**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 $\mu$-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 $\|\cdot\|$).

We know that $f^*$ is $C^1$ with $\nabla f^*$ $(1/\mu)$-Lipschitz, from the lecture notes (this is obtained by considering the conjugate of the second characterization of strong convexity). The estimate is then deduced in a standard way (by considering Taylor's expansion of $f^*$ near $p$).

**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.

We use the inequality in the previous question. We first remark that

$$\langle p, x\rangle = f(x) + f^*(p)\,, \quad \langle q, y\rangle = f(y) + f^*(q),$$

in addition, $x = \nabla f^*(p)$ and $y = \nabla f^*(q)$. Then, we replace $f^*(q)$ with $\langle q, y\rangle - f(y)$ and $f^*(p)$ with $\langle p, x\rangle - f(x)$ in that inequality. We easily conclude.

# 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

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

One has

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^*\|^2 - 2\eta^k \nabla f(x^k) \cdot (x^k - x^*) + (\eta^k)^2 \|\nabla f(x^k)\|^2.$$

By convexity, $f(x^*) \ge f(x^k) + \nabla f(x^k) \cdot (x^* - x^k)$, that is, $-\nabla f(x^k) \cdot (x^k - x^*) \le -(f(x^k) - f(x^*))$, hence:

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^*\|^2 - 2\eta^k (f(x^k) - f(x^*)) + (\eta^k)^2 \|\nabla f(x^k)\|^2.$$

The choice $\eta^k$ in the algorithm is precisely the $\eta$ which minimizes the right-hand side, and the optimal value answers the question.

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

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

Summing (‡) from $k = 0$ to $K - 1$, one finds

$$\|x^K - x^*\|^2 + \frac{K}{G^2} \min_{0 \le k \le K} (f(x^k) - f(x^*))^2 \le \|x^K - x^*\|^2 + \sum_{k=0}^{K-1} \frac{(f(x^k) - f(x^*))^2}{\|\nabla f(x^k)\|^2} \le \|x^0 - x^*\|^2$$

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

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

The right hand side comes from (Lip), replacing $x$ with $x^*$ (and using $\nabla f(x^*) = 0$) and $y$ with $x$.

The left-hand side comes from the co-coercivity: one has

$$f(y) \ge f(x) + \nabla f(x) \cdot (y - x) + \frac{1}{2\beta}\|\nabla f(y) - \nabla f(x)\|^2$$

which can be deduced from the fact that $f^*$ is $(1/\beta)$-convex (see slides and lecture notes). Again, replacing $x$ with $x^*$ and $y$ by $x$ shows the inequality.

**4.** Deduce that

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

and then that

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

One first uses that:

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

so that from the first question,

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

Reasoning then as in the second question, we have

$$K \min_{0 \le k \le K} \frac{f(x^k) - f(x^*)}{2\beta} + d_K^2 \le \sum_{k=0}^{k-1} \frac{f(x^k) - f(x^*)}{2\beta} + d_K^2 \le d_0^2.$$

The result follows.

**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.**

Now, the left-hand side follows trivially from (SC) (for $x$ replaced with $x^*$ and $y$ with $x$). The right-hand side follows from the previous exercise, second question: one has for any $x, y$:

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

Choosing $y = x^*$, we get the estimate.

**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.**

Now, we use that $(f(x^k) - f(x^*))^2 \ge \alpha^2 d_k^4/4$ to find that:

$$d_{k+1}^2 \le d_k^2 \left(1 - \frac{\alpha^2}{4\|\nabla f(x^k)\|^2} d_k^2\right)$$

We deduce that if $a_k = \frac{\alpha^2}{4G^2} d_k^2$,

$$a_{k+1} \le a_k(1 - a_k)$$

so that $a_k \le \frac{1}{k+1}$. Hence for all $k$,

$$d_k^2 \le \frac{4G^2}{\alpha^2(k+1)}.$$

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

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

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

As in question **4.**, we have

$$d_{k+1}^2 \leq d_k^2 - \frac{f(x^k) - f(x^*)}{2\beta}.$$

Using then question **7.**, left hand side, we obtain

$$d_{k+1}^2 \leq d_k^2 - \frac{\alpha d_k^2}{4\beta}.$$

In particular, $d_k \leq (1 - \alpha/(4\beta))^{k/2} d_0$.

Again using that $\nabla f$ is Lipschitz, we have $f(x^k) - f(x^*) \leq \beta/2 d_k^2$, and the conclusion follows.

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

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

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

Summing (‡) from $\lceil k/2 \rceil$ to $k$, we find:

$$\sum_{\ell=\lceil k/2 \rceil}^{k} \frac{(f(x^\ell) - f(x^*))^2}{G^2} + d_k^2 \leq d_{\lceil k/2 \rceil}^2 \leq \frac{8G^2}{\alpha^2 k}$$

where we have used the estimate in question **8.** for $d_{\lceil k/2 \rceil}^2$.

It follows that

$$\frac{k}{2G^2} \min_{0 \leq \ell \leq k} (f(x^\ell) - f(x^*))^2 \leq \frac{8G^2}{\alpha^2 k}$$

so that

$$\min_{0 \leq \ell \leq k} f(x^\ell) - f(x^*) \leq \frac{4G^2}{\alpha k}.$$

## 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 $x < 0$.

$f'(x) = \ln(x)$, hence $(f')^{-1}(y) = \exp(y)$, and this should be $(f^*)'$, which shows that $f^*(y) = \exp(y)$ (plus a constant, but $f^*(0) = -\min_x f$ is reached for $\ln x = 0$, that is $x = 1$, and the value is 1).

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

For $f(x) = -\sqrt{x}$ ($x \geq 0$) we can have $y = f'(x) = -1/(2\sqrt{x})$ only if $y < 0$. Actually, if $y \geq 0$, $yx + \sqrt{x} \to +\infty$ as $x \to +\infty$, so that $f^*(y) = +\infty$. Otherwise, $x = 1/(4y^2)$ and $f^*(y) = -1/(4|y|) + 1/(2|y|) = 1/(4|y|) = -1/(4y)$.

**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 $f^*(y) = \max_i y_i$.

First, denoting $\Sigma$ the unit simplex, one has $f^*(y) = \sup_{x\in\Sigma} x \cdot y \leq (\max_i y_i)(\sum_j x_j) = \max_i y_i$ since $y_i \leq \max_j y_j$ for all $i = 1, \dots, n$ and $x_i \geq 0$ for $x \in \Sigma$. Then, if $x_i = 0$ for all $i$ except $x_{\bar\imath} = 1$ for $\bar\imath$ such that $y_{\bar\imath} = \max_i y_i$, then $x \cdot y = \max_i y_i$. Hence, $f^*(y) = \max_i y_i$.

**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 $x \in \mathbb{R}^n$.

First, the sup is reached at an extremal point, hence one can assume $\sum_i |x_i| = 2$ at a maximum. Then for such $x$, $x_i = x_i^+ - x_i^-$ (positive – negative part) with $\sum_i x_i^+ = \sum_i x_i^-$ and $\sum_i x_i^\pm = 1$. Hence,

$$f^*(y) = \sup_{x_i^+\geq 0, \sum_i x_i^+ = 1} \sum_i x_i^+ y_i + \sup_{x_i^- \geq 0, \sum_i x_i^- = 1} \sum_i -x_i^- y_i$$

One has $\sum_i x_i^+ y_i \leq \max_i y_i$ and if $x_i^+ = 0$ except $x_{\bar\imath}^+ = 1$ for $\bar\imath$ such that $y_{\bar\imath}$ is the maximal value, $\sum_i x_i^+ y_i = y_{\bar\imath} = \max_i y_i$. Then also, $\sup_{x_i^-\geq 0, \sum_i x_i^- = 1} \sum_i -x_i^- y_i = \max_i(-y_i) = -\min_i y_i$. Hence, $f^*(y) = \max_i y_i - \min_i y_i$.

## 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 $H_0^1(\Omega)$, 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:

$$\lambda_1(\Omega) = \inf \left\{ \int_\Omega |\nabla u|^2 dx : u \in H_0^1(\Omega), \int_\Omega |u|^2 dx = 1 \right\}. \qquad (\lambda_1)$$

We admit that $\lambda_1 > 0$ and we want to show that the problem $(\lambda_1)$ 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 <u>almost every</u> $x \in \Omega$:

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

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

**1.** Show that in $(\lambda_1)$ one can consider only non-negative and boudned functions. One first has to show that the set:

$$\Lambda_1 := \left\{ \int_\Omega |\nabla u|^2 dx : u \in H_0^1(\Omega), \int_\Omega |u|^2 dx = 1 \right\} = \left\{ \int_\Omega |\nabla u|^2 dx : u \in H_0^1(\Omega), u \geq 0 \text{ a.e.}, \int_\Omega |u|^2 dx = 1 \right\}$$

and then that

$$\Lambda_1 = \text{closure of} \left\{ \int_\Omega |\nabla u|^2 dx : u \in H_0^1(\Omega), \exists M, u \leq M \text{ a.e.}, u \geq 0 \text{ a.e.}, \int_\Omega |u|^2 dx = 1 \right\}$$

The first statemet is obvious since $\int |\nabla u|^2 = \int |\nabla |u||^2$. For the second, we consider $u \geq 0$ and the sequence $u_k = \min\{u, k\}$ for $k \geq 1$. Then, it is a sequence of non-negative bounded functions in $H_0^1(\Omega)$. One has $u_k \leq u$, hence:

$$1 \geq \int_\Omega u_k^2 dx = \int_{\{u\leq k\}} u^2 dx + k^2 |\{u > k\}| = 1 - \int_{\{u>k\}} u^2 dx + k^2 |\{u > k\}|.$$

In particular $|\{u>k\}|\le 1/k^2\to 0$ (so that $\bigcap_k\{u>k\}$ has zero measure) and one also deduces $\int_{\{u>k\}}u^2dx\to 0$, which shows that $\int u_k^2dx\to 1$. We let $u_k'=u_k/\sqrt{\int u_k^2}$ so that $u_k'$ is bounded, non-negative, and $\int {u_k'}^2dx=1$.

In addition, $\nabla u_k=\nabla u$ a.e. in $\{u\le k\}$, 0 in $\{u>k\}$, so that

$$\int_\Omega|\nabla u_k|^2dx=\int_{\{u\le k\}}|\nabla u|^2dx=\int_\Omega|\nabla u|^2dx-\int_{\{u>k\}}|\nabla u|^2dx.$$

The last integral goes to zero again. We also deduce that $\int|\nabla u_k'|^2dx\to\int|\nabla u|dx$.

**2.** We assume $u\in\Lambda_1$, and that $u$ is bounded and non-negative. We let $v=u^2$. We admit that $v\in H_0^1(\Omega)$ (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 $\Omega$.

One has $\nabla v=2u\nabla u=2\sqrt{v}\nabla u$ so that $|\nabla v|=2\sqrt{v}|\nabla u|$. In addition, $v=0$ if and only if $u=0$. From the introduction, $\nabla u=0$ a.e. in $\{u=0\}$. This shows the result.

**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 $\Phi$ is positively 1-homogeneous.

This is obvious as for $t>0$, $\lambda>0$, $|\lambda p|^2/(4\lambda t)=\lambda(|p|^2/(4t))$, etc.

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

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

What can we say about the function $\Phi$?

- If $t<0$, $\sup_{|q|^2+s\le 0}p\cdot q+ts=+\infty$ since one can send $s\to-\infty$.
- If $t=0$, $\sup_{|q|^2+s\le 0}p\cdot q+ts=\sup_q p\cdot q=+\infty$ if $p\neq 0$, 0 else (since one can always choose $s=-|q|^2$).
- If $t>0$, $\sup_{|q|^2+s\le 0}p\cdot q+ts=\sup_q p\cdot q-t|q|^2=t\sup_q p\cdot(q/t)-|q|^2=|q|^2/(4t)$.

In particular, $\Phi$ is convex, lsc, as the conjugate of the characteristic function of $\{(q,s):|q|^2+s\le 0\}$.

**5.** Deduce that:

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

and show that this set is convex. Conclude.

Given $u\in H_0^1(\Omega)$, bounded, nonnegative, with $u\in\Lambda_1$, we have that the function $v=u^2$ is in $H_0^1(\Omega)$, nonnegative and bounded, and $\int_\Omega v\,dx=1$. In addition, thanks to question **2.**, $|\nabla u|^2$ is $|\nabla v|^2/(4v)$ if $v>0$, and 0 if $v=0$, almost-everywhere. As a result, $\int|\nabla u|^2=\int\Phi(\nabla v,v)$. Hence the set in the right-hand side of the question is the same as the set in the last equation of question **1.** Now since $\Phi$ is convex, this set is convex. Hence the minimization problem for finding $\lambda_1$ can be written as a convex minimization problem.