

Numerical Optimization: General Review

Ignasi Cos – ignasi.cos@ub.edu

18th September 2023

Done

Abstract

This laboratory focuses on the unconstrained optimization of a function $f(x)$, and on understanding some of the basic conditions a minimum must satisfy. Indeed, a minimum x^* of a function $f(x)$ must satisfy that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite. Which are the implications from a geometrical standpoint? Let's analyse this with several examples:

1 One dimensional case

We begin with the one dimensional case. Assume that $x \in R$ and that

$$f(x) = x^3 - 2x + 2$$

You are asked to follow the next steps:

1. Plot this function within the range $x \in [-2, 2]$, for instance. For that purpose use the `matplotlib` from Python using the examples included within this document¹.
2. Compute analytically the points x^* that satisfy $f'(x) = 0$. Observe if the obtained result is congruent with the plot of the previous point.
3. We are now going to check which of the latter points x^* are minima (or a maxima). For that purpose let us perform a 2nd order Taylor expansion around point x^*

$$f(x^* + d) \approx f(x^*) + d f'(x^*) + \frac{1}{2} d^2 f''(x^*) \quad (1)$$

where $d \in R$ is the perturbation around x^* . Since we are dealing with a one dimensional function, $f''(x^*)$ is a real number which may be positive or negative.

Equation (1) approximates the function $f(x)$ around x^* (with a small value of d) using a quadratic function. The value of the second derivative will tell us whether the point x^* is a minimum or a maximum.

In order for x^* to be a minimum, you need $f''(x^*)$ to be positive. In other words, you need $f''(x^*)$ to be convex at that point. This can be expressed in another way: you need $d^2 f''(x^*) > 0$ for any $d \neq 0$. Despite the trivial of the one dimensional case, this latter check becomes very much relevant at higher dimensions.

¹You may find a gallery here: <http://matplotlib.org/gallery.html>.

4. You may also plot $f''(x)$ within the range $x \in [-2, 2]$. If $f''(x)$ is positive, the function may be approximated with a convex 2^{nd} order Taylor expansion at x . On the other hand, if $f''(x)$ is negative the function is concave at that point.

2 Two dimensional case

2.1 A simple two-dimensional function

We are now going to focus on simple two-dimensional functions, $\mathbf{x} \in R^2$, $x = (x_1, x_2)^T$ (vectors are expressed column-wise). Let us begin with the next quadratic expression

$$f(\mathbf{x}) = x_1^2 + x_2^2$$

Follow the next steps:

1. Plot this function. Please, note that this function has a minimum at $(x_1, x_2) = (0, 0)$.
2. Analytically compute the gradient of the function, $\nabla f(\mathbf{x})$, and compute the point \mathbf{x}^* at which $\nabla f(\mathbf{x}^*) = 0$.
3. Let $\mathbf{d} \in R^2$ be the perturbation around \mathbf{x}^* . The Taylor expansion, up to second order, of a function of several variables can be compactly expressed as

$$f(\mathbf{x}^* + \mathbf{d}) \approx f(\mathbf{x}^*) + \mathbf{d}^T \nabla f(\mathbf{x}^*) + \frac{1}{2} \mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} \quad (2)$$

Analyze the previous expression and be sure to understand the operations that are done at each of the terms.

Compute the Hessian matrix, $\nabla^2 f(\mathbf{x})$, at the point $\mathbf{x} = \mathbf{x}^*$. You should obtain

$$\nabla^2 f(\mathbf{x}^*) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

The latter matrix is giving us information about the shape of the quadratic approximation at $\mathbf{x} = \mathbf{x}^*$ in a similar way as has been done for the one dimensional case.

For the one-dimensional case it is easy to check if we have a minimum, $f''(x^*) > 0$, or a maximum, $f''(x^*) < 0$. For a higher dimensional problem we are sure that the quadratic approximation is convex and that we have a minimum if

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} > 0 \quad \mathbf{d} \neq \mathbf{0} \quad (3)$$

We have a maximum if

$$\mathbf{d}^T \nabla^2 f(\mathbf{x}^*) \mathbf{d} < 0 \quad \mathbf{d} \neq \mathbf{0} \quad (4)$$

The previous conditions can be verified by computing the eigenvalues of $\nabla^2 f(\mathbf{x}^*)$. If all eigenvalues are strictly positive, equation (3) is satisfied. If all eigenvalues are strictly negative, equation (4) is satisfied. For this example, which are the eigenvalues of the Hessian matrix? Do we have a minimum or a maximum at x^* ?

4. The question that may arise now is: what happens if some eigenvalues are positive and some negative? What happens if the eigenvalue is zero? For that issue you are asked to analyze the following functions:

$$f_A(\mathbf{x}) = -x_1^2 - x_2^2 \quad f_B(\mathbf{x}) = x_1^2 - x_2^2 \quad f_C(\mathbf{x}) = x_1^2$$

You are recommended to draw the contour plot of the previous functions. Observe the shape they have. Then answer the following questions:

- Perform a plot of the function. At which point \mathbf{x}^* is the gradient zero?
- At the points where the gradient is zero, what kind of information is giving us the Hessian matrix? Is this a minimum? A maximum? None of both? You may use the `eigvals` function of Python to compute the eigenvalues of the Hessian matrix (i.e. there is no need to compute them analytically).

In a similar way as for the one dimensional case, the eigenvalues of the Hessian $\nabla^2 f(\mathbf{x})$ gives us information about the local shape of the function $f(\mathbf{x})$ at point \mathbf{x} . This information will be used by numerical methods to accelerate descent to the optimal point we are looking for! In this lab we will focus, however, on the optimal points.

2.2 A two dimensional function with multiple minima

You are proposed to study the function that has been given in the lectures

$$f(x_1, x_2) = x_1^2 (4 - 2.1 x_1^2 + \frac{1}{3} x_1^4) + x_1 x_2 + x_2^2 (-4 + 4 x_2^2)$$

Follow these steps:

- Plot the previous function within the range $x_1 \in [-2, 2]$ and $x_2 \in [-1, 1]$ using, for instance, a step of e.g. 0.1. Be sure that the plot is correct: just look at the plot of the lectures and compare them with the result you obtain. Observe where the minima (and maxima) may be. There may be multiple minima and maxima!
- Analytically compute the gradient $\nabla f(\mathbf{x})$.
- Numerically compute an approximation of the points \mathbf{x}^* at which $\nabla f(\mathbf{x}^*) = 0$. For that issue
 - Evaluate $\|\nabla f(\mathbf{x})\|^2$ at the previous range using a step of e.g. 0.005 or smaller if you prefer (but not too small!). You may create a matrix that stores all the latter values to be able to analyze them in the next steps.
 - Using brute force, search for those points $\tilde{\mathbf{x}}$ within the previous range at which the value of $\|\nabla f(\mathbf{x})\|^2$ is strictly smaller than the value of its 8 neighbors². Our purpose here is to find the those points at which the gradient is small. We thus find all the “candidate” points that may be a minimum, a maximum or a saddle point!
 - Which are the values of $\tilde{\mathbf{x}}$ you have obtained? Which is the value of $\|\nabla f(\mathbf{x})\|^2$ at those points?

²For that issue just ignore those points that are at the border of the grid.

4. Compute the analytical expression of the Hessian of $f(x_1, x_2)$ and evaluate it at the values $\tilde{\mathbf{x}}$ you have found. What kind of information is giving you the Hessian? Does it correspond to a minimum (the Hessian is positive definite)? To a maximum (the Hessian is negative definite)? Or may be a saddle point? You may use the *eigvals* function of Python to compute the eigenvalues of the Hessian matrix (no need to compute them analytically). Take into account that there may be several minima, maxima and saddle points for the function you are analyzing.

Report

You are requested to deliver a report (PDF, MSWord or a cleaned python notebook), preferably in *pairs*. Comment the experiments you have performed in sections 1 and 2, as well as the results and plots obtained. Bear in mind that the professor has to be able to readily interpret your results.

If you want to include some parts of code, please include them along with your report. Do not include separate files.