A review of important concepts in optimization

Lluís Garrido – lluis.garrido@ub.edu

September 2016

Abstract

This laboratory is focused on unconstrained optimization of a function $f(x)$ and, in particular, on understanding some of the basic condition a minimum has to satisfy. Indeed, recall that a minimum x^* has to satisfy that $\nabla f(x^*) = 0$ and that $\nabla^2 f(x^*)$ is positive definite. What does this mean from a geometrical point of view? Let us analyze 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

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 document[1].

2. Compute analytically the points x^* that satisfy $f'(x) = 0$. Observe if the obtained result is congruent with the plot performed in the previous point.

3. We are now going to check which of the latter points x^* are a minimum (or a maximum). For that purpose let us perform a Taylor expansion around point x^*

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

   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) tells us that the function $f(x)$ can be approximated around x^* (with a small value of d) using a quadratic function. The value of the second derivative will tell us if the point x^* is a minimum or a maximum.

[1] You may find a gallery here: `http://matplotlib.org/gallery.html`.

4. You are asked to plot the quadratic approximation over the original function. For that purpose use the following Taylor expression which is equivalent to equation (1):

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

In order for x^* to be a minimum, you need $f''(x^*)$ to be positive. This can be expressed in another way: you need $d^2 f''(x^*) > 0$ for any $d \neq 0$. The latter sentence is obvious (and may seem stupid) in one dimension but has a high importance in higher dimensions.

2 Two dimensional case

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

$$f(x) = x_1^2 + x_2^2$$

You are asked to

1. Plot this function. It should be noted that this function has a minimum at $(x_1, x_2) = (0, 0)$.

2. Compute the gradient of the function, $\nabla f(x)$, and analytically compute the points x^* at which $\nabla f(x^*) = 0$.

3. Let $d \in R^2$ be the perturbation around x^*. The Taylor expansion, up to second order, of a function of several variables can be compactly expressed as

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

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

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

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

The latter matrix is giving us information about the shape of the quadratic approximation at $x = x^*$ 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 just need to check if

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

In the latter case, we are sure that the quadratic approximation is convex and that we have a minimum. We have a maximum if

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

The previous conditions can be verified by computing the eigenvalues of $\nabla^2 f(x^*)$. 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^*?

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

$$f_A(x) = -x_1^2 - x_2^2 \qquad f_B(x) = x_1^2 - x_2^2 \qquad f_C(x) = x_1^2$$

Answer the following questions:

1. At which point x^* the gradient is zero?
2. 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?

3 The exercise

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

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

Follow these steps:

1. 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.
2. Analytically compute the gradient $\nabla f(x)$, and plot $\|\nabla f(x)\|^2$ within the previous range. You are recommended to perform a contour plot within the previous range. Observe where the minimum (and maximum) may be.
3. Numerically compute an approximation of the points x^* at which $\nabla f(x^*) = 0$. For that issue
   (a) Evaluate $\|\nabla f(x)\|^2$ at the previous range using a step of e.g. 0.01 or smaller if you prefer (but not too small!).
   (b) Using brute force, search for those points $\tilde{x} \in R^2$ at which the value of $\|\nabla f(x)\|^2$ is strictly smaller than the value of its neighbors[2]. Our purpose here is to find those points at which the gradient is small: this is the principle of the gradient descent algorithms, which will be analyzed in later laboratories.

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

(c) Which are the values of $\tilde{x}$ you have obained? Which is the value of $\|\nabla f(x)\|^2$ at those points? Do they correspond to the "possible" minimums you have seen in step 2?

4. Analytically compute the Hessian of $f(x_1, x_2)$ and evaluate it at the values $\tilde{x}$ you have found. What kind of information is giving you the Hessian? Does it correspond to a minimum? To a maximum? To neither of both?

Report

You are asked to deliver an *individual* report of section 3. Just explain each of the steps you have followed. There is no need to include nothing else. If you want to include some parts of code, please include it within the report. Do not include it as separate files.

The deadline to deliver this report is October 4th at 3 p.m. (15h).