1.1.1 Gradient and Hessian of f_1 [Report]

Derive the analytical expressions for the gradient and Hessian (with respect to x) of the following scalar function:

$$f_1(x) = \phi(Ax) \tag{1}$$

Where:

- $f_1 : \mathbb{R}^n \to \mathbb{R}$
- $x \in \mathbb{R}^{n\times 1}$
- $A \in \mathbb{R}^{m\times n}$
- $\phi : \mathbb{R}^m \to \mathbb{R}$

פתרון :

הפונקציה שנתונה : $f_1(x) = \varphi(A\cdot x)$

נסמן : $u = A\cdot x$

ולכן נקבל ש $du = A\cdot dx$

ידועים לנו הגרדיאנט וההסיאן של הפונקציה φ ולכן אנו יודעים את הגרדיאנט וההסיאן של הפונקציה f_1 כתלות במשתנה u נשתמש בהגדרה החיצונית כפי שראינו בהרצאה למציאת הגרדיאנט וההסיאן של הפונקציה f_1 כתלות במשתנה x :

נרצה להגיע לביטוי מהצורה $df = \langle \nabla f | dx \rangle$ כדי לחלץ את הגרדיאנט

נסתכל ב :

$$df = \langle \nabla\varphi | du \rangle = \langle \nabla\varphi | Adx \rangle = (Adx)^T \nabla\varphi = dx^T A^T \nabla\varphi = \langle \mathrm{A}^\mathrm{T} \nabla\varphi | dx \rangle$$

גראדיינט $\boxed{\rightarrow \nabla f_1(x) = \mathrm{A}^\mathrm{T}\nabla\varphi}$

נחשב עכשיו את הנגזרת מסדר 2 :

נגדיר :

$$g = \nabla f = \mathrm{A}^\mathrm{T}\nabla\varphi$$

נרצה להגיע לביטוי מהצורה כדי לחלץ את ההיסיאן של f_1

$$dg = Hdx$$

שיטה 1

$$dg = \mathrm{d}\nabla f_1(x) = d\mathrm{A}^\mathrm{T}\nabla\varphi = \mathrm{A}^\mathrm{T}\mathrm{d}\nabla\varphi = \mathrm{A}^\mathrm{T}\mathrm{d}\nabla f(u) = A^T H(u) du = A^T HAdx$$

$$\boxed{\rightarrow \nabla^2 f_1(\mathrm{x}) = \mathrm{A}^T H A}$$

שיטה 2

$$dg = \mathrm{d}\nabla f_1(x) = d\mathrm{A}^{\mathrm{T}}\nabla\varphi = \mathrm{A}^{\mathrm{T}}\mathrm{d}\nabla\varphi = \{chain\ rule\} = \mathrm{A}^{\mathrm{T}}\nabla\varphi^2 Adx = A^T H_\varphi Adx$$

$$\boxed{\rightarrow \nabla^2 f_1(\mathrm{x}) = \mathrm{A}^T H_\varphi A}$$

1.1.2 Gradient and Hessian of f_2 [Report]

Derive the analytical expressions for the gradient and Hessian of the following scalar function:

$$f_2(x) = h(\phi(x)) \tag{2}$$

Where:

- $f_2 : \mathbb{R}^n \rightarrow \mathbb{R}$
- $x \in \mathbb{R}^{n\times 1}$
- $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$
- $h : \mathbb{R} \rightarrow \mathbb{R}$

You can assume that the gradient and Hessian of ϕ and the first and second derivatives of h are given.

פתרון:

נתונה:

$$f_2(x) = h(\varphi(x))$$

נשתמש בכלל השרשרת:

$$\frac{\partial f}{\partial x_i} = h'(\varphi(x))\cdot\frac{\partial\varphi}{\partial x_i} \quad \Rightarrow \quad \nabla f(x) = h'(\varphi(x))\cdot\nabla\varphi$$

$$\left.\begin{aligned} \frac{\partial^2 f}{\partial {x_i}^2} &= h''(\varphi(x)) \cdot \left(\frac{\partial \varphi}{\partial x_i}\right)^2 + h'(\varphi(x)) \cdot \frac{\partial^2 \varphi}{\partial {x_i}^2} \\ \frac{\partial^2 f}{\partial x_i \partial x_j} &= h''(\varphi(x)) \cdot \frac{\partial \varphi}{\partial x_i} \cdot \frac{\partial \varphi}{\partial x_j} + h'(\varphi(x)) \cdot \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \quad \nabla^2 f(x) = h''(\varphi(x)) \cdot \nabla\varphi \cdot \nabla\varphi^T + h'(\varphi(x)) \cdot \nabla^2 \varphi$$

1.1.3 Gradient and Hessian of ϕ [Report]

Derive the analytical expressions for the gradient and Hessian of the following scalar function $\phi : \mathbb{R}^3 \to \mathbb{R}$:

$$\phi\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \sin(x_1 x_2 x_3) \tag{3}$$

$$\nabla\phi = \left(\frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2}, \frac{\partial \phi}{\partial x_3}\right) = \left(\frac{\partial(\sin(x_1x_2x_3))}{\partial x_1}, \frac{\partial(\sin(x_1x_2x_3))}{\partial x_2}, \frac{\partial(\sin(x_1x_2x_3))}{\partial x_3}\right)$$

$$\Rightarrow \text{grad} = \nabla\phi = (x_2x_3\cos(x_1x_2x_3) \;,\; x_1x_3\cos(x_1x_2x_3) \;,\; x_1x_2\cos(x_1x_2x_3))$$

$$\nabla^2\phi = \begin{pmatrix} \frac{\partial \phi}{\partial x_1 \partial x_1} & \frac{\partial \phi}{\partial x_1 \partial x_2} & \frac{\partial \phi}{\partial x_1 \partial x_3} \\ \frac{\partial \phi}{\partial x_2 \partial x_1} & \frac{\partial \phi}{\partial x_2 \partial x_2} & \frac{\partial \phi}{\partial x_2 \partial x_3} \\ \frac{\partial \phi}{\partial x_3 \partial x_1} & \frac{\partial \phi}{\partial x_3 \partial x_2} & \frac{\partial \phi}{\partial x_3 \partial x_3} \end{pmatrix} =$$

$$H = \nabla^2\phi =$$

$$\begin{pmatrix} -x_2^2x_3^2\sin(x_1x_2x_3) & x_3\cos(x_1x_2x_3) - x_2x_1x_3^2\sin(x_1x_2x_3) & x_2\cos(x_1x_2x_3) - x_1x_2^2x_3\sin(x_1x_2x_3) \\ x_3\cos(x_1x_2x_3) - x_2x_1x_3^2\sin(x_1x_2x_3) & -x_1^2x_3^2\sin(x_1x_2x_3) & x_1\cos(x_1x_2x_3) - x_1^2x_2x_3\sin(x_1x_2x_3) \\ x_2\cos(x_1x_2x_3) - x_1x_2^2x_3\sin(x_1x_2x_3) & x_1\cos(x_1x_2x_3) - x_1^2x_2x_3\sin(x_1x_2x_3) & -x_1^2x_2^2\sin(x_1x_2x_3) \end{pmatrix}$$

1.1.4 First and Second Derivatives of h [Report]

Derive the analytical expressions for first and second derivatives of the following function $h : \mathbb{R} \to \mathbb{R}$:

$$h(x) = \sqrt{(1+x^2)} \tag{4}$$

פה נתונה פונקציה במשתנה יחיד (מימד 1)

נגדיר את :

$$u = x^2 \to du = 2xdx$$

נשתמש בכלל השרשרת:

$$dh = h'(x)dx = g'(u)du \quad ; \quad when\ g(u) = \sqrt{(1+u)}$$

$$h'(x)dx = \frac{1}{2}(1+u)^{-\frac{1}{2}}du = \frac{1}{2}(1+x^2)^{-\frac{1}{2}} \cdot 2xdx = \frac{x}{\sqrt{1+x^2}}dx$$

$$\to\ h'(x) = \frac{x}{\sqrt{1+x^2}}$$

$$\underset{\therefore}{\Rightarrow} h'(x) = \frac{x}{\sqrt{1+x^2}}$$

$$\nabla^2 h = h''(x) = \frac{\sqrt{1+x^2} - x \cdot \frac{x}{\sqrt{1+x^2}}}{1+x^2} = \frac{1+x^2-x^2}{\sqrt{1+x^2}(1+x^2)}$$

$$\underset{\therefore}{\Rightarrow} h''(x) = \frac{1}{\sqrt{1+x^2}(1+x^2)} = \frac{1}{(1+x^2)^{\frac{3}{2}}}$$

1.3 Comparison [Code and Report]

In this section, we will compare the difference between the analytical and numerical evaluation of the gradient and Hessian of functions f_1 and f_2.
Draw a random vector $x \in \mathbb{R}^3$ and a random matrix $A \in \mathbb{R}^{3\times3}$, and evaluate the infinity norm of the difference between the numerical and analytical gradient and Hessian of f_1 and f_2 at the point x (given by $||\nabla_x^{\text{analytical}} f_{1,2}(x) - \nabla_x^{\text{numerical}} f_{1,2}(x)||_\infty$ for the gradient, and $||\nabla_x^{2\,\text{analytical}} f_{1,2}(x) - \nabla_x^{2\,\text{numerical}} f_{1,2}(x)||_\infty$ for the Hessian) for values of ε between 2^0 to 2^{-60}. Plot your results on four logarithmic-scaled line-charts (two for f_1 (gradient and Hessian) and two for f_2 (gradient and Hessian)) where the x-axis represents ε and the y-axis represents $||\cdot||_\infty$. **Clarification (25/3/2021)**: Only the y-axis should be logarithmic scaled. The x-axis should be the absolute value of the exponent (0, 1, 2, ..., 60).

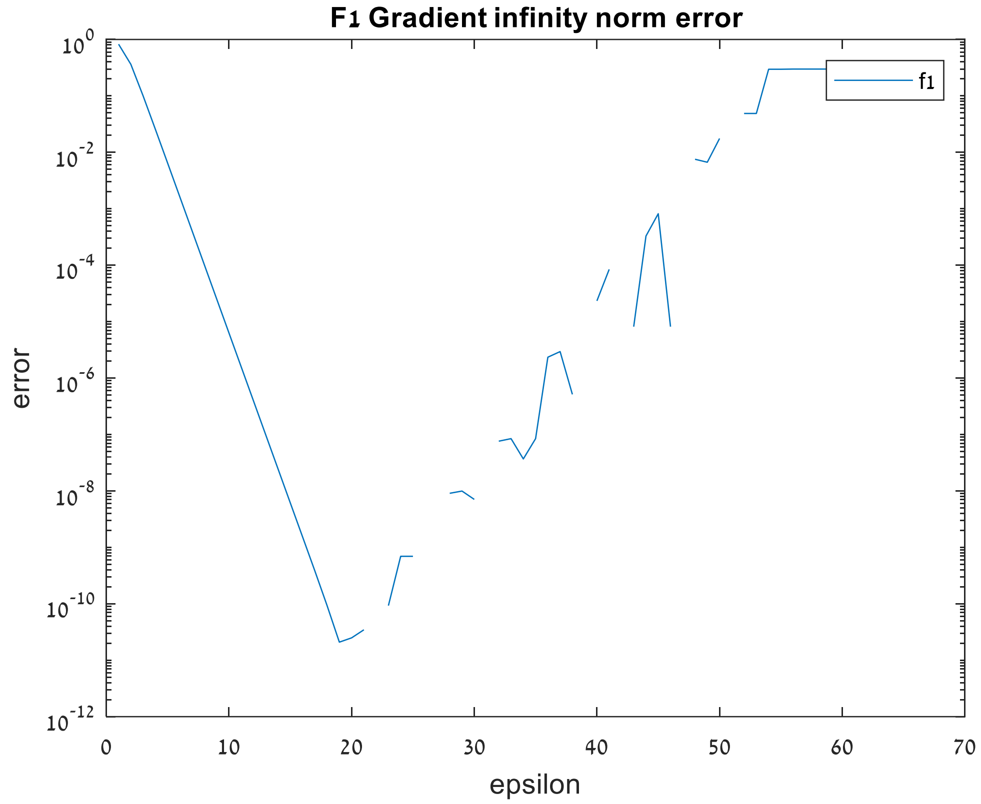
פתרון :

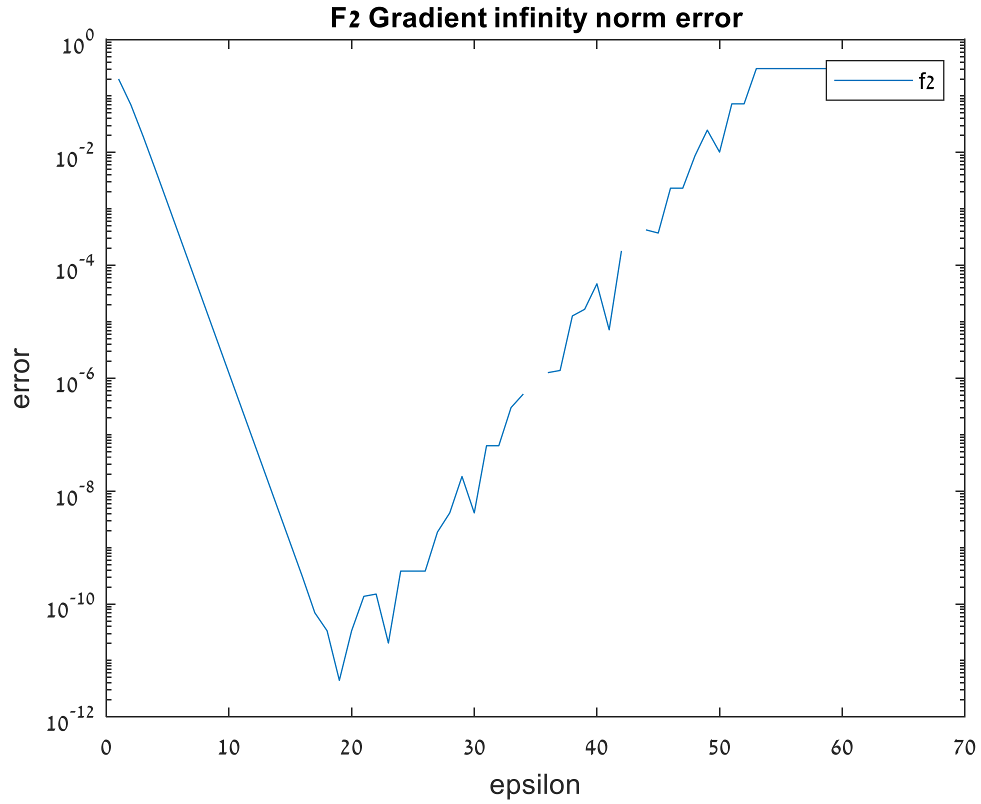
ציר X מתאר $\varepsilon = [2^{-60}, 2^0]$ לפי הערך המוחלט של החזקות $[0:60]$

בכל הגרפים אנו רואים שהשגיאה קטנה מאוד בין הפונקציה בחישוב אנליטי לנומרי כאשר אפסילון שלנו הוא קטן, אבל השגיאה הולכת ונהיית גדולה בסקלה לוגרתמית כאשר האפסילון הולך וגדל(לכיוון שמאל לציר האופקי בגרפים פה), ולכן הקירוב הנומרי הוא קירוב טוב יחסית לקירוב האנליטי עבור אפסילונים קטנים יחסית.

ניתן לראות כי השגיאה שואפת לאפס בסביבה השמאלית לנקודה $\varepsilon \approx 2^{-19}$ בכל הגרפים(19 בציר האופקי), ייתכן שזה בגלל שהפונקציות עצמן הן בקירוב נומרי ולכן התוצאות הן נומריות

הגרפים בעמודים הבאים


F1 Gradient infinity norm error
error
epsilon
f1
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
0
10
20
30
40
50
60
70


F2 Gradient infinity norm error
error
epsilon
f2
10^{0}
10^{-2}
10^{-4}
10^{-6}
10^{-8}
10^{-10}
10^{-12}
0
10
20
30
40
50
60
70


F1 Hessian infinity norm error
f1
error
epsilon
10^2
10^0
10^-2
10^-4
10^-6
10^-8
10^-10
10^-12
0
10
20
30
40
50
60
70


F2 Hessian infinity norm error
f2
error
epsilon
10^0
10^-2
10^-4
10^-6
10^-8
10^-10
10^-12
0
10
20
30
40
50
60
70