TD 8 – Gradient descent and smooth functions

Mathematics of data

04/12/24

Exercise 1. Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a convex function. The gradient descent method is defined by the update rule

$$x_{t+1} = x_t - \rho \nabla f(x_t).$$

Assume that f is L-smooth: for all $x, y \in \mathbb{R}^d$, we have

$$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2} \|x - y\|^2 .$$

Let x^* be a minimizer of f. The aim of this exercise is to obtain a convergence rate for $f(x_t) - f(x^*)$.

1. Prove that for any $x \in \mathbb{R}^d$, we have $f(x) - f(x^*) \leq \langle \nabla f(x), x - x^* \rangle$.
2. Prove that $f(x_{t+1}) \leq f(x_t) - \rho(1 - L\rho/2) \|\nabla f(x_t)\|^2$.
3. Given the previous inequality, provide the best choice for ρ (i.e. that minimizes the r.h.s.) and prove the descent lemma:
$$f(x_{t+1}) \leq f(x_t) - \frac{1}{2L} \|\nabla f(x_t)\|^2 .$$
We keep this choice of ρ in what follows.
4. Let
$$V_t \coloneqq t(f(x_t) - f(x^*)) + \frac{L}{2} \|x_t - x^*\|^2 .$$
Using questions 1 and 3, prove that
$$V_{t+1} - V_t \leq -\frac{t}{2L} \|\nabla f(x_t)\|^2 .$$
5. Derive a bound on $f(x_t) - f(x^*)$.

Exercise 2. The aim of this exercise is to establish the co-coercivity inequality

$$\frac{1}{2L} \|\nabla f(z) - \nabla f(y)\|^2 \leq f(z) - f(y) + \langle \nabla f(y), y - z \rangle$$

characterizing L-smooth convex functions $f\colon \mathbb{R}^d \to \mathbb{R}$.

1. Show that f is convex and L-smooth if and only if, for all (x, y, z):
$$f(y) + \langle \nabla f(y), x - y \rangle \leq f(z) + \langle \nabla f(z), x - z \rangle + \frac{L}{2} \|x - z\|^2 .$$
2. Show that f is convex and L-smooth if and only if for all (y, z):
$$0 \leq f(z) - f(y) + \langle \nabla f(y), y - z \rangle - \frac{1}{2L} \|\nabla f(z) - \nabla f(y)\|^2 .$$
3. Show that f is convex and L-smooth if and only if for all (y, z):
$$\frac{1}{L} \|\nabla f(z) - \nabla f(y)\|^2 \leq \langle \nabla f(y) - \nabla f(z), y - z \rangle.$$
Hint: prove that f is convex iff for all θ, η it holds $\langle \theta - \eta, \nabla f(\theta) - \nabla f(\eta) \rangle \geq 0$. For this (reverse implication), you can consider $g(t) = f(\theta + t(\eta - \theta))$, prove that $g'(t) \geq g'(0)$ for $t \geq 0$ and deduce $f(\eta) \geq f(\theta) + \langle \nabla f(\theta), \eta - \theta \rangle$.