

236330 - Introduction to Optimization: Homework #1

April 17, 2020

Amit Rotner
123456789
Or Steiner
123456789

Analytical Differentiation:**Task 1:**

Find the Gradient and Hessian of:

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

Where,

$$\begin{aligned} f_1 &: \mathbb{R}^n \rightarrow \mathbb{R} \\ x &\in \mathbb{R}^n \\ A &\in \mathbb{R}^{m \times n} \\ \varphi &: \mathbb{R}^m \rightarrow \mathbb{R} \end{aligned}$$

Given $\nabla \varphi$ and $\nabla^2 \varphi$.

Solution:

We start by finding the gradient:

Denoting $u = Ax$. Hence $du = dAx$.

From the external definition of gradient, we know that,

$$d\varphi = \langle \nabla \varphi, du \rangle = \nabla \varphi^T du = \nabla \varphi^T A dx$$

Since $df = \nabla f^T dx$ and $d\varphi = df$ (because $f(x) = \varphi(Ax)$) we get that:

$$g = \nabla f = A^T \nabla \varphi(Ax)$$

Now for the Hessian,

$$dg = d\nabla f = dA^T \nabla \varphi(u) = A^T d\nabla \varphi(u)$$

From the external definition of Hessian we know that $d\nabla \varphi(u) = H du$.

$$A^T d\nabla \varphi(u) = A^T H(u) du = A^T H(Ax) A dx$$

Since $H = \nabla^2 \varphi$, it yields that:

$$H(x) = A^T \nabla^2 \varphi(Ax) A$$

Task 2:

Develop the Gradient and Hessian of:

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

Where,

$$\begin{aligned} f_2 : \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\in \mathbb{R}^n \\ \varphi : \mathbb{R}^m &\rightarrow \mathbb{R} \\ h : \mathbb{R} &\rightarrow \mathbb{R} \end{aligned}$$

Given $\nabla\varphi$, $\nabla^2\varphi$, h' , h'' .

Solution:

We start by finding the gradient:

Denoting $u = \varphi(x)$. From the external definition of the gradient, we know that:

$$du = \langle \nabla\varphi(x), dx \rangle$$

$$dh = h'(u) du$$

Combining these two together yields:

$$dh = h'(u) \langle \nabla\varphi(x), dx \rangle = \langle h'(u) \nabla\varphi(x), dx \rangle = \langle h'(\varphi(x)) \nabla\varphi(x), dx \rangle$$

Since $df = \nabla f^T dx$ we get that:

$$g = \nabla f = (h'(\varphi(x)) \nabla\varphi(x))^T \stackrel{(1)}{=} h'(\varphi(x)) \nabla\varphi(x)$$

(1) - Since $h'(\varphi(x))$ is a scalar and $\nabla\varphi(x) = \nabla\varphi^T(x)$.

Now for the Hessian,

$$\begin{aligned}
dg &= d\nabla f = d(h'(u) \nabla \varphi(x)) \\
&= dh'(u) \nabla \varphi(x) + h'(u) d\nabla \varphi(x) \\
&= h''(u) du \nabla \varphi(x) + h'(u) \nabla^2 \varphi(x) dx \\
&\stackrel{(1)}{=} h''(u) \nabla \varphi(x) du + h'(u) \nabla^2 \varphi(x) dx \\
&= h''(u) \cdot \nabla \varphi(x) \cdot \langle \nabla \varphi(x), dx \rangle + h'(u) \nabla^2 \varphi(x) dx \\
&= h''(u) \cdot \nabla \varphi(x) \cdot \nabla \varphi^T(x) dx + h'(u) \nabla^2 \varphi(x) dx \\
&= (h''(u) \cdot \nabla \varphi(x) \cdot \nabla \varphi^T(x) + h'(u) \nabla^2 \varphi(x)) \cdot dx
\end{aligned}$$

(1) - Since $du = \langle \nabla \varphi(x), dx \rangle$ is a scalar.

From the external definition of Hessian we know that $dg = Hdx$, hence:

$$H = h''(u) \cdot \nabla \varphi(x) \cdot \nabla \varphi^T(x) + h'(u) \nabla^2 \varphi(x)$$

Task 3:

1. Let:

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

The gradient of φ is:

$$\nabla \varphi(x_1, x_2, x_3) = \begin{pmatrix} \frac{\partial \varphi}{\partial x_1} \\ \frac{\partial \varphi}{\partial x_2} \\ \frac{\partial \varphi}{\partial x_3} \end{pmatrix} = \begin{pmatrix} x_2 x_3 \cos(x_1 x_2 x_3) \\ x_1 x_3 \cos(x_1 x_2 x_3) \\ x_1 x_2 \cos(x_1 x_2 x_3) \end{pmatrix}$$

The Hessian of φ is:

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

Denoting $u = x_1 x_2 x_3$, we get:

$$H = \begin{pmatrix} -x_2^2 x_3^2 \sin(u) & x_3 (\cos(u) - u \sin(u)) & x_2 (\cos(u) - u \sin(u)) \\ x_3 (\cos(u) - u \sin(u)) & -x_1^2 x_3^2 \sin(u) & x_1 (\cos(u) - u \sin(u)) \\ x_2 (\cos(u) - u \sin(u)) & x_1 (\cos(u) - u \sin(u)) & -x_1^2 x_2^2 \sin(u) \end{pmatrix}$$

2. Let:

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

It is known that:

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

Explanation about our implementation:

Using phi() function we have calculated the value, gradient and hessian of $\varphi(Ax)$.

Calling phi() in f1(), and then calculating gradient and hessian of $f_1(x)$ according to task 1.

Using h() and phi() functions we have calculated the value, gradient and hessian of $h(x)$ and $\varphi(x)$.

In f2(), calling h() and phi() and calculating gradient and hessian of $f_2(x)$ according to task 2.

We have used nargout to avoid redundant calculations.

Task 4:

The required function to calculate numerically the gradient and hessian is passed as a parameter to the numdiff function as well as required ϵ . We calculate f's value at $(x - \epsilon)$, $(x + \epsilon)$, and calculate the numerical gradient and hessian according to the formula presented in the assignment.

Task 5:

Using the magic matrix:

$$A = \begin{pmatrix} 8 & 1 & 6 \\ 3 & 5 & 7 \\ 4 & 9 & 2 \end{pmatrix}$$

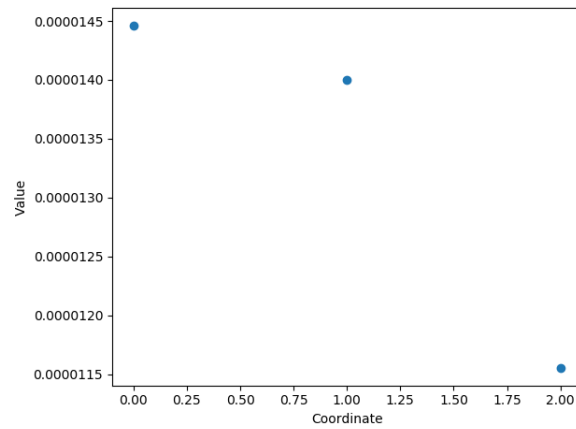
And a random vector:

$$x = \begin{pmatrix} 0.06892 \\ 0.24130 \\ 0.46690 \end{pmatrix}$$

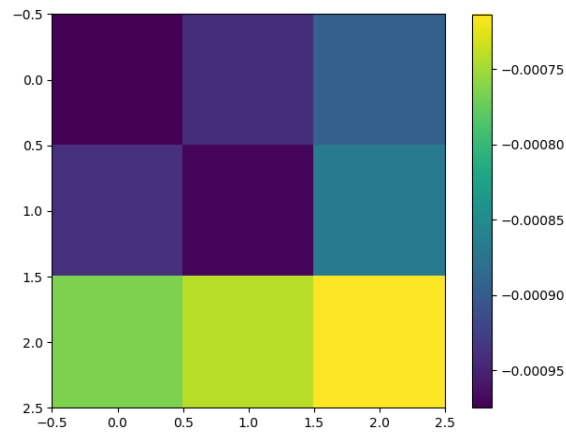
We got the following results:

The function f_1 :

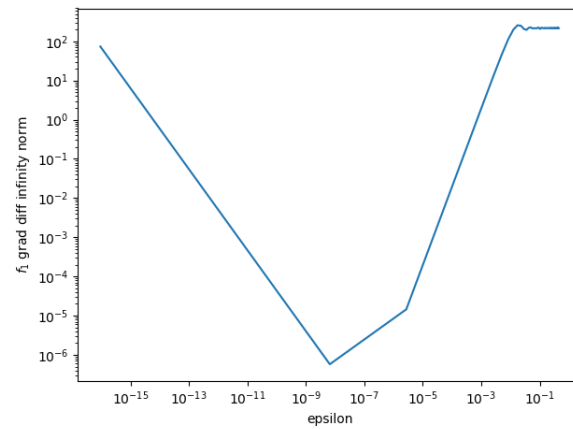
The difference between the numerical and analytical gradient:



The difference between the numerical and analytical Hessian:

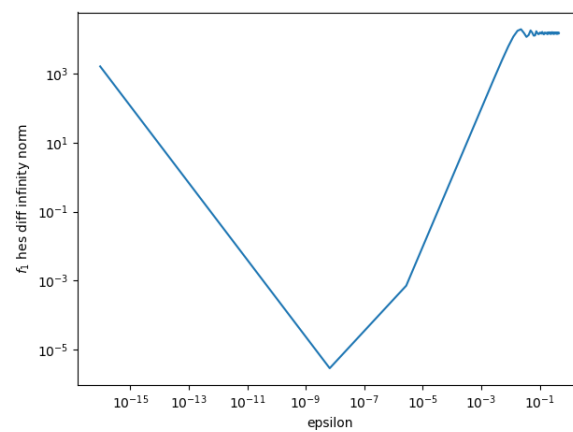


The infinity norm of the error between the numerical and analytical gradient as a function of ϵ :



The best error rate is 5.7478×10^{-7} , for $\epsilon = 6.60302 \times 10^{-9}$.

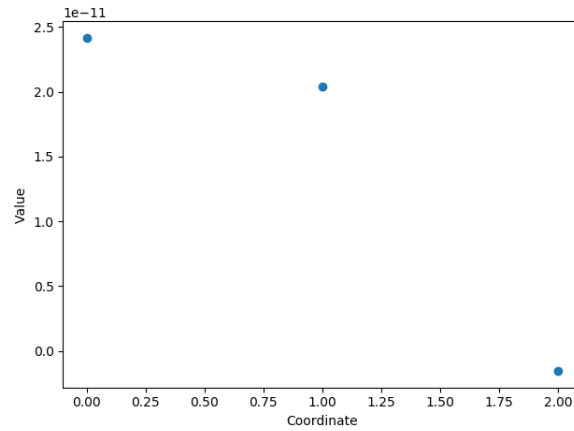
The infinity norm of the error between the numerical and analytical Hessian as a function of ϵ :



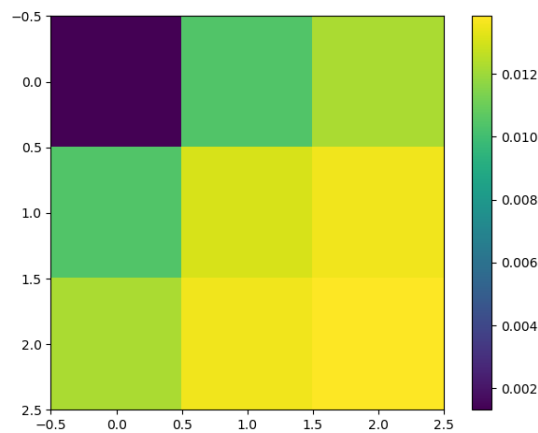
The best error rate is 2.8968×10^{-6} , for $\epsilon = 6.60302 \times 10^{-9}$.

The function f_2 :

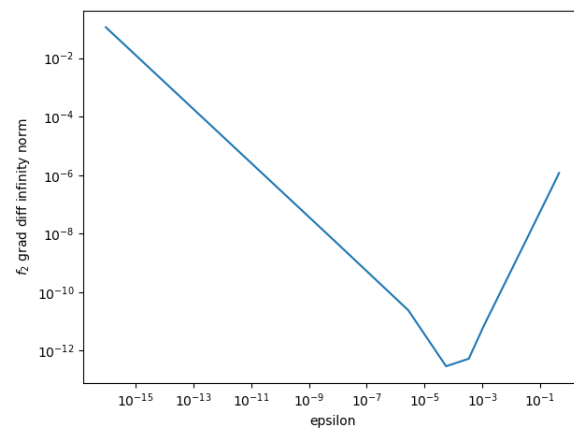
The difference between the numerical and analytical gradient:



The difference between the numerical and analytical Hessian:

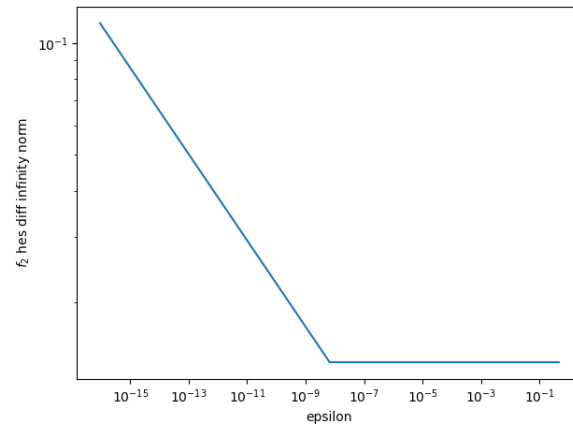


The infinity norm of the error between the numerical and analytical gradient as a function of ϵ :



The best error rate is 2.9484×10^{-13} , for $\epsilon = 5.5524 \times 10^{-5}$.

The infinity norm of the error between the numerical and analytical Hessian as a function of ϵ :



The best error rate is 0.01383, for $\epsilon = 2.7304 \times 10^{-6}$.