ECE 805 – Machine Learning

Tutorial 1, Friday, March 19, 2021

Norms:

- **Definition:** Let (X, F) be a linear vector space. A real-valued function is called a norm (and is denoted by || . ||) if the following properties hold:
  (i) $||x|| \geq 0$ and $||x||=0 \Rightarrow x=0_x$ for each $x \in X$
  (ii) $||\alpha \cdot x|| = |a| \cdot ||x||$ for each $x \in X$ and $\alpha \in F$
  (iii) $||x + y|| \leq ||x|| + ||y||$ for each $x, y \in X$ (triangle inequality)

- **Examples of norms:**
  1. $X = \Re^n$ (X is a vector with n elements)
     1-norm: $\|x\|_1 = \sum_{i=1}^{n}|x_i|$
     2-norm or Euclidean norm: $\|x\|_2 = \sqrt[2]{\sum_{i=1}^{n}|x_i|^2}$
     ∞-norm: $\|x\|_\infty = \max_{1\leq i\leq n}|x_i|$
     p-norm: $\|x\|_p = \sqrt[p]{\sum_{i=1}^{n}|x_i|^p}$

  2. X = C [0, T] (X is a continuous function in the given interval [0, T])
     1-norm: $\|x\|_1 = \int_0^T |x(t)|\, dt$
     2-norm or Euclidean norm: $\|x\|_2 = \sqrt[2]{\int_0^T |x(t)|^2\, dt}$
     ∞-norm: $\|x\|_\infty = \max_{t\in[0,T]}|x(t)|$
     p-norm: $\|x\|_p = \sqrt[p]{\int_0^T |x(t)|^p\, dt}$

  3. $X \in \Re^{n \times n}$ (Matrix norms)
     1-norm: $\|X\|_1 = \max_{1\leq j\leq n}\sum_{i=1}^{n}|x_{ij}|$
     2-norm or Euclidean norm: $\|X\|_2 = \sqrt[2]{\lambda_{max}\,(A^T A)}$
     ∞-norm: $\|X\|_\infty = \max_{1\leq i\leq n}\sum_{j=1}^{n}|x_{ij}|$

Gradient Method – Steepest Descent Method: $\begin{bmatrix} x_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \end{bmatrix} - \lambda_k^* \begin{bmatrix} \nabla F\,(x_k) \end{bmatrix}$

Newton's Method: $\begin{bmatrix} x_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \end{bmatrix} - H(x_k)^{-1} \begin{bmatrix} \nabla F\,(x_k) \end{bmatrix}$

Exercises:

1. (a)

$\|A_1\|_1 = \max\{1,3\} = 3$ $\qquad$ $\|A_1\|_\infty = \max\{3,1\} = 3$

$\|A_1\|_2 = \sqrt[2]{\lambda_{max}(A_1^T A_1)}$

$A_1^T A_1 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$

$\lambda I - A_1^T A_1 = \begin{bmatrix} \lambda - 1 & -2 \\ -2 & \lambda - 5 \end{bmatrix} \rightarrow \det(\lambda I - A_1^T A_1) = (\lambda - 1)(\lambda - 5) - 4$

$= \lambda^2 - 6\lambda + 5 - 4 = \lambda^2 - 6\lambda + 1$

$\lambda_{1,2} = \dfrac{6 \pm \sqrt{36 - 4}}{2} = \dfrac{6 \pm \sqrt{32}}{2} = \dfrac{6 \pm 2\sqrt{8}}{2} = 3 \pm \sqrt{8}$

$\rightarrow \lambda_1 = 3 + \sqrt{8} = 5.828$ $\qquad$ $\lambda_2 = 3 - \sqrt{8} = 0.172$

$\rightarrow \lambda_{max} = max\{\lambda_1, \lambda_2\} = \lambda_1 = 5.828$

$\rightarrow \|A_1\|_2 = \sqrt[2]{5.828} = 2.414$

(b)

$\|A_2\|_1 = \max\{4,6\} = 6$ $\qquad$ $\|A_2\|_\infty = \max\{3,7\} = 7$

$\|A_2\|_2 = \sqrt[2]{\lambda_{max}(A_2^T A_2)}$

$A_2^T A_2 = \begin{bmatrix} 1 & 3 \\ 2 & -4 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} = \begin{bmatrix} 10 & -10 \\ -10 & 20 \end{bmatrix}$

$\lambda I - A_2^T A_2 = \begin{bmatrix} \lambda - 10 & 10 \\ 10 & \lambda - 20 \end{bmatrix} \rightarrow \det(\lambda I - A_2^T A_2) = (\lambda - 10)(\lambda - 20) - 100$

$= \lambda^2 - 30\lambda + 200 - 100 = \lambda^2 - 30\lambda + 100$

$\lambda_{1,2} = \dfrac{30 \pm \sqrt{30^2 - 400}}{2} = 15 \pm 5\sqrt{5}$

$\rightarrow \lambda_1 = 15 + 5\sqrt{5} = 26.18$ $\qquad$ $\lambda_2 = 15 - 5\sqrt{5} = 3.82$

$\rightarrow \lambda_{max} = max\{\lambda_1, \lambda_2\} = \lambda_1 = 26.18$

$\rightarrow \|A_2\|_2 = \sqrt[2]{26.18} = 5.117$

2. (a) $f(x) = x_1^2 + x_2^2 - x_1$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 - 1 \\ 2x_2 \end{bmatrix} = 0 \rightarrow \begin{matrix} 2x_1 - 1 = 0 \rightarrow x_1 = \frac{1}{2} \\ 2x_2 = 0 \rightarrow x_2 = 0 \end{matrix}$$

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \rightarrow \det(\nabla^2 f) = 4 - 0 = 4 > 0$$

$$\nabla^2 f \; is\ a\ positive\ definite \; \rightarrow \left(\frac{1}{2}, 0\right) \; is\ a\ minimum$$

(b) $f(x) = x_1^4 + x_1 x_2 + \frac{1}{2} x_2^2$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 4x_1^3 + x_2 \\ x_1 + x_2 \end{bmatrix} = 0 \rightarrow \begin{matrix} 4x_1^3 + x_2 = 0 \rightarrow 4x_1^3 - x_1 = x_1(4x_1^2 - 1) = 0 \\ x_1 + x_2 = 0 \rightarrow x_2 = -x_1 \end{matrix}$$

$$x_1 = 0 \; \rightarrow x_2 = 0$$

$$x_1 = \pm\frac{1}{2} \rightarrow x_2 = \mp\frac{1}{2}$$

$$Points\colon (0,0), \left(\frac{1}{2}, -\frac{1}{2}\right), \left(-\frac{1}{2}, \frac{1}{2}\right)$$

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 12x_1^2 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\nabla^2 f|_{(0,0)} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \rightarrow \mathrm{det} = 0 - 1 = -1 < 0 \; \rightarrow \text{negative definite}$$

$$\rightarrow (0,0) \; is\ not\ a\ minimum$$

$$\nabla^2 f|_{\left(\frac{1}{2}, -\frac{1}{2}\right)} = \nabla^2 f|_{\left(-\frac{1}{2}, \frac{1}{2}\right)} = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix} \rightarrow \mathrm{det} = 3 - 1 = 2 > 0 \; \rightarrow \text{positive definite}$$

$$\rightarrow \left(\frac{1}{2}, -\frac{1}{2}\right) \; and \; \left(-\frac{1}{2}, \frac{1}{2}\right) \; are\ minimum\ points$$

(c) $f(x) = x_1^2 + 2x_2^2 + 4x_1 + 4x_2$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 + 4 \\ 4x_2 + 4 \end{bmatrix} = 0 \rightarrow \begin{matrix} 2x_1 + 4 = 0 \\ 4x_2 + 4 = 0 \end{matrix} \rightarrow \begin{matrix} x_1 = -2 \\ x_2 = -1 \end{matrix} \rightarrow (-2, -1)$$

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix} \rightarrow \det(\nabla^2 f) = 8 - 0 = 8 > 0$$

$\nabla^2 f$ *is a positive definite* $\rightarrow (-2,-1)$ *is a minimum*

(d) $f(x) = 4x_1^2 - 2x_2^4 + \frac{1}{3}x_1^6 + x_1x_2 + \frac{1}{4}x_2^2$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 8x_1 - 8x_1^3 + 2x_1^5 + x_2 \\ x_1 + \frac{1}{2}x_2 \end{bmatrix} = 0 \rightarrow x_1 + \frac{1}{2}x_2 = 0 \rightarrow x_2 = -2x_1$$

$\rightarrow 8x_1 - 8x_1^3 + 2x_1^5 + x_2 = 0 \rightarrow 8x_1 - 8x_1^3 + 2x_1^5 - 2x_1 = 6x_1 - 8x_1^3 + 2x_1^5 = 0$

$\rightarrow 2x_1(x_1^4 - 4x_1^2 + 3) = 2x_1(x_1^2 - 3)(x_1^2 - 1) = 0$

$\rightarrow x_1 = 0 \rightarrow x_2 = 0$

$\rightarrow x_1 = \pm\sqrt{3} \rightarrow x_2 = \mp 2\sqrt{3}$

$\rightarrow x_1 = \pm 1 \rightarrow x_2 = \mp 2$

Points: $(0,0), (\sqrt{3}, -2\sqrt{3}), (-\sqrt{3}, 2\sqrt{3}), (1,-2), (-1,2)$

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 8 - 24x_1^2 + 10x_1^4 & 1 \\ 1 & \frac{1}{2} \end{bmatrix}$$

$\nabla^2 f|_{(0,0)} = \begin{bmatrix} 8 & 1 \\ 1 & 1/2 \end{bmatrix} \rightarrow$ det $= 4 - 1 = 3 > 0 \rightarrow$ positive definite $\rightarrow (0,0)$ *is a minimum*

$\nabla^2 f|_{(1,-2)} = \nabla^2 f|_{(-1,2)} = \begin{bmatrix} -6 & 1 \\ 1 & \frac{1}{2} \end{bmatrix} \rightarrow$ det $= -3 - 1 = -4 < 0 \rightarrow$ negative definite

$\rightarrow (1,-2)$ *and* $(-1,2)$ *are not minimum points*

$\nabla^2 f|_{(\sqrt{3},-2\sqrt{3})} = \nabla^2 f|_{(-\sqrt{3},2\sqrt{3})} = \begin{bmatrix} 26 & 1 \\ 1 & \frac{1}{2} \end{bmatrix} \rightarrow$ det $= 13 - 1 = 12 < 0 \rightarrow$ positive definite

$\rightarrow (\sqrt{3}, -2\sqrt{3})$ *and* $(-\sqrt{3}, 2\sqrt{3})$ *are minimum points*

3. Find the point on the line 3x +2y = 5 in two-dimensional space closest to the origin when distance is measured by each of the following three norms:
   a. The 1-norm
   b. The 2-norm
   c. The ∞-norm

Solution:

$$3\mathrm{x} + 2y = 5 \Longrightarrow x = a \Longrightarrow y = \frac{5-3a}{2}$$

$$Any\ point\ on\ the\ line\ 3x + 2y = 5\ can\ be\ represented\ by\ z = \begin{bmatrix} a \\ \frac{5-3a}{2} \end{bmatrix}$$

We are going to use the following equations given on the first page, as we consider z to be a vector with 2 elements and represents any point on the line $3x + 2y = 5$.

z = $\Re^2$ (z is a vector with 2 elements)
1-norm: $\|z\|_1 = \sum_{i=1}^{n}\|z_i\|$
2-norm or Euclidean norm: $\|z\|_2 = \sqrt[2]{\sum_{i=1}^{n}\|z_i\|^2}$
∞-norm: $\|z\|_\infty = \max_{1\le i\le n}\|z_i\|$

<u>1-norm:</u>

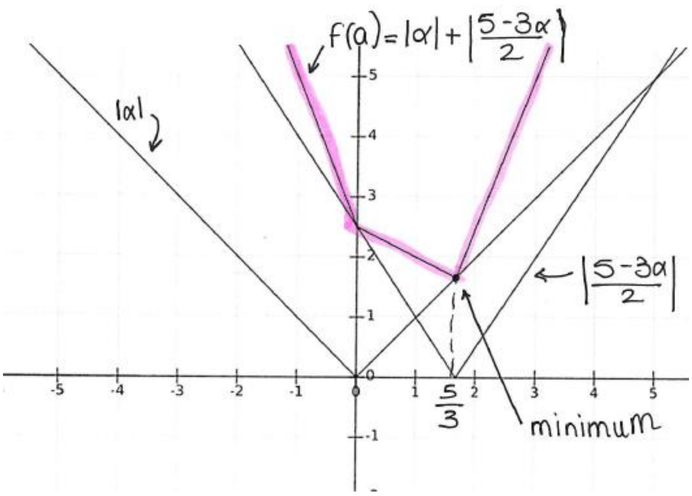
Distance in 1-norm of z from origin is $\|z_1\| = |a| + \left|\frac{5-3a}{2}\right|$

We want to find α that minimizes $f(a) = |a| + \left|\frac{5-3a}{2}\right|$

Note: This is a non-smooth optimization problem, so we cannot take derivatives.

We can obtain the minimum by sketching $f(a)$ and it occurs at $a = \frac{5}{3}$.

$\|z_1\|$ is minimized at $z^* = \begin{bmatrix} 5/3 \\ 0 \end{bmatrix}$



2-norm:

(smooth problem, we can take derivatives)

Distance in 2-norm of z from origin is $\|z_2\| = \sqrt[2]{a^2 + \left(\frac{5-3a}{2}\right)^2} \Rightarrow F(a) = a^2 + \left(\frac{5-3a}{2}\right)^2$

$$\frac{\partial F}{\partial a} = 0 \Rightarrow 2a + 2\left(\frac{-3}{2}\right)\left(\frac{5-3a}{2}\right) = 0 \Rightarrow 2a - \frac{15}{2} + \frac{9a}{2} = 0 \Rightarrow \boxed{a = \frac{15}{13}}$$

$$\frac{\partial^2 F}{\partial a^2} = \frac{13}{2} > 0 \ \Rightarrow z = \begin{bmatrix} 15/13 \\ 10/13 \end{bmatrix} \Rightarrow \|z_2\| \textit{ is minimize at } z = \begin{bmatrix} 15/13 \\ 10/13 \end{bmatrix}.$$

∞-norm:

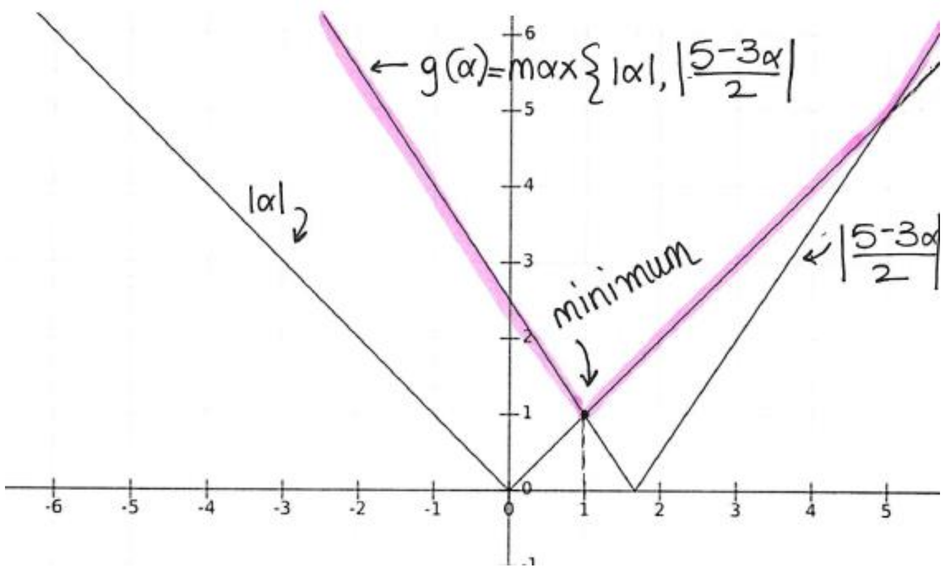
Again, it is a non-smooth optimization problem, so by sketching $g(a)$ we can obtain the minimum.

$$\|z_\infty\| = \max\left\{|a|, \left|\frac{5-3a}{2}\right|\right\}$$

$$\textit{Find } a \textit{ that minimizes } g(a) = \max\left\{|a|, \left|\frac{5-3a}{2}\right|\right\}$$

$$g(a) \textit{ is minimized at } a^* \textit{ where } a^* = \frac{5-3a}{2} \Rightarrow \boxed{a^* = 1} \Rightarrow z = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$\|z_\infty\|$ is minimized at $z^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$



4. (a) Let $x = [x_1 \ x_2 \ x_3]^T$ and $F(x) = x_1x_2^2 + x_2x_3^2 + x_3^3$. Compute the Gradient and Hessian of F and find all values x^* for which $\nabla F(x^*) = 0$.
Is the Hessian singular or nonsingular at these values?

(b) Using the first order necessary conditions ($\nabla F(x^*) = 0$) find a minimum point of the function:
$F(x_1, x_2, x_3) = 2x_1^2 + x_1x_2 + x_2^2 + x_2x_3 + x_3^2 - 6x_1 - 7x_2 - 8x_3 + 9.$
Verify that this point is a minimum by checking the second order sufficiency conditions.

Solution:

(a)

$$Gradient = \nabla \mathrm{F} = \begin{bmatrix} \dfrac{\partial F}{\partial x_1} \\ \dfrac{\partial F}{\partial x_2} \\ \dfrac{\partial F}{\partial x_3} \end{bmatrix} = \begin{bmatrix} x_2^2 \\ x_3^2 + 2x_1x_2 \\ 2x_2x_3 + 3x_3^2 \end{bmatrix}$$

$$Hessian = \nabla^2 F = \begin{bmatrix} \frac{d^2F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \frac{\partial^2 F}{\partial x_1 \partial x_3} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{d^2F}{\partial x_2^2} & \frac{\partial^2 F}{\partial x_2 \partial x_3} \\ \frac{\partial^2 F}{\partial x_3 \partial x_1} & \frac{\partial^2 F}{\partial x_3 \partial x_2} & \frac{d^2F}{\partial x_3^2} \end{bmatrix} = \begin{bmatrix} 0 & 2x_2 & 0 \\ 2x_2 & 2x_1 & 2x_3 \\ 0 & 2x_3 & 2x_2 + 6x_3 \end{bmatrix}$$

$$\nabla F(x^*) = 0 \Rightarrow \left.\begin{matrix} x_2^2 = 0 \\ 2x_1x_2 + x_3^2 = 0 \\ 2x_2x_3 + 3x_3^2 = 0 \end{matrix}\right\} \Rightarrow \begin{matrix} x_1^* = a \\ x_2^* = 0 \\ x_3^* = 0 \end{matrix} \Rightarrow x^* = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix}, \alpha \in \Re$$

$$\nabla^2 F(x^*) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2a & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \det\left(\nabla^2 F(x^*)\right) = 0 \Rightarrow Hessian\ is\ singular$$

[A matrix is singular if the determinant of the matrix is equal to zero.]

(b)

$$Gradient = \nabla \mathrm{F}\ (\mathrm{x}) = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \frac{\partial F}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 4x_1 + x_2 - 6 \\ x_1 + 2x_2 + x_3 - 7 \\ x_2 + 2x_3 \end{bmatrix} \Rightarrow \nabla \mathrm{F}\ (x^*) = 0 \Rightarrow \begin{matrix} 4x_1^* + x_2^* - 6 = 0 \\ x_1^* + 2x_2^* + x_3^* - 7 = 0 \\ x_2^* + 2x_3^* - 8 = 0 \end{matrix}$$

Solve the equations:

$$\begin{aligned} &4x_1^* + x_2^* - 6 = 0 && (1) \\ &x_1^* + 2x_2^* + x_3^* - 7 = 0 && (2) \\ &x_2^* + 2x_3^* - 8 = 0 && (3) \end{aligned}$$

$$\left.\begin{array}{l} 4x_1 + x_2 = 6 \;\; (1) \\ x_2 = 8 - 2x_3 \;\; (3) \end{array}\right\} \Rightarrow\; 4x_1 = 6 - x_2 \Rightarrow (3) \longrightarrow (1) \Rightarrow 4x_1 = 6 - 8 + 2x_3 \;\Rightarrow\; 4x_1 = 2x_3 - 2$$

$$\Rightarrow x_1 = \frac{1}{2}(x_3 - 1) \quad (4)$$

$$(4) \rightarrow (2) \Rightarrow \frac{1}{2}(x_3 - 1) + 16 - 4x_3 + x_3 - 7 = 0 \Rightarrow x_3 - 1 + 32 - 8x_3 + 2x_3 - 14 = 0$$

$$\Rightarrow -5x_3 + 17 = 0 \Rightarrow \boxed{x_3 = \frac{17}{5}} \quad (5)$$

$$(5) \rightarrow (3) \Rightarrow x_2 = 8 - 2\left(\frac{17}{5}\right) \;\Rightarrow\; \boxed{x_2 = \frac{6}{5}} \quad (6)$$

$$(5) \rightarrow (4) \Rightarrow\; x_1 = \frac{1}{2}\left(\frac{17}{5} - 1\right) = \frac{12}{5} \times \frac{1}{2} \Rightarrow \boxed{x_1 = \frac{6}{5}} \;\Rightarrow \boxed{x^* = \begin{bmatrix} 6/5 \\ 6/5 \\ 17/5 \end{bmatrix}}$$

$$Hessian = \nabla^2 F(x) = \begin{bmatrix} \frac{d^2F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \frac{\partial^2 F}{\partial x_1 \partial x_3} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{d^2F}{\partial x_2^2} & \frac{\partial^2 F}{\partial x_2 \partial x_3} \\ \frac{\partial^2 F}{\partial x_3 \partial x_1} & \frac{\partial^2 F}{\partial x_3 \partial x_2} & \frac{d^2F}{\partial x_3^2} \end{bmatrix} = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow \nabla^2 F(x^*) = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Eigenvalues (calculated using Matlab) are: $\begin{array}{l} \lambda_1 = 0.8299 \\ \lambda_2 = 2.6889 \\ \lambda_3 = 4.4812 \end{array}$

All eigenvalues are positive, $\lambda_i(A) > 0 \Rightarrow \nabla^2 F(x^*) > 0,\; positive\ definite\ \Rightarrow x^*\ is\ a\ minimum$

5. Write a script to implement the Gradient method – Steepest Descent Method for minimizing the function:

   $F(x_1, x_2) = e^{x_1}(4x_1^2 + 2x_2^2 + 4x_1x_2 + 2x_2 + 1).$

   Let your initial estimate be something close to the origin. Choose the step-size λ to be a constant. Run a few simulations with different values of λ to see what happens as you vary the step-size from a small to a large value.

Solution:

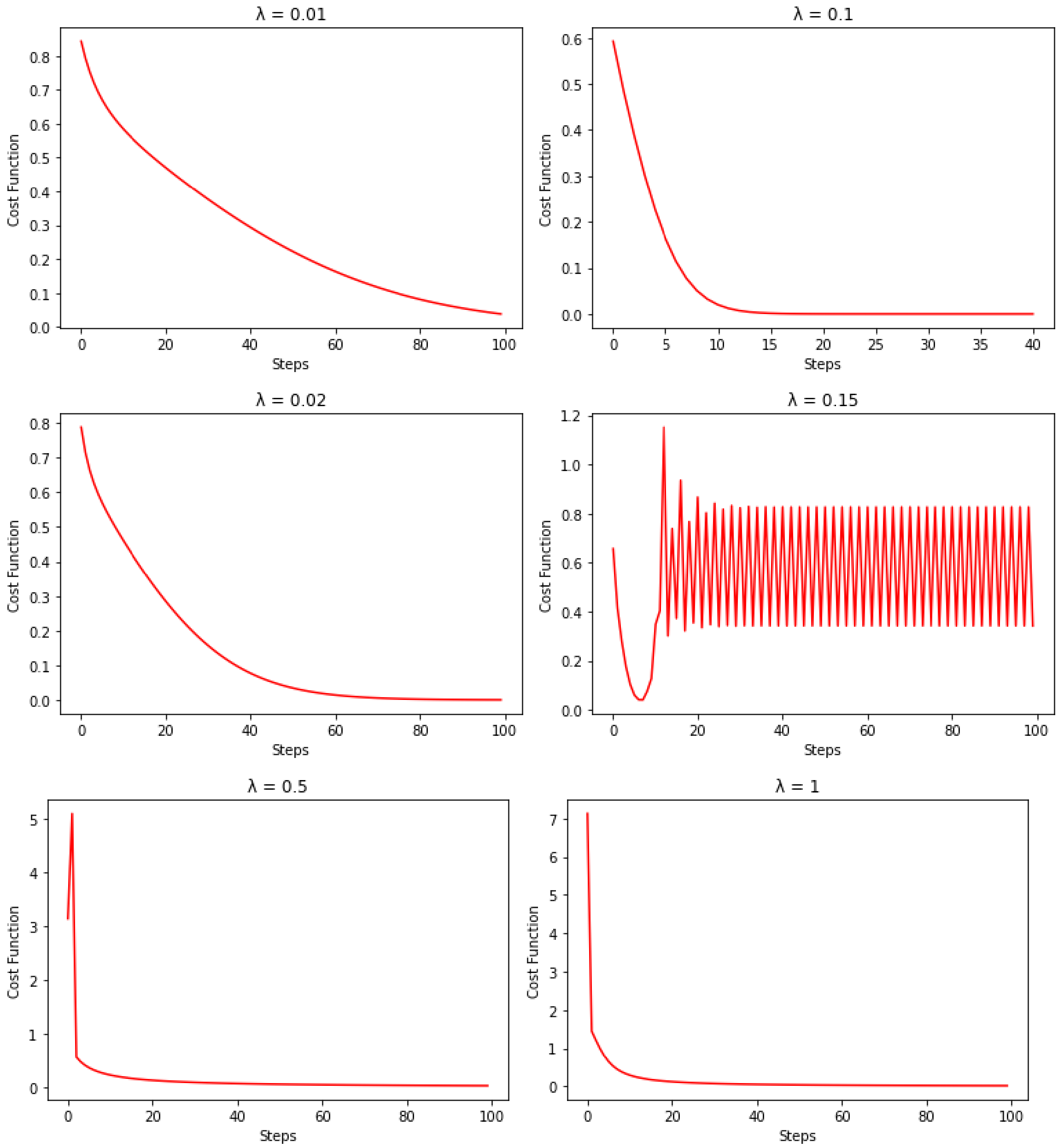
$$\nabla F(x_1, x_2) = \begin{bmatrix} \frac{\partial \mathrm{F}}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \end{bmatrix} = \begin{bmatrix} e^{x^1}(4x_1^2 + 2x_2^2 + 4x_1x_2 + 6x_2 + 8x_1 + 1) \\ e^{x_1}(4x_2 + 4x_1 + 2) \end{bmatrix}$$

Gradient Method – Steepest Descent Method: $\begin{bmatrix} x_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \end{bmatrix} - \lambda_k^* \begin{bmatrix} \nabla F\ (x_k) \end{bmatrix}$

```
import numpy as np
import pandas as pd
import math
import matplotlib.pyplot as plt

def my_function(x):
   z = math.exp(x[0]) * (4*x[0]**2 + 2*x[1]**2 + 4*x[0]*x[1]+2*x[1]+1)
   return z
def gradient1(x):
   z = math.exp(x[0]) * (4*x[0]**2 + 2*x[1]**2 + 4*x[0]*x[1] + 8*x[0] + 6*x[1] + 1)
   return z
def gradient2(x):
   z = math.exp(x[0]) * (4*x[0] + 4*x[1] + 2)
   return z
l = 1
tolerance = 0.000000001
maxsteps = 100
cost_function = np.zeros((1,maxsteps))
x = np.zeros((1,maxsteps))
y = np.zeros((1,maxsteps))
x_initial1 = 0.1
x_initial2 = -0.1
count = 0
w = [x_initial1, x_initial2]
x[0,count] = w[0]
y[0,count] = w[1]
cost_function[count] = my_function(w)
while ((count<maxsteps)&(abs(my_function(w))>tolerance)):
   d = [gradient1(w), gradient2(w)]
   w[0] = w[0] - l * d[0]
```

```
    w[1] = w[1] - l * d[1]
    x[0, count] = w[0]
    y[0, count] = w[1]
    cost_function[0, count] = my_function(w)
    count = count + 1
plt.plot([r for r in range(count)], cost_function[0,:count], 'r-')
plt.xlabel('Steps')
plt.ylabel('Cost Function')
plt.title('λ = 1')
plt.show()
```



It can be seen that the rate of convergence keeps increasing as λ increases until about λ = 0.1 in which oscillations are not observed. If we choose λ = 0.15 we observe some oscillations.