Constant stepsize gradient descent on smooth convex function

Samuel Vaiter

Created: 2023-11-10.
Last update: 2023-11-10.
Status: inprogress.

Convex L-smooth function

Proposition

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex L-smooth function. For any initial guess $x^{(0)}$, the GD iterates $x^{(t)}$ with step-size $\eta^{(t)} = \eta \in (0, \frac{2}{L})$ converge towards an optimal point $x^\star$ of f, and moreover

$$f(x^{(t)}) - f^\star \leq \frac{2(f(x^0) - f^\star)\|x^{(0)} - x^\star\|^2}{t\eta(2 - L\eta)(f(x^0) - f^\star) + 2\|x^{(0)} - x^\star\|^2}.$$

Moreover, if $\eta = \frac{1}{L}$, then

$$f(x^{(t)}) - f^\star \leq \frac{2L\|x^{(0)} - x^\star\|^2}{t + 4}.$$

Proof based on the co-coercivity of the gradient.

Co-coercivity of ∇f

This result is sometimes coined Baillon–Haddad theorem.

Proposition

Let f be a full-domain convex L-smooth function. Then, ∇f is $1/L$-co-coercive, i.e.,

$$\forall x, y \in \mathbb{R}^n, \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{1}{L} \|\nabla f(x) - \nabla f(y)\|^2.$$