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
$$V_t := t(f(x_t) - f(x^*)) + \frac{L}{2} \|x_t - x^*\|^2 .$$
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 $f(\eta) \geq f(\theta) + \langle \nabla f(\theta), \eta - \theta\rangle$.

Solutions

Exercise 1.

1. It is just convexity: $f(x^*) \geq f(x) + \langle \nabla f(x), x^* - x \rangle$.
2. Apply the L-smoothness inequality with $x = x_{t+1}$ and $y = x_t$.
3. The function $\rho \mapsto -\rho(1 - L\rho/2)$ is minimal for $\rho = 1/L$, which leads to
$$f(x_{t+1}) \leq f(x_t) - \frac{1}{2L} \left\| \nabla f(x_t) \right\|^2 .$$
4. We have
$$\begin{aligned} V_{t+1} - V_t &= (t+1)(f(x_{t+1}) - f(x_t)) + f(x_t) - f(x^*) + \frac{L}{2} \left\| x_{t+1} - x^* \right\|^2 - \frac{L}{2} \left\| x_t - x^* \right\|^2 \\ &= (t+1)(f(x_{t+1}) - f(x_t)) + f(x_t) - f(x^*) + \frac{1}{2L} \left\| \nabla f(x_t) \right\|^2 - \langle x_t - x^*, \nabla f(x_t) \rangle \\ &\leq -(t+1)\frac{1}{2L} \left\| \nabla f(x_t) \right\|^2 + \langle x_t - x^*, \nabla f(x_t) \rangle + \frac{1}{2L} \left\| \nabla f(x_t) \right\|^2 - \langle x_t - x^*, \nabla f(x_t) \rangle \\ &= -\frac{t}{2L} \left\| \nabla f(x_t) \right\|^2 , \end{aligned}$$
where we bounded $f(x_{t+1}) - f(x_t)$ with the descent Lemma and $f(x_t) - f(x^*)$ with question 1.
5. Using that V_t is decreasing, we get:
$$f(x_t) - f(x^*) \leq \frac{1}{t} V_t \leq \frac{1}{t} V_0 = \frac{L}{2t} \left\| x_0 - x^* \right\| .$$

Exercise 2.

1. If f is convex and L-smooth, then the inequality is true as $f(x)$ is larger than the left-hand side and smaller than the right-hand side. Conversely, taking $z = x$ in the inequality gives convexity, then $y = x$ gives smoothness.
2. With question 1, f is convex and L-smooth if and only if
$$\begin{aligned} & 0 \leq f(z) - f(y) + \langle \nabla f(z), x - z \rangle - \langle \nabla f(y), x - y \rangle + \frac{L}{2} \left\| x - z \right\|^2 \\ \Leftrightarrow\ & 0 \leq f(z) - f(y) + \frac{L}{2} \left\| x - z + \frac{1}{L}(\nabla f(z) - \nabla f(y)) \right\|^2 - \frac{1}{2L} \left\| \nabla f(z) - \nabla f(y) \right\|^2 + \langle \nabla f(y), y - z. \rangle \end{aligned}$$
Taking $x = -z + \frac{1}{L}(\nabla f(z) - \nabla f(y))$ we recover the desired inequality, which, conversely, implies the inequality above.
3. If f is convex and L-smooth, writing the characterization of question 2 for (z, y) and (y, z) (i.e., switching z and y) and summing the two inequalities gives the result. Conversely, assume that
$$\frac{1}{L} \left\| \nabla f(z) - \nabla f(y) \right\|^2 \leq \langle \nabla f(y) - \nabla f(z), y - z \rangle.$$
Admitting the result in the hint, we have convexity of f. Moreover, applying Cauchy-Schwarz inequality to the right-hand side leads to
$$\frac{1}{L} \left\| \nabla f(z) - \nabla f(y) \right\|^2 \leq \left\| \nabla f(z) - \nabla f(y) \right\| \left\| z - y \right\| ,$$
and dividing both sides by $\left\| \nabla f(z) - \nabla f(y) \right\|$ gives that ∇f is L-Lipschitz, so that f is L-smooth.
Proof of the hint: Let $g(t) = f(\theta + t(\eta - \theta))$. Then
$$g'(t) = \langle \nabla f(\theta + t(\eta - \theta)), \eta - \theta \rangle.$$
If f is convex, then g is convex so g' is non-decreasing. Therefore
$$g'(1) = \langle \nabla f(\eta), \eta - \theta \rangle \geq g'(0) = \langle \nabla f(\theta), \eta - \theta \rangle,$$

which gives the result. Conversely, assume that $\langle \nabla f(\theta) - \nabla f(\eta), \theta - \eta \rangle \geq 0$ for all θ, η. Then

$$g'(t) - g'(0) = \frac{1}{t} \langle \nabla f(\theta + t(\eta - \theta)) - f(\theta), \theta + t(\eta - \theta) - \theta \rangle \geq 0.$$

Applying the fundamental theorem of analysis we get

$$g(1) \geq g(0) + \int_0^1 g'(t) \mathrm{d}t \geq g(0) + g'(0),$$

which reads

$$f(\eta) \geq f(\theta) + \langle \nabla f(\theta), \eta - \theta \rangle.$$