FTML practical session 5

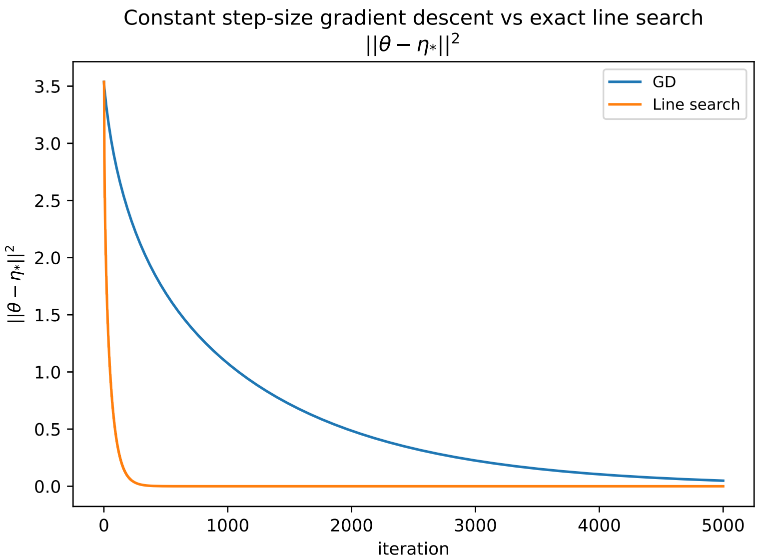

TABLE DES MATIÈRES

1 GRADIENT DESCENT AND LINE SEARCH FOR A LEAST SQUARES PROBLEM

We come back to a least squares problem. We recall the setting here :

— $\mathcal{X} = \mathbb{R}^d$
— $\mathcal{Y} = \mathbb{R}$
— $\theta = \mathbb{R}^d$ defines the estimator, which is the function $x \to x^T\theta$.
— design matrix : $X \in \mathbb{R}^{n,d}$
— vector of outputs (labels) : $y \in \mathbb{R}^n$.

We want to minimize the function f representing the empirical risk :

$$f(\theta) = \frac{1}{2n}||X\theta - y||^2 \tag{1}$$

1.1 Gradient descent

We recall that the gradient and the Hessian write :

$$\begin{aligned}\nabla_\theta f(\theta) &= \frac{1}{n}X^T(X\theta - y)\\ &= H\theta - \frac{1}{n}X^Ty\end{aligned} \tag{2}$$

$$H = \frac{1}{n}X^TX \tag{3}$$

An **iteration** of the gradient algorithm writes :

$$\theta_{t+1} = \theta_t - \gamma\nabla_\theta f(\theta_t) \tag{4}$$

1.2 Global minimizers

We consider the global minimizers of f, noted η^*. They all necessary verify

$$\nabla_\theta f(\eta^*) = 0 \tag{5}$$

Which means that

$$H\eta* = \frac{1}{n}X^Ty \tag{6}$$

As $\theta \to f(\theta)$ is convex, we know that this condition is also sufficient : any vector that cancels the gradient is a global minimum.

If H is not invertible, there might be several minimizes. However, if f is strongly convex, then H is invertible and η^* is unique. We will consider such a case here. In general, H is a positive semi-definite matrix, and here we hence consider the case where it is definite positive.

1.3 Line search : A nested optimization problem

Considering a fixed iteration step θ_t, we note

$$\alpha(\gamma) = \theta_t - \gamma\nabla_\theta f(\theta_t) \tag{7}$$

The **exact line seach** method attempts to find the optimal step γ^*, at each iteration. This means, given the position θ_t, the parameter γ that minimizes the function defined by

$$\begin{aligned} g(\gamma) &= f(\theta_t - \gamma \nabla_\theta f(\theta_t)) \\ &= f(\alpha(\gamma)) \end{aligned} \tag{8}$$

Is $g : \mathbb{R} \to \mathbb{R}$ a convex function ?

Find the value γ^* that minimizes $\gamma \to g(\gamma)$ for a given θ_t, and implement both a constant step GD and a GD where γ is set optimally thanks to this method (exact line search). Compare the convergence speeds by measuring the distance between the iterate and the OLS estimator η^* at each iteration.

You should observe a result like figures 1 and 2. Some template files can be found in **exercice_1_line_search/**

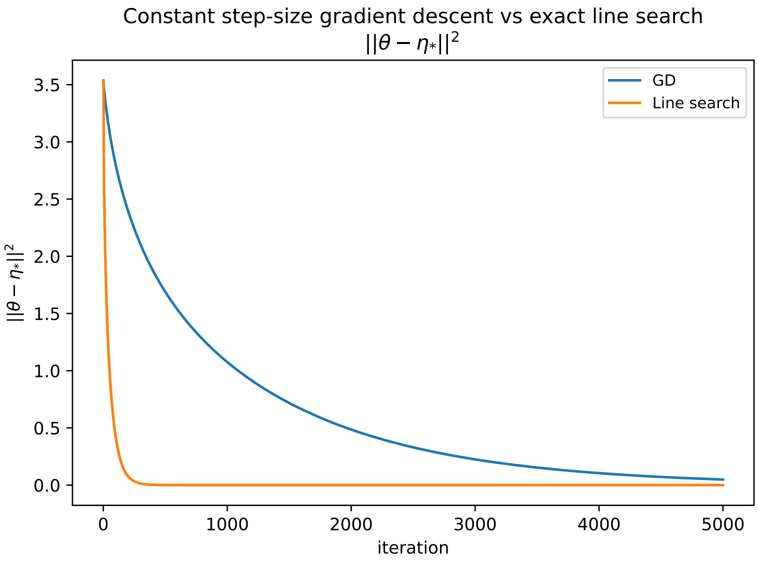

FIGURE 1 – Convergence to the global minimizer.

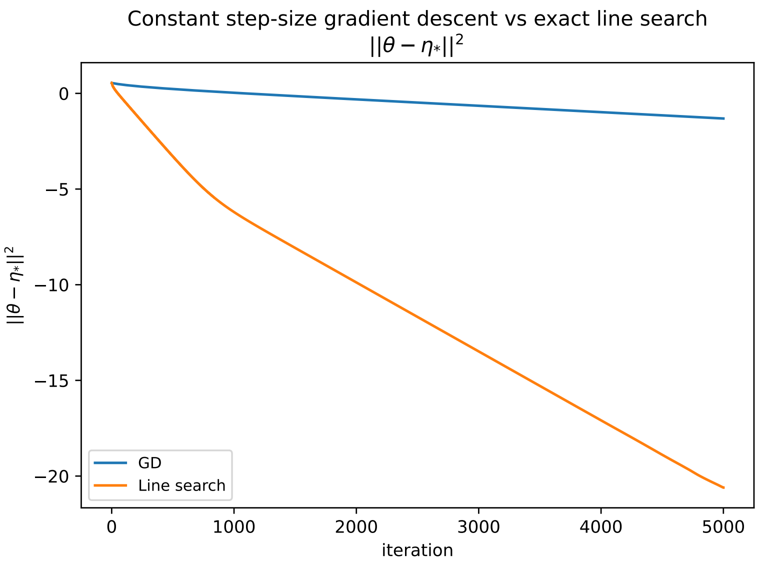

FIGURE 2 – Convergence to the global minimizer in log scale.