Numerical Optimization – Tutorial on the rates of first-order methods

L. Desbat & F. Iutzeler

In the whole tutorial, we will assume that $f : \mathbb{R}^n \to \mathbb{R}$ is an L-smooth *convex* function with minimizers.

A. Convergence rates in the strongly convex case

Exercise 1 (Some other descent lemmas)**.**

The goal of this exercise is to provide useful lemmas for proving convergence rates. Let $x^\star$ be a minimizer of f.

a. Show that for all $x, y \in \mathbb{R}^n$,

$$f(x) - f(y) \leq \langle x - y; \nabla f(x) \rangle - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|^2$$

and thus

$$\frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \leq \langle x - y; \nabla f(x) - \nabla f(y) \rangle \leq L\|x - y\|^2.$$

Hint: Define $z = y - \frac{1}{L}(\nabla f(y) - \nabla f(x))$.
Use convexity to bound $f(x)-f(z)$ and smoothness to bound $f(z)-f(y)$ and sum both inequalities.

b. Let f be in addition μ-strongly convex; that is, $f - \frac{\mu}{2}\|\cdot\|^2$ is convex. Show that for all $x \in \mathbb{R}^n$,

$$(x - x^\star)^{\mathrm{T}}\nabla f(x) \geq \frac{\mu L}{\mu + L}\|x - x^\star\|^2 + \frac{1}{\mu + L}\|\nabla f(x)\|^2.$$

Hint: Use the fact that $f - \frac{\mu}{2}\|\cdot\|^2$ is $(L - \mu)$-smooth and question a.

Exercise 2 (Strongly convex case)**.**

The goal of this exercise is to investigate the convergence rate of the fixed stepsize gradient algorithm on a μ-strongly convex, L-smooth function:

$$x_{k+1} = x_k - \frac{2}{\mu + L}\nabla f(x_k)$$

which will introduce us to the mechanics of Optimization theory.

a. From 1b., prove that

$$\begin{aligned}\|x_{k+1} - x^\star\|^2 &\leq \left(1 - \frac{4\mu L}{(\mu + L)^2}\right)\|x_k - x^\star\|^2 \\ &= \left(\frac{\kappa - 1}{\kappa + 1}\right)^2 \|x_k - x^\star\|^2\end{aligned}$$

where $\kappa = L/\mu$ is the *conditionning number* of the problem.

b. Show that

$$f(x_k) - f(x^\star) \leq \frac{L}{2}\|x_k - x^\star\|^2.$$

c. Conclude that for the gradient algorithm with stepsize $2/(\mu + L)$ we have

$$f(x_k) - f(x^\star) \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^{2k} \frac{L\|x_0 - x^\star\|^2}{2}.$$

B. Convergence rates in the non-strongly convex case

Exercise 3 (Smooth case)**.**
The goal of this exercise is to investigate the convergence rate of the fixed stepsize gradient algorithm on an L-smooth function:

$$x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k)$$

which will introduce us to the mechanics of Optimization theory.

a. Prove that

$$\|x_{k+1} - x^\star\|^2 \leq \|x_k - x^\star\|^2 - \frac{1}{L^2}\|\nabla f(x_k)\|^2 = \|x_k - x^\star\|^2 - \|x_{k+1} - x_k\|^2.$$

b. Show that

$$\delta_k := f(x_k) - f(x^\star) \leq \|x_k - x^\star\| \cdot \|\nabla f(x_k)\| \leq \|x_1 - x^\star\| \cdot \|\nabla f(x_k)\|.$$

Hint: Use convexity then a.

c. Use smoothness and b. to show that

$$0 \leq \delta_{k+1} \leq \delta_k - \underbrace{\frac{1}{2L\|x_1 - x^\star\|^2}}_{:=\omega} \delta_k^2.$$

d. Deduce that

$$\frac{1}{\delta_{k+1}} - \frac{1}{\delta_k} \geq \omega.$$

Hint: Divide c. by $\delta_k \delta_{k+1}$.

e. Conclude that for the gradient algorithm with stepsize $1/L$ we have

$$f(x_k) - f(x^\star) \leq \frac{2L\|x_1 - x^\star\|^2}{k-1}.$$

Optimization inequalities cheatsheet

For any function f:

(convex) convex
(diff) differentiable
(min) with minimizers $X^\star$, $x^\star \in X^\star$
(smooth) L-smooth (differentiable with ∇f L Lipschitz continuous)
(strong) μ-strongly convex (μ can be taken equal to 0 below)

$$\begin{aligned} & f(y) \geq f(x) + (y-x)^{\mathrm{T}} \nabla f(x) \text{ (convex) + (diff)} \\ \Rightarrow & \langle x-y; \nabla f(x) - \nabla f(y) \rangle \geq 0 \text{ (convex) + (diff)} \end{aligned}$$

$$\begin{aligned} & f(x^\star) \leq f(x) \forall x \text{ (minimizer)} \\ \Rightarrow & \nabla f(x^\star) = 0 \text{ (convex) + (diff) + (minimizer)} \end{aligned}$$

$$\begin{aligned} & \|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \text{ (smooth)} \\ \Rightarrow & f(x) \leq f(y) + (x-y)^{\mathrm{T}} \nabla f(y) + \frac{L}{2}\|x-y\|^2 \text{ (smooth)} \\ \Rightarrow & \langle x-y; \nabla f(x) - \nabla f(y) \rangle \leq L\|x-y\|^2 \text{ (smooth)} \end{aligned}$$

$$\begin{aligned} & f(x) - \frac{\mu}{2}\|x\|^2 \text{ is convex (strong)} \\ \Rightarrow & f(y) + (x-y)^{\mathrm{T}} \nabla f(y) + \frac{\mu}{2}\|x-y\|^2 \leq f(x) \text{ (strong) + (diff)} \\ \Rightarrow & \mu\|x-y\|^2 \leq \langle x-y; \nabla f(x) - \nabla f(y) \rangle \text{ (strong) + (diff)} \end{aligned}$$

Combining the above, when f is μ-strongly convex and L-smooth:

$$f(y) + (x-y)^{\mathrm{T}} \nabla f(y) + \frac{\mu}{2}\|x-y\|^2 \leq f(x) \leq f(y) + (x-y)^{\mathrm{T}} \nabla f(y) + \frac{L}{2}\|x-y\|^2$$

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

If in addition, f is twice differentiable,

$$\mu I \leq \nabla^2 f(x) \leq LI$$