Introduction to Numerical Optimization

Assignment 1

March 2022

General Instructions:

- Submissions are in pairs.
- You are free to use either MATLAB or Python for any coding exercise.
- Sections labeled by [**Report**] should be included in your final report, and sections labeled with [**Code**] include a short coding task.
- Sections with no associated label include explanatory text only.
- More details about what should be included in your final report and your submission file are available in sections 2 and 3.
- Due date: Wed, April 6, 2022 (23:59).

1 Analytical and Numerical Differentiation

In this section we will derive and evaluate the analytic expressions for the gradient and Hessian of scalar functions, and compare our findings with the numerical differentiation results. First, watch the video **"Homework on analytical and numerical computation of gradient and Hessian"**. Note: There is a mistake in the video at time 8:00, where numerical differentiation is described. Optimal ϵ should be equal to $\|x\|_\infty \sqrt[3]{\text{Machine\_Epsilon}}$

1.1 Analytical Differentiation

Derive and evaluate analytical expressions for the gradient and Hessian of given multivariate scalar functions.

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 vector $x \in \mathbb{R}^n$, matrix $A \in \mathbb{R}^{m \times n}$, function $\phi : \mathbb{R}^m \to \mathbb{R}$

We assume that the gradient and Hessian of ϕ are given.

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 vector $x \in \mathbb{R}^n$, functions $\phi : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$

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

1.1.3 Gradient and Hessian of ϕ [Report]

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

$$\phi\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \cos^2\left(x_1 x_2^2 x_3\right) \tag{3}$$

1.1.4 First and Second Derivatives of h [Report]

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

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

1.1.5 Analytical Evaluation [Code]

1. Write a function that receives as input a vector $x \in \mathbb{R}^3$ and a square matrix $A \in \mathbb{R}^{3\times 3}$, and returns the evaluated value, gradient and Hessian at the point x of the function f_1 from section 1.1.1, when ϕ is given by section 1.1.3. For evaluating the gradient and Hessian of f_1, use the expressions you have derived in section 1.1.1 and section 1.1.3.

2. Write a function that receives as input a vector $x \in \mathbb{R}^3$, and returns the value, gradient and Hessian at the point x of the function f_2 from section 1.1.2, when ϕ is given by section 1.1.3 and when h is given by section 1.1.4. For evaluating the gradient and Hessian of f_2, use the expressions you have derived in sections 1.1.1, 1.1.3 and 1.1.4.

Please try to avoid unnecessary computations - that is, design your functions interface to accept additional flags that determine whether the function's value, gradient or Hessian should be evaluated or not.

1.2 Numerical Differentiation

In this part we will evaluate numerically the gradient and Hessian of the scalar functions described in section 1.1 using the method of finite differences. For more information about numerical differentiation, please see the following two Wikipedia pages: Numerical Differentiation and Finite Difference.

Computationally, numerical derivatives are much more expensive, than analytical. One of the reasons to use them is to debug correctness of the analytical derivatives.

1.2.1 Numerical Gradient

Given a scalar function $f : \mathbb{R}^n \to \mathbb{R}$, we can approximate its gradient at a point x by evaluating the ratio df to dx of each component of the gradient ∇f, between two infinitesimally close points to x. This method is called *Finite Differences* using *Central Differences*, and is given as follows:

$$\nabla f(x) \approx \begin{bmatrix} \frac{f(x+\varepsilon e_1)-f(x-\varepsilon e_1)}{2\varepsilon} \\ \frac{f(x+\varepsilon e_2)-f(x-\varepsilon e_2)}{2\varepsilon} \\ \vdots \\ \frac{f(x+\varepsilon e_n)-f(x-\varepsilon e_n)}{2\varepsilon} \end{bmatrix} \tag{5}$$

Where $x \in \mathbb{R}^n$, e_i is the ith standard basis vector of $\mathbb{R}^n$, and $\varepsilon \in \mathbb{R}$ is infinitesimally small, such that equation (5) becomes an equality when $\varepsilon \longrightarrow 0$.

1.2.2 Numerical Hessian

We can also estimate the Hessian matrix of a scalar function $f : \mathbb{R}^n \to \mathbb{R}$ at a point x by repetitive evaluation of f at nearby points, entry by entry, as explained in Finite Differences for the Hessian. However, if the analytical expression for the gradient ∇f is known, we can exploit it to estimate the Hessian $\nabla^2 f$ at a point x more efficiently as follows:

$$\nabla^2 f^{(i)}(x) \approx \frac{\nabla f(x+\varepsilon e_i) - \nabla f(x-\varepsilon e_i)}{2\varepsilon} \tag{6}$$

Where $\nabla^2 f^{(i)}(x)$ is the ith column of $\nabla^2 f(x)$.

1.2.3 Numerical Evaluation of Gradient and Hessian [Code]

Write a function that receives as input a function reference $f(x, \ldots)$ (Python Example, MATLAB Example), a vector $x \in \mathbb{R}^n$, a scalar $\varepsilon \in \mathbb{R}$ and any other additional required argument (depends on your implementation), and returns back the **numerical** evaluation of the gradient and Hessian at the point x of the input function f, as explained in sections 1.2.1 and 1.2.2.

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, taken as a long vector of concatenated columns) 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 $||\cdot||_\infty$. Only the y-axis should be logarithmic scaled. The x-axis should be the absolute value of the exponent (0, 1, 2, ..., 60).

2 What Should Be Included In Your Written Report?

1. Detailed derivations of sections 1.1.1, 1.1.2, 1.1.3 and 1.1.4.
2. Chart plots of section 1.3. Please provide an explanation for the behavior of each chart, and specify for which ε you get an optimal result.

3 Assignment Submission

Please zip your report's PDF file along with the source code you have written in sections 1.1.5, 1.2.3 and 1.3 into a file named **hw1.zip**, and submit your file through Coursera.

Screen-recording video-report Explain all your mathematical derivations and the program code in detail using screen-recorded video with voice. Afterward run your code with screen recording and explain the results. You can create several videos and put them into common file hw1_videos.zip. If you do the homework in pair, each partner should prepare his own video-report and upload separately to Coursera.

You can perform screen-recording with Zoom, Free-Cam-8 `https://ispring-free-cam.software.informer.com/8.3/` or any other tool.

Anonymity of the submission In order to allow blind peer review, your report, program code and video-report should contain no personal information.

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 vector $x \in \mathbb{R}^n$, matrix $A \in \mathbb{R}^{m\times n}$, function $\phi : \mathbb{R}^m \to \mathbb{R}$

We assume that the gradient and Hessian of ϕ are given.

$$f_1(x) = \phi(\underbrace{Ax}_{u}) = \phi(u)$$

$$du = A\,dx \qquad (du = A(x+dx) - Ax = A\,dx)$$

$$df_1 = \nabla\phi^T du$$

$$\Rightarrow df_1 = \nabla\phi^T \underbrace{A\,dx}_{} = \nabla f_1^T dx$$

$$\Rightarrow \nabla f_1 = (\nabla\phi^T A)^T = A^T\nabla\phi \;//$$

$$d\nabla f_1 = H\,dx$$

$$d\nabla f_1 = A^T\nabla^2\phi\,du = (A^T\nabla^2\phi A)\,dx$$

$\Rightarrow$

$\nabla^2 f_1 = H = A^T \nabla^2 \phi A$ //

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 vector $x \in \mathbb{R}^n$, functions $\phi : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$

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

$df_2 = g^T(x)\, dx =$

$d\phi = \nabla\phi^T dx$

$df_2 = h'(\phi(x))\, d\phi = h'(\phi(x))\, \nabla\phi^T dx$

$\Rightarrow \quad g(x) = \nabla\phi\, h'(\phi(x)) = \nabla\phi\, h'$ //

$dg = H\, dx$ (seen at class)

$$dg = d\,\nabla\phi\, h'$$

(by using the product rule of differentials)

$$= \nabla^2\phi\, dx\, h' + \nabla\phi\, h''\, d\phi$$

$$= h'\,\nabla^2\phi\, dx + \nabla\phi\, h''\, \nabla\phi^T dx$$

$$= \left(h'\,\nabla^2\phi + \nabla\phi\, h'' \cdot \nabla\phi^T\right) dx$$

$$\Rightarrow H = \left(h'\,\nabla^2\phi + h''\,\nabla\phi \cdot \nabla\phi^T\right) \;//$$

1.1.3 Gradient and Hessian of ϕ [Report]

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

$$\phi\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \cos^2\left(\underbrace{x_1 x_2^2 x_3}_{u}\right) \tag{3}$$

$$d\phi = d(\cos^2(x_1 x_2^2 x_3))$$

$$= d(\cos^2(u))$$

$$= 2\cos(u) \cdot -\sin(u)\, du$$

$$du = \nabla(x_1 x_2^2 x_3)^T dx = (x_2^2 x_3, 2x_1 x_2 x_3, x_1 x_2^2)\, dx$$

$$\nabla u = (x_2^2 x_3, 2x_1 x_2 x_3, x_1 x_2^2)^T$$

$$\Rightarrow d\phi = -2\cos(u)\sin(u) \cdot (x_2^2 x_3, 2x_1 x_2 x_3, x_1 x_2^2)\, dx$$

$$\Rightarrow \nabla\phi = -\nabla u \cdot 2\cos(u) \cdot \sin(u)$$

by trigo identity

$$= -\nabla u \sin(2u) //$$

$d\nabla\phi = dg = H\,dX$ (seen at class)

by using the product rule of differentials

$$dg = -\left(\nabla^2 u\, dX \sin(2u) + \nabla u \cos(2u)\cdot 2du\right)$$

$$= -\left(\nabla^2 u\, dX\cdot \sin(2u) + \nabla u \cos(2u)\cdot 2\nabla u^T dX\right)$$

$$\left(-\sin(2u)\nabla^2 u^T - 2\cos(2u)\nabla u\, \nabla u^T\right) dX$$

$$H = -\sin(2u)\nabla^2 u - 2\cos(2u)\nabla u\,\nabla u^T \quad //$$

s.t. $\nabla^2 u = \nabla(x_2^2 X_3,\ 2X_1X_2X_3,\ X_1X_2^2)^T$

$$= \begin{pmatrix} 0, & 2X_2X_3, & X_2^2 \\ 2X_2X_3, & 2X_1X_3, & 2X_1X_2 \\ X_2^2, & 2X_1X_2, & 0 \end{pmatrix}$$

1.1.4 First and Second Derivatives of h [Report]

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

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

$$dh = d(1+\sin^2 x)^{\frac{1}{2}} = d(u^{\frac{1}{2}})$$

$$= \frac{1}{2} u^{-\frac{1}{2}} \cdot du$$

$$du = (2 \sin x \cdot \cos x) dx = \sin(2x)\, dx$$

$$\Rightarrow dh = \frac{1}{2} u^{-\frac{1}{2}} \cdot \sin 2x\, dx$$

$$= \frac{1}{2}(1+\sin^2 x)^{-\frac{1}{2}} \cdot \sin 2x\, dx$$

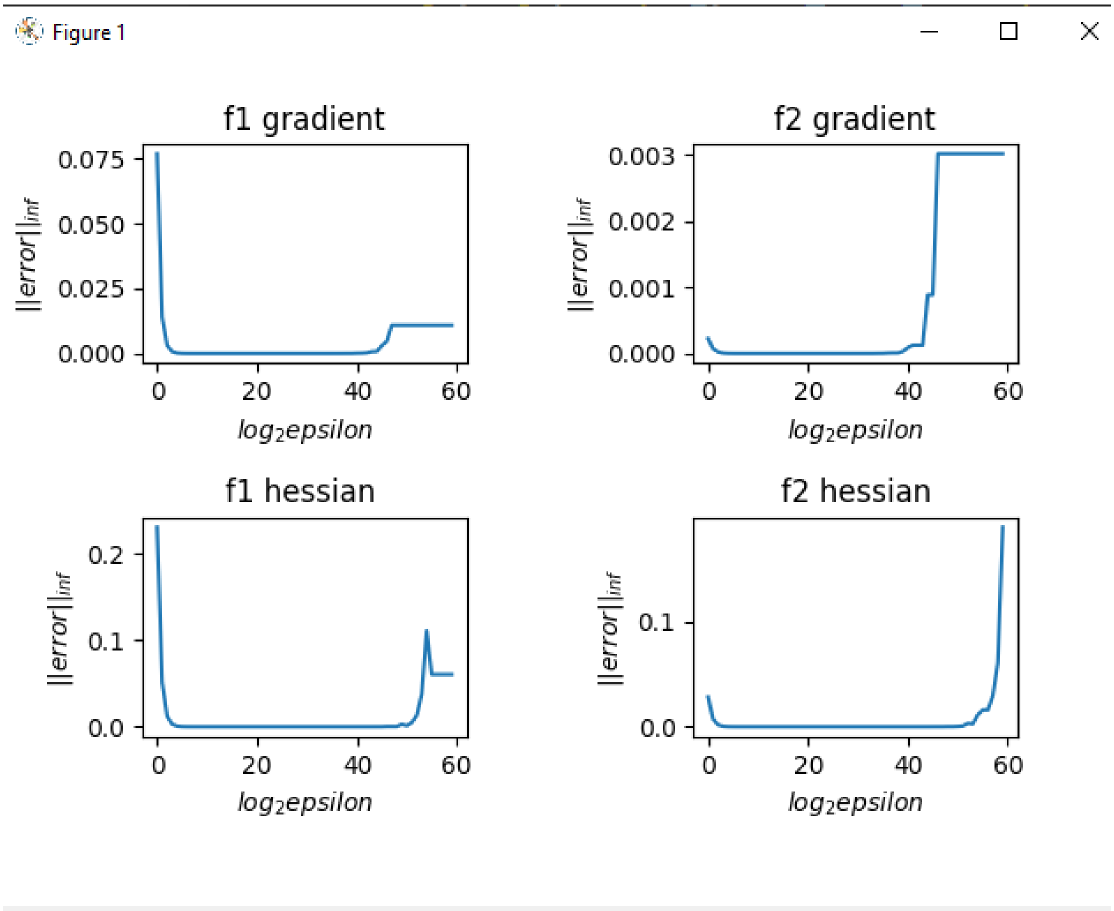
$$\Rightarrow h' = \frac{1}{2}(1+\sin^2 x)^{-\frac{1}{2}} \cdot (\sin 2x) \quad //$$

$$dh' = \Big(\frac{1}{2} \cdot -\frac{1}{2}(1+\sin^2 x)^{-\frac{3}{2}} \cdot \overbrace{2\sin x \cos x}^{\sin(2x)} \cdot \sin 2x$$

$$+ \frac{1}{2}(1+\sin^2 x)^{-\frac{1}{2}} \cos 2x \cdot 2\Big)\, dx$$

$$= \Big(-\frac{1}{4}(1+\sin^2 x)^{-\frac{3}{2}} \sin^2(2x) + (1+\sin^2 x)^{-\frac{1}{2}} \cdot \cos(2x)\Big)\, dx$$

$$h'' = -\frac{1}{4}(1+\sin^2(x))^{-\frac{3}{2}} \cdot \sin^2(2x) + (1+\sin^2 x)^{-\frac{1}{2}} \cos(2x)$$




```
best epsilon for f1 gradient 2^- 16
best epsilon for f2 hessian 2^- 14
best epsilon for f1 gradient 2^- 19
best epsilon for f2 hessian 2^- 19
```

אפשר לראות מגמת ירידה בשגיאה ככל שהאפסילון עולה תחילה בכל הגרפים שהאפסילון הכי טוב מתקבל באיזור 2^20 2^14 עבור אפסילון קטן מידי מתחילות להתקבל שגיאות נומריות אפשר לראות את זה בf1 בגרדיאנט החל מאפסילון 2^45כנל בגרדיאנט של f2 .

אפשר לראות שההסיאנים באופן דומה מתחילים לקבל שגיאות החל מאפסילון 2^50 .