial^2 g}{\partial x_1 x_2}(x_1, x_2) \\ \frac{\partial^2 g}{\partial x_2 x_1}(x_1, x_2) & \frac{\partial^2 g}{\partial x_2^2}(x_1, x_2) \end{bmatrix}.$$

The first order Taylor series approximation around $(1, 1)$ can be computed as:

$$g(x_1, x_2) \approx g(1, 1) + (\nabla g(1, 1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix}$$
$$= \frac{13}{36} + \frac{x_1}{2} - \frac{1}{2} + \frac{2x_2}{9} - \frac{2}{9} = \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36}.$$

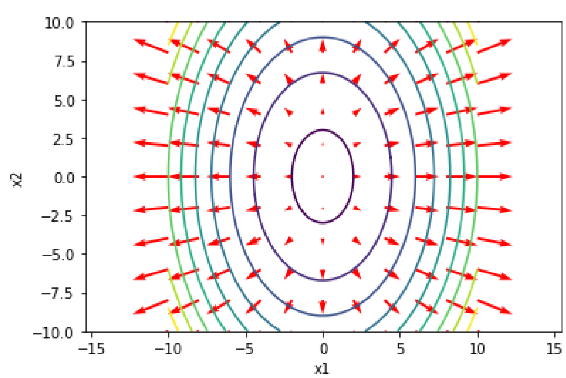

Figure 2: Level sets and gradient directions for the function $g(x_1, x_2) = \frac{x_1^2}{4} + \frac{x_2^2}{9}$.

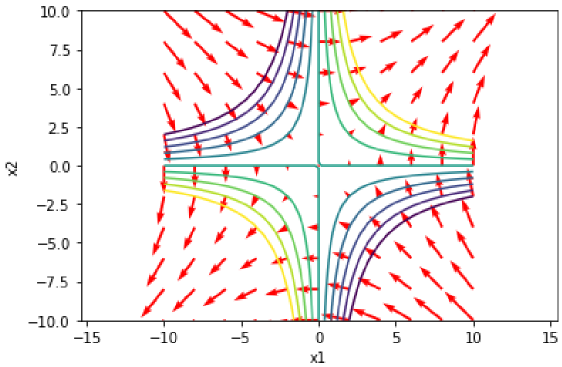

Figure 3: Level sets and gradient directions for the function $g(x_1, x_2) = x_1 x_2$.

The second order Taylor series approximation around $(0, 0)$ can be computed as:

$$\begin{aligned} g(x_1,x_2) &\approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 - 1 & x_2 - 1 \end{bmatrix} H(1,1) \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \\ &= \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36} + \frac{1}{2}\left(\frac{1}{2}(x_1-1)^2 + \frac{2}{9}(x_2-1)^2\right) \\ &= \frac{(x_1-1)^2}{4} + \frac{(x_2-1)^2}{9} + \frac{x_1}{2} + \frac{2x_2}{9} - \frac{13}{36}. \\ &= \frac{x_1^2}{4} + \frac{x_2^2}{9} \end{aligned}$$

The original function at (1.1,1.1) takes on the value 0.437. The first order approximation returns, evaluated at (1.1,1.1): $\frac{1.1}{2} + \frac{1.2}{9} - \frac{13}{36} = 0.433$. Additionally, observe that the second order approximation simplifies to return the original function!

(b) $g(x_1, x_2) = x_1 x_2$

Solution:

We follow the same steps as in the previous part of the problem. The partial derivatives for this g are given by:

$$\begin{aligned} \frac{\partial g}{\partial x_1}(x_1,x_2) &= x_2, & \frac{\partial g}{\partial x_2}(x_1,x_2) &= x_1, \\ \frac{\partial^2 g}{\partial x_1^2}(x_1,x_2) &= 0, & \frac{\partial^2 g}{\partial x_2 x_1}(x_1,x_2) &= 1, \\ \frac{\partial^2 g}{\partial x_2^2}(x_1,x_2) &= 0, & \frac{\partial^2 g}{\partial x_1 x_2}(x_1,x_2) &= 1. \end{aligned}$$

The first order Taylor series approximation around $(1, 1)$ can be computed as:

$$\begin{aligned} g(x_1,x_2) &\approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \\ &= 1 + x_1 - 1 + x_2 - 1 = x_1 + x_2 - 1. \end{aligned}$$

The second order Taylor series approximation around $(1, 1)$ can be computed as:

$$\begin{aligned}
g(x_1, x_2) &\approx g(1,1) + (\nabla g(1,1))^\top \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x_1 - 1 & x_2 - 1 \end{bmatrix} H(1,1) \begin{bmatrix} x_1 - 1 \\ x_2 - 1 \end{bmatrix} \\
&= x_1 + x_2 - 1\frac{1}{2} \left(2(x_1 - 1)(x_2 - 1)\right) \\
&= (x_1 - 1)(x_2 - 1) + x_1 + x_2 - 1. \\
&= x_1 x_2
\end{aligned}$$

The original function evaluated at (1.1,1.1) is 1.21. The first order approximation around $(1.1, 1.1)$ is 1.2, but the second order approximation again exactly represents the function!

5. Jupyter Notebook Setup

Note: Please feel free to ask for help from the TAs during office hours and homework parties to ensure your conda environment works as it should

Conda Download

If you already, have anaconda, skip to the Conda Environment header. The goal of this problem is to confirm that you are proficient with the software environment, which you will need to complete the class.

(a) To get started, download anaconda from this link: `https://www.anaconda.com/distribution/`

(b) Download the command line installer

(c) Run the downloaded script

(d) Follow the instructions on the terminal

(e) Quit the terminal and restart it — this refreshes the environment and lets your terminal see Conda

Conda Environment
To create the environment we will be using in this class, run the following command:
`conda env create -f ee127_***.yml`, where `***` corresponds to your system (mac, linux, or windows).

For windows users, you may need to download Visual Studio. If for any reason, the environment creation fails, please refer to this link for instructions on how to manually install the correct libraries.

Then, run `conda activate ee127`. **You will need to run this command every time you start doing any code work for this class**

Please refer to the corresponding jupyter notebook for the rest of the question: `intro_to_jupyter.ipynb`