HW1
236330 – Introduction to Optimization and Deep Learning

Nerya Hadad
Yair Rada

1. **analytical differentiation:**
   gradient and hessian calculation:

Task 1:

$f_1(x) = \phi(Ax)$

Define:

$f(x) = \varphi(u)$ Such that $u = Ax$

$du = A \cdot dx$

Gradient:

$df = \nabla^T \varphi(u) \cdot du$

$df = \nabla^T \varphi(u) \cdot A \cdot dx$

$df = (A^T \cdot \nabla\varphi(u))T \cdot dx$

Using the external definition $df = g(x)^T \cdot dx$:

$\boldsymbol{g_{f1} = A^T \nabla \phi(Ax)}$

Hessian:

$dg = A^T \cdot d\nabla\varphi(Ax)$

$dg = A^T \cdot H_\varphi(Ax)Adx$

Using the external definition $dg = H(x) \cdot dx$:

$\boldsymbol{H_\varphi(x) = A^T \cdot H_\varphi(Ax)A}$

Task 2:

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

Gradient:

$$df = \nabla h^T(\varphi(x)) \cdot d\varphi = \nabla h^T(\varphi(x)) \cdot \nabla\varphi^T \cdot dx$$

Using the external definition $df = g(x)^T \cdot dx$:

$$\boldsymbol{g(x) = \nabla\varphi(x) \cdot \nabla h(\varphi(x))}$$

Hessian:

$$\begin{aligned} dg &= d\nabla\varphi(x) \cdot \nabla h(\varphi(x)) + \nabla\varphi(x) \cdot d\nabla h(\varphi(x)) \\ &= H_\varphi(x) \cdot dx \cdot \nabla h(\varphi(x)) + \nabla\varphi(x) \cdot H_h(\varphi(x)) \cdot d\varphi \end{aligned}$$

$d\varphi = \nabla\varphi^T \cdot dx$ =>

$$dg = H_\varphi(x) \cdot dx \cdot \nabla h(\varphi(x)) + \nabla\varphi(x) \cdot H_h(\varphi(x)) \cdot \nabla\varphi^T \cdot dx$$

h: R->R is a scalar function, hence it's derivative in the point $\varphi(x)$ is a scalar too =>

$$\begin{aligned} dg &= H_\varphi(x) \cdot \nabla h(\varphi(x)) \cdot dx + \nabla\varphi(x) \cdot H_h(\varphi(x)) \cdot \nabla\varphi^T \cdot dx \\ &= \left[H_\varphi(x) \cdot \nabla h(\varphi(x)) + \nabla\varphi(x) \cdot H_h(\varphi(x)) \cdot \nabla\varphi^T\right] \cdot dx \end{aligned}$$

Using the external definition $dg = H(x) \cdot dx$:

$$\boldsymbol{H_\varphi(x) = H_\varphi(x) \cdot \nabla h(\varphi(x)) + \nabla\varphi(x) \cdot H_h(\varphi(x)) \cdot \nabla\varphi^T}$$

Task 3:

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

$$\bar{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$df(\bar{x}) = \nabla \sin(\mathrm{prod}(\bar{x}))^t dx$$

$$\nabla \sin(\mathrm{prod}(\bar{x})) = \begin{pmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{pmatrix} \cos(\mathrm{prod}(\bar{x}))$$

Using the external definition $df = g(x)^t dx$:

$$g(x) = \begin{pmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{pmatrix} \cos(\mathrm{prod}(\bar{x})) \quad \Big| \quad \mathrm{prod}(\bar{x}) = x_1 x_2 x_3$$

$$\nabla g(x) = \begin{pmatrix} 0 & x_3 & x_2 \\ x_3 & 0 & x_1 \\ x_2 & x_1 & 0 \end{pmatrix} \cdot \cos(\mathrm{prod}(\bar{x})) - \begin{pmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{pmatrix} \begin{pmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{pmatrix}^t \cdot \sin(\mathrm{prod}(\bar{x}))$$

since $\nabla g(x)$ is symetric($\nabla g(x)^t = \nabla g(x)$) :

$dg = \nabla g(x)^t dx = \nabla g(x) dx$

Using the external definition $dg = H(x)\, dx$:

$$H(x) = \begin{pmatrix} 0 & x_3 & x_2 \\ x_3 & 0 & x_1 \\ x_2 & x_1 & 0 \end{pmatrix} \cos(\mathrm{prod}(\bar{x}) - \begin{pmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{pmatrix} \begin{pmatrix} x_2 x_3 \\ x_1 x_3 \\ x_1 x_2 \end{pmatrix}^t \cdot \sin(\mathrm{prod}(\bar{x}))$$

$$h(x) = \exp(x)$$

$dh = \nabla(e^x)^t\, dx = (e^x)^t dx$

Using the external definition $dh = g(x)^t dx$:

$g(x) = e^x = h(x)$

$dg = dh = (e^x)^t dx = e^x dx$

Using the external definition $dg = H(x)\, dx$:

$H(x) = e^x$

/*TODO*/
(continue of Task 3):

5. Implement functions to evaluate the analytical expressions you have derived for the value, gradient and hessian of the given instances of f1 and f2.
Please add a very short explanation of your implementation.

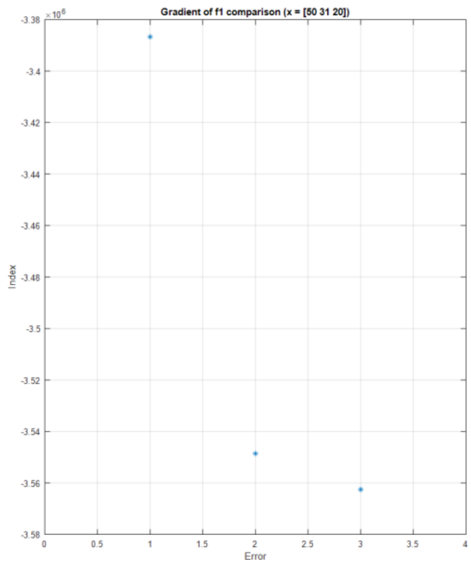
Task 4:

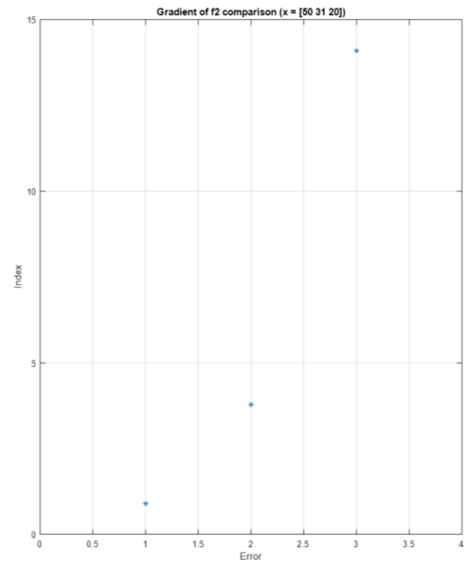
1. Implement functions to numerically calculate the gradient and hessian of f1 and f2, as instructed in the assignment (numerical differentiation).
Please add a very short explanation of your implementation.

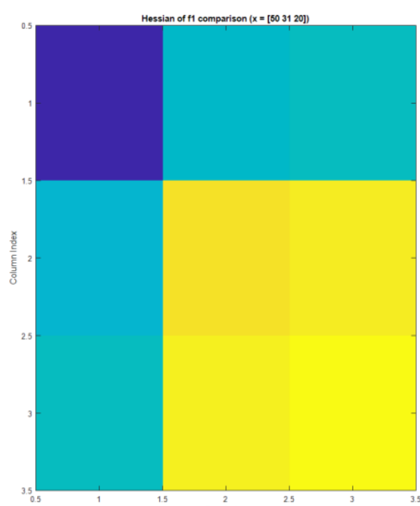
במשימה 5 – אולי להוסיף סקאלה לצבע, אבל לא קריטי כי רשמו שאפשר להשתמש ב imagesc
בנוסף במשימה 5 רשום:
make sure you explain the results of each plot
לא ברור אם הכוונה לרשום על הצירים ובכותרת מה זה כל גרף או גם לנתח את זה

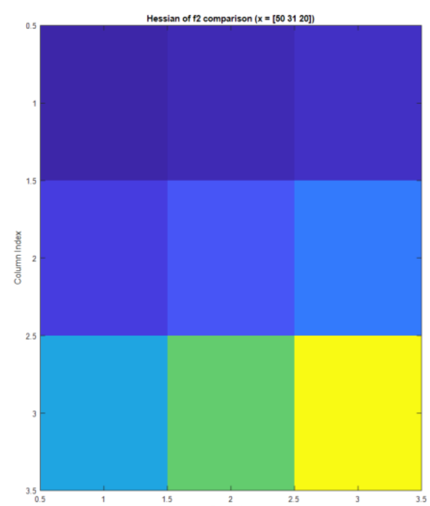
/*TODO*/

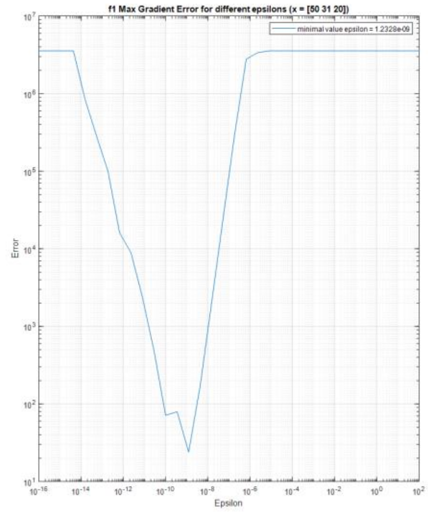
Task 5:

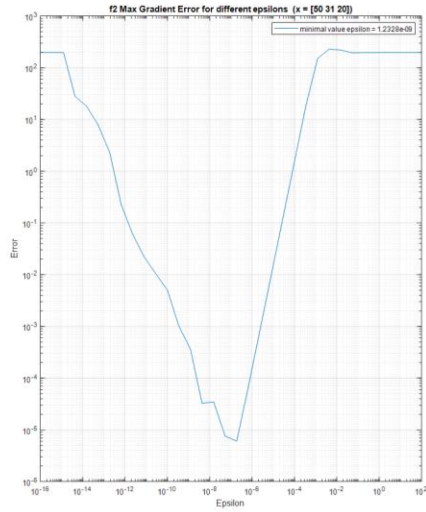

Gradient of f1 comparison (x = [50 31 20])
Index
Error

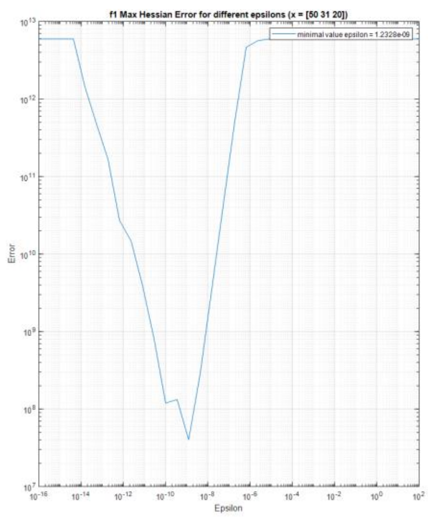

Gradient of f2 comparison (x = [50 31 20])
Index
Error

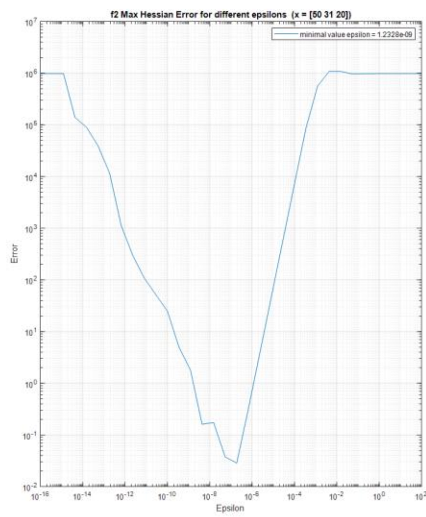

Hessian of f1 comparison (x = [50 31 20])
Column Index
Row Index


Hessian of f2 comparison (x = [50 31 20])
Column Index
Row Index


f1 Max Gradient Error for different epsilons (x = [50 31 20])
minimal value epsilon = 1.2328e-09
Error
Epsilon


f2 Max Gradient Error for different epsilons (x = [50 31 20])
minimal value epsilon = 1.2328e-09
Error
Epsilon


f1 Max Hessian Error for different epsilons (x = [50 31 20])
minimal value epsilon = 1.2328e-09
Error
Epsilon


f2 Max Hessian Error for different epsilons (x = [50 31 20])
minimal value epsilon = 1.2328e-09
Error
Epsilon