Introduction to Continuous optimization
Assessment
(10 December 2024)
Durée: 3h

Exercise I: Strongly convex functions

One considers a convex function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$, proper, lsc, and one assumes it is μ-strongly convex: $\forall x, y \in \mathcal{X}, t \in [0, 1]$,

$$f(tx + (1-t)y) \le tf(x) + (1-t)f(y) - \mu\frac{t(1-t)}{2}\|x-y\|^2.$$

Here, $\mathcal{X}$ can be $\mathbb{R}^N$, a Hilbert space, or even a Banach space, without changing much the results. We recall (see lecture notes or slides) that for all $x, y \in \mathcal{X}$ and $p \in \partial f(x)$,

$$f(y) \ge f(x) + \langle p, y-x\rangle + \frac{\mu}{2}\|x-y\|^2$$

We introduce the conjugate:

$$f^*(p) := \sup_{x\in\mathcal{X}} \langle p, x\rangle - f(x)$$

1. What can we say about f^*? Deduce that for any $p, q \in \mathcal{X}^*$,

$$f^*(q) \le f^*(p) + \langle \nabla f^*(p), q-p\rangle + \frac{1}{2\mu}\|q-p\|_*^2$$

(here the norm $\|\cdot\|_*$ on $\mathcal{X}$ is the dual norm of $\|\cdot\|$, which, in the Euclidean or Hilbert cases, is the same as $\|\cdot\|$).

2. We consider $x, y \in \mathcal{X}$ with nonempty subgradients, and $p \in \partial f(x)$, $q \in \partial f(y)$. Deduce from the previous question that

$$f(x) \le f(y) + \langle q, x-y\rangle + \frac{1}{2\mu}\|q-p\|_*^2.$$

Hint: use Legendre-Fenchel's identity.

Exercise II: Polyak Step Sizes

[After Polyak, 1987, Hazan & Kakade, 2022] We consider a variant of the subgradient descent algorithm, with the assumption that the value of the minimizer is known. The idea is due to Boris Polyak (*Introduction to Optimization*, 1987).

Let $f \in C^1(\mathbb{R}^N)$ ($\mathbb{R}^N$ is endowed with the standard Euclidean scalar product and norm) a convex function, with a minimizer x^*, and we assume the value $f(x^*)$ is known. We will denote $\alpha \ge 0$ the convexity parameter of f:

$$\forall x, y\,, \quad f(y) \ge f(x) + \nabla f(x)\cdot(y-x) + \frac{\alpha}{2}\|y-x\|^2 \tag{SC}$$

($\alpha = 0$ means f is just convex, $\alpha > 0$ means strongly convex), and $\beta \in [0, +\infty]$ the Lipschitz constant of the gradient ($\beta = +\infty$ means f might not have a Lipschitz gradient):

$$\forall x, y\,, \quad f(y) \le f(x) + \nabla f(x)\cdot(y-x) + \frac{\beta}{2}\|y-x\|^2. \tag{Lip}$$

We consider the algorithm, given x_0 and a number of iterations K:

1. let $k = 0$ and then do:
2. let $\eta^k = \dfrac{f(x^k) - f(x^*)}{\|\nabla f(x^k)\|^2}$

3. $x^{k+1} = x^k - \eta^k \nabla f(x^k)$,

4. $k \leftarrow k+1$,

5. stop if $k = K$ or some ending criterion is satisfied, else return to 2

6. return the point x_k such that $f(x^k) = \min_\ell f(x^\ell)$.

1. We let $d_k := \|x^k - x^*\|$. Show that

$$d_{k+1}^2 \le d_k^2 - \frac{(f(x^k) - f(x^*))^2}{\|\nabla f(x^k)\|^2}. \tag{‡}$$

2. Deduce that, letting $G = \sup_{\|x-x^*\| \le \|x^0 - x^*\|} \|\nabla f(x)\|$,

$$\min_{0 \le k \le K} f(x^k) - f(x^*) \le \frac{G\|x^0 - x^*\|}{\sqrt{K}}.$$

3. We now assume $\beta < +\infty$, that is, f has Lipschitz gradient. Show that

$$\frac{1}{2\beta}\|\nabla f(x)\|^2 \le f(x) - f(x^*) \le \frac{\beta}{2}\|x - x^*\|^2$$

4. Deduce that

$$\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - \frac{f(x^k) - f(x^*)}{2\beta},$$

and then that

$$\min_{0 \le k \le K} f(x^k) - f(x^*) \le \frac{2\beta\|x^0 - x^*\|^2}{K}.$$

7. One now assumes that $\alpha > 0$. Show that for any x,

$$\frac{\alpha}{2}\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{1}{2\alpha}\|\nabla f(x)\|^2.$$

Hint: one possibility is to use Exercise **I**, Question **2.**

8. Show that for all $k \ge 1$,

$$\|x^k - x^*\| \le \frac{2G}{\alpha\sqrt{k+1}}.$$

where G is as in question **2.**

9. Now we consider the case where both $\alpha > 0$ and $\beta < +\infty$. Show that for all $k \ge 0$,

$$d_{k+1}^2 \le \left(1 - \frac{\alpha}{4\beta}\right) d_k^2.$$

Deduce that $f(x^k) - f(x^*) \le \beta(1 - \alpha/(4\beta))^k \|x^0 - x^*\|^2/2$.

10. Returning to the strongly convex case ($\alpha > 0$, β possibly $+\infty$), using (‡) and question **8.**, show that

$$f(x^k) - f(x^*) \le \frac{C}{k}$$

for some constant C depending on G, α. Hint: sum (‡) from $k/2$ (or $(k+1)/2$ if k odd) to k.

Exercise III - conjugates

Find the convex (Legendre-Fenchel) conjugate $f^*(y) = \sup_x \langle x, y\rangle - f(x)$ in the following cases:

1 $f(x) = x\ln x - x$ for $x \geq 0$ (and $0\ln 0 = 0$), and $+\infty$ for $x < 0$.

2. $f : x \mapsto -\sqrt{x}$ if $x \geq 0$, $+\infty$ if $x < 0$.

3. $f(x) = 0$ if $x_i \geq 0$, $\sum_i x_i = 1$, and $+\infty$ else, for $x \in \mathbb{R}^n$ (f is the characteristic function of the $(n-1)$-dimensional unit simplex): show that $f^*(y) = \max_i y_i$.

4. $f(x) = 0$ if $\sum_{i=1}^n x_i = 0$ and $\sum_{i=1}^n |x_i| \leq 2$, and $f(x) = +\infty$ else, for $x \in \mathbb{R}^n$.

Exercise IV: First eigenvalue of the Dirichlet Laplace operator

We consider $\Omega \subset \mathbb{R}^N$, $N \geq 1$, a bounded, connected open set, and the (Hilbert) space $H^1(\Omega)$ of Sobolev functions such that $\nabla u \in L^2(\Omega;\mathbb{R}^N)$ and

$$\int_\Omega |u|^2 + |\nabla u|^2\, dx =: \|u\|^2$$

is finite. We consider more precisely $H^1_0(\Omega)$, which is the closure in $H^1(\Omega)$ of the smooth functions with compact support. The first eigenvalue of the (Dirichlet) Laplacian is defined as the number:

$$\lambda_1(\Omega) = \inf\left\{\int_\Omega |\nabla u|^2 dx : u \in H^1_0(\Omega), \int_\Omega |u|^2 dx = 1\right\}. \qquad (\lambda_1)$$

We admit that $\lambda_1 > 0$ and we want to show that the problem (λ_1) is in fact a convex problem.

We recall an important fact about Sobolev functions: if $u, v \in H^1(\Omega)$, then $w = \max\{u, v\} \in H^1(\Omega)$ with, for almost every $x \in \Omega$:

$$\nabla w(x) = \begin{cases} \nabla u(x) & \text{if } u(x) > v(x)\,, \\ \nabla v(x) & \text{if } u(x) > v(x)\,, \\ \nabla u(x) = \nabla v(x) & \text{if } u(x) = v(x). \end{cases}$$

and in particular, $|\nabla |u|| = |\nabla u|$ almost everywhere.

1. Show that in (λ_1) one can consider only non-negative and boudned functions. One first has to show that the set:

$$\Lambda_1 := \left\{\int_\Omega |\nabla u|^2 dx : u \in H^1_0(\Omega), \int_\Omega |u|^2 dx = 1\right\} = \left\{\int_\Omega |\nabla u|^2 dx : u \in H^1_0(\Omega), u \geq 0 \text{ a.e.}, \int_\Omega |u|^2 dx = 1\right\}$$

and then that

$$\Lambda_1 = \text{closure of} \left\{\int_\Omega |\nabla u|^2 dx : u \in H^1_0(\Omega), \exists M, u \leq M \text{ a.e.}, u \geq 0 \text{ a.e.}, \int_\Omega |u|^2 dx = 1\right\}$$

2. We assume $u \in \Lambda_1$, and that u is bounded and non-negative. We let $v = u^2$. We admit that $v \in H^1_0(\Omega)$ (this is because u is bounded). Show that:

$$|\nabla u(x)| = \begin{cases} \frac{|\nabla v(x)|}{2\sqrt{v(x)}} & \text{if } v(x) > 0, \\ 0 & \text{if } v(x) = 0 \end{cases}$$

almost everywhere in Ω.

3. We consider the function, defined for $(p,t) \in \mathbb{R}^N \times \mathbb{R}$:

$$\Phi(p,t) := \begin{cases} \frac{|p|^2}{4t} & \text{if } t > 0, \\ 0 & \text{if } (v,t) = (0,0), \\ +\infty & \text{if } t < 0 \text{ or } t = 0, v \neq 0. \end{cases}$$

Show that Φ is positively 1-homogeneous.

4. Show that for $(p,t) \in \mathbb{R}^N \times \mathbb{R}$,

$$\Phi(p,t) = \sup \left\{ p \cdot q + ts : (q,s) \in \mathbb{R}^N \times \mathbb{R}\,, |q|^2 + s \leq 0 \right\}$$

What can we say about the function Φ?

5. Deduce that:

$$\Lambda_1 = \text{closure of} \left\{ \int_\Omega \Phi(\nabla v, v) dx : v \in H_0^1(\Omega), \exists M, 0 \leq v \leq M \text{ a.e.}, \int_\Omega v\, dx = 1 \right\}.$$

and show that this set is convex. Conclude.