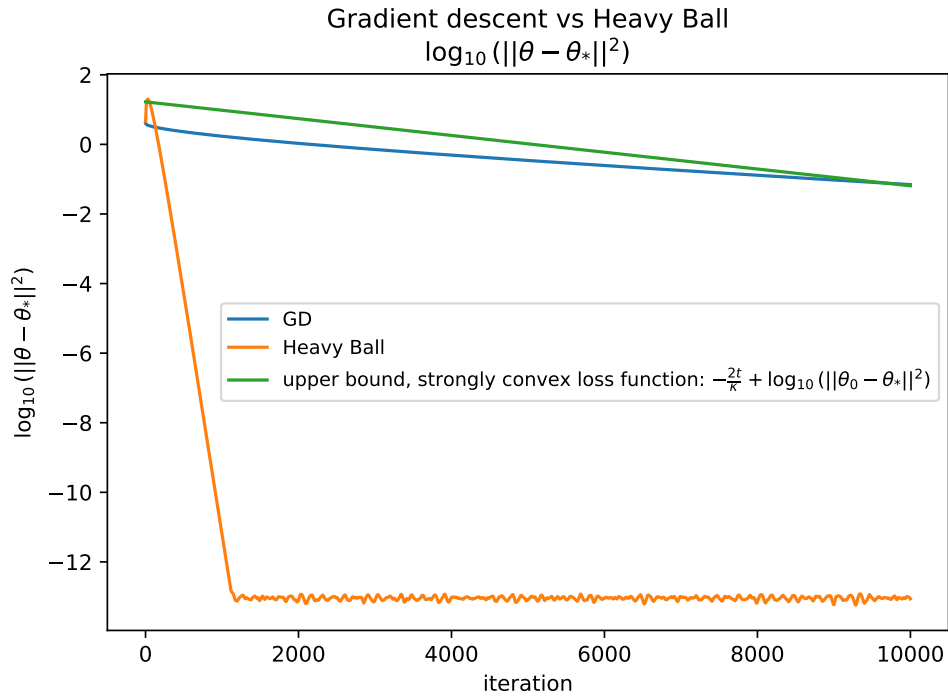


FTML Exercices 5

Pour le 4 mai 2023



1 HEAVY-BALL

In this exercise we study the values of the γ and β parameters in a Heavy-Ball algorithm, which is an acceleration of gradient descent, in the case where f is a least squares loss function.

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

The update $\theta_{t+1} - \theta_t$ is then a combination of the gradient $\nabla_{\theta_t} f$ and of the previous update $\theta_t - \theta_{t-1}$. The goal of this method it might balance the effet of oscillations in the gradient. We will use these parameters :

$$\gamma = \frac{4}{(\sqrt{L} + \sqrt{\mu})^2} \quad (2)$$

and

$$\beta = \left(\frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}} \right)^2 \quad (3)$$

We keep the same notations as in the former practical sessions dedicated to the least squares problem. μ is the smallest eigenvalue of the Hessian H and L is the largest. Assuming $\mu > 0$ (strongly convex function), we will show that the characteristic convergence time with the heavy-ball momentum term is $\sqrt{\kappa}$ instead of κ .

Let λ be an eigenvalue of H and u_λ a eigenvector for this eigenvalue. We are interested in the evolution of $\langle \theta_t - \eta^*, u_\lambda \rangle$.

We note

$$a_t = \langle \theta_t - \eta^*, u_\lambda \rangle \quad (4)$$

Exercise 1: Show that

$$a_{t+1} = (1 - \gamma\lambda + \beta)a_t - \beta a_{t-1} \quad (5)$$

Exercise 2: Compute the constant-recursive sequence a_t , and show that there exists a constant C_λ that depends on the initial conditions, such that

$$\forall t, a_t \leq (\sqrt{\beta})^t C_\lambda \quad (6)$$

https://en.wikipedia.org/wiki/Constant-recursive_sequence

If u_i is a basis of orthonormal vectors with eigenvalues λ_i , we have that

$$\begin{aligned} \|\theta_t - \eta^*\|^2 &= \sum_{i=1}^d (\langle \theta_t - \eta^*, u_i \rangle)^2 \\ &\leq \sum_{i=1}^d (\sqrt{\beta})^{2t} C_{\lambda_i}^2 \\ &= (\sqrt{\beta})^{2t} D \end{aligned} \quad (7)$$

with

$$D = \sum_{i=1}^d C_{\lambda_i}^2 \quad (8)$$

We can now remark that

$$\begin{aligned} \sqrt{\beta} &= \frac{\sqrt{L} - \sqrt{\mu}}{\sqrt{L} + \sqrt{\mu}} \\ &= \frac{1 - \sqrt{\frac{\mu}{L}}}{1 + \sqrt{\frac{\mu}{L}}} \\ &\leq 1 - \sqrt{\frac{\mu}{L}} \\ &= 1 - \frac{1}{\sqrt{\kappa}} \end{aligned} \quad (9)$$

Exercise 3: Conclude