Continuous Optimization
Introduction à l'optimisation continue
Assessment
(4th January 2021)

1. Convex analysis: Exercise

1. Evaluate the convex conjugate (Legendre-Fenchel conjugate) of the functions:

1. $x \mapsto |x|^3/3$;
2. $x \mapsto 3x$;
3. $x \mapsto \langle Ax, x\rangle /2$ where $x \in \mathbb{R}^n$ and A is a symmetric, positive definite operator;
4. $x \mapsto -\sqrt{x}$ if $x \geq 0$, $+\infty$ if $x < 0$.

In general, we have $f^*(y) = \sup_y xy - f(x)$ and the sup is reached for $y = f'(x)$, or $x = (f')^{-1}(y)$, if this makes sense (in the strictly convex case, it should since f' is an increasing function, invertible). For $f(x) = |x|^3/3$, we write $y = |x|x$ so that $x = \sqrt{|y|}\mathrm{sign}\,(y)$. We find $f^*(y) = |y|^{3/2} - |y|^{3/2}/3 = (2/3)|y|^{3/2}$.

For $f(x) = 3x$, we can write that $f(x) = \sup_{y=3} yx$, that is, f is the conjugate of the characteristic function:

$$\delta_{\{3\}}(y) = \begin{cases} 0 & \text{if } y = 3 \\ +\infty & \text{else.} \end{cases}$$

We find that $f^* = \delta_{\{3\}}$.

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 $f^*(y) = -1/(4|y|) + 1/(2|y|) = 1/(4|y|) = -1/(4y)$.

For $f(x) = \langle Ax, x\rangle /2$, $x \in \mathbb{R}^n$, we write

$$f^*(y) = \sup_x \langle y, x\rangle - \langle Ax, x\rangle /2$$

and since f is strongly convex the sup is reached at some point x, and one has $y - Ax = 0$, that is $x = A^{-1}y$. We find that $f^*(y) = \left\langle A^{-1}y, y\right\rangle /2$.

2. Evaluate $y = \mathrm{prox}_{\tau f}(x)$ for $f(x) = |x|^3/3$, $\tau > 0$.

That is, we have to solve

$$\min_y \frac{|y|^3}{3} + \frac{|y-x|^2}{2\tau}.$$

The minimizer satisfies $\tau|y|y + y - x = 0$, that is $y(1 + \tau|y|) = x$. In particular y has the same sign as x, and $\mathrm{prox}_{\tau f}(-x) = -\mathrm{prox}_{\tau f}(x)$. Hence we may assume that $x > 0$, and $y > 0$. We solve $\tau y^2 + y - x = 0$ which has a positive and a negative solution. We are interested only in the positive solution, it is

$$y = \frac{\sqrt{1 + 4\tau x} - 1}{2\tau}.$$

3. (More difficult) Evaluate the convex conjugate of the "Entropy" function:

$$S\ :\ \mathbb{R}^n \to [0,+\infty]\ ; \qquad x \mapsto \begin{cases} \sum_{i=1}^n x_i \ln x_i & \text{if } x_i \geq 0\ \forall i, \sum_i x_i = 1\,, \\ +\infty & \text{else,} \end{cases}$$

where here by convention we let $t\ln t = 0$ when $t = 0$. (Hint: introduce a Lagrange multiplier for the constraint $\sum_i x_i = 1$.)

We have to compute, for $y \in \mathbb{R}^n$,

$$S^*(y) = \sup_{x_i \geq 0, \sum_i x_i = 1} \sum_i x_i y_i - x_i \ln x_i.$$

At the maximum point x (which exists since x is in a compact set) one should have $y_i - \ln x_i - 1 = \lambda$ where $\lambda \in \mathbb{R}$ is the Lagrange multiplier. That is, $x_i = \exp(y_i - 1 - \lambda)$. One has $\sum_i x_i = (1/\exp(1+\lambda))\sum_i \exp(y_i) = 1$ so that $1 + \lambda = \ln \sum_i \exp(y_i)$. (Incidentally, $x_i = \exp(y_i)/(\sum_j \exp(y_j))$.)

Then, we use $\sum_i (x_i y_i - x_i \ln x_i - x_i) = \lambda \sum_i x_i$ to deduce that $S^*(y) = 1 + \lambda$. It follows

$$S^*(y) = \ln \sum_{i=1}^n e^{y_i}$$

(the "soft-max" or "log-sum-exp" function).

2. Convex analysis: Moreau-Yosida regularization

Given $f : \mathbb{R}^n \to]-\infty,+\infty]$ a convex, lower-semicontinuous function, which is proper (that is, $f > -\infty$ and $\operatorname{dom} f \neq \emptyset$), we recall that the Moreau-Yosida regularization of f with parameter $\tau > 0$ is given by:

$$f_\tau(x) = \min_y f(y) + \frac{1}{2\tau}\|y - x\|^2$$

We recall that for any x, this problem has a unique minimizer y (because the function to minimize is strongly convex, lower-semicontinuous) and that the minimizer is also known as $y = \operatorname{prox}_{\tau f}(x)$, the "proximity operator" of τf evaluated at x. In particular, $f_\tau(x) \in \mathbb{R}$ and $\operatorname{dom} f_\tau = \mathbb{R}^n$. Further properties of the $\operatorname{prox}_{\tau f}$ operator are described in the lecture notes.

1. Show (by giving a proof or invoking the appropriate result in the lecture notes) that f_τ is convex, lower-semicontinuous.

The function f_τ is trivially convex as if $x, x' \in \mathbb{R}^n$ and $t \in [0,1]$, letting $y = \operatorname{prox}_{\tau f}(x)$, $y' = \operatorname{prox}_{\tau f}(x')$, and $y_t = ty + (1-t)y'$,

$$\begin{aligned} f_\tau(tx + (1-t)x') &\leq f(y_t) + \frac{1}{2\tau}\|y_t - (tx + (1-t)x')\|^2 \\ &\leq tf(y) + (1-t)f(y') + t\frac{1}{2\tau}\|y - x\|^2 + (1-t)\frac{1}{2\tau}\|y' - x'\|^2 \\ &= tf_\tau(x) + (1-t)f_\tau(x'). \end{aligned}$$

It is lower-semicontinous as the inf-convolution of a quadratic function and a convex, lower-semicontinuous and proper function (Lemma 4.20 in the notes). This can be easily re-proved in this particular, simpler case: if $x_n \to x$ then since $\mathrm{prox}_{\tau f}$ is 1-Lipschitz (see for instance Thm 4.28), $y_n := \mathrm{prox}_{\tau f}(x_n) \to \mathrm{prox}_{\tau f}(x) =: y$ and one has

$$f_\tau(x) \le f(y) + \frac{1}{2\tau}\|y-x\|^2 \le \liminf_{n\to\infty} f(y_n) + \frac{1}{2\tau}\|y_n - x_n\|^2 = \liminf_{n\to\infty} f_\tau(x_n).$$

2. Let $x \in \mathbb{R}^n$ and $p \in \partial f_\tau(x)$. Show that for any $h \in \mathbb{R}^n$,

$$p \cdot h \le \left(\frac{x - \mathrm{prox}_{\tau f}(x)}{\tau}\right) \cdot h.$$

(Hint: bound from below and above $f_\tau(x+th)$, for $t > 0$ small, then send t to zero.) Deduce that f_τ is differentiable at x, with $\nabla_\tau f(x) = (x - \mathrm{prox}_{\tau f}(x))/\tau$.

One has

$$f_\tau(x+th) \ge f_\tau(x) + tp\cdot h$$

and for $y = \mathrm{prox}_{\tau f}(x)$,

$$\begin{aligned} f_\tau(x+th) \le f(y) + \frac{1}{2\tau}\|x+th-y\|^2 &= f(y) + \frac{1}{2\tau}\|x-y\|^2 + t\frac{1}{\tau}(x-y)\cdot h + \frac{t^2}{2\tau}h^2 \\ &= f_\tau(x) + t\frac{1}{\tau}(x-y)\cdot h + \frac{t^2}{2\tau}h^2. \end{aligned}$$

Hence, combining both inequalities and dividing by $t > 0$,

$$p\cdot h \le \frac{1}{\tau}(x-y)\cdot h + \frac{t}{2\tau}h^2$$

and letting $t \to 0$ we deduce the required inequality. Then, since this is true for any h, and in particular for both h and $-h$, it is an equality, and it shows that $p = (x-y)/\tau$. In particular, there is only a unique subgradient at each point which shows that f_τ is differentiable at x and $p = \nabla f_\tau(x)$.

3. Recall why $\mathrm{prox}_{\tau f}$ is "firmly non-expansive". Deduce that ∇f_τ is $(1/\tau)$-Lipschitz.

Thm 4.28 asserts that $\mathrm{prox}_{\tau f} = (I + \tau\partial f)^{-1}$ is "firmly non-expansive" as the "resolvent" of the maximal-monotone operator $A = \tau\partial f$. (We recall that by minimality, $\mathrm{prox}_{\tau f}(x) = y$ solves

$$\frac{y-x}{\tau} + \partial f(y) \ni 0 \ \Leftrightarrow\ y = (I + \tau\partial f)^{-1}(x)$$

and is the resolvent of a maximal monotone operator.) This means that

$$\|x - \mathrm{prox}_{\tau f}(x) - (x' - \mathrm{prox}_{\tau f}(x'))\|^2 + \|\mathrm{prox}_{\tau f}(x) - \mathrm{prox}_{\tau f}(x')\|^2 \le \|x - x'\|^2$$

and in particular $\|\tau\nabla f_\tau(x) - \tau\nabla f_\tau(x')\| \le \|x - x'\|$.

4. Recall "Moreau's" identity. Deduce that $\nabla f_\tau(x) = \text{prox}_{\frac{1}{\tau}f^*}(x/\tau)$ where f^* is the convex conjugate (Legendre-Fenchel transform) of f.

This is in the notes:

$$x = \text{prox}_{\tau f}(x) + \tau \text{prox}_{\frac{1}{\tau}f^*}(\tfrac{x}{\tau}).$$

And the other identity follows from this and the previous results.

In what follows, ****to simplify**** we let $\tau = 1$.

5. Deduce from the previous results that for any x,

$$\nabla f_1(x) + \nabla (f^*)_1(x) = x$$

(here $(f^*)_1$ is the Moreau-Yosida regularization with parameter $\tau = 1$ of the conjugate f^* of f, and *not!!* the conjugate of f_1).

One has

$$\nabla f_1(x) + \nabla (f^*)_1(x) = (x - \text{prox}_f(x)) + (x - \text{prox}_{f^*}(x))$$
$$= x - [x - \text{prox}_f(x) - \text{prox}_{f^*}(x)] = x$$

thanks again to Moreau's identity.

6. Therefore by integration one finds: $f_1(x) + (f^*)_1(x) = \|x\|^2/2 + C$ for some constant C, with $C = f_1(0) + (f^*)_1(0)$. Let $y = \text{prox}_f(0)$, $z = \text{prox}_{f^*}(0)$. Show that $y = -z$. Deduce that $C = 0$.

First, if $y = \text{prox}_f(0), z = \text{prox}_{f^*}(0)$, with Moreau's identity we have $0 = y + z$, that is $y = -z$. By definition, $y + \partial f(y) \ni 0$, that is, $z = -y \in \partial f(y)$ and $y \in \partial f^*(z)$. In particular, $f(y) + f^*(z) = y \cdot z = -\|y\|^2$ so that

$$C = f_1(0) + (f_1^*)(0) = \frac{\|y\|^2}{2} + f(y) + \frac{\|z\|^2}{2} + f^*(z) = 0.$$

3. Optimization: Nonlinear gradient descent

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$, possibly different from the standard Euclidean 2-norm: for instance, $\|x\| = \sum_{i=1}^n |x_i|$ (the 1-norm), or $\|x\| = \max\{|x_1|, \dots, |x_n|\}$ (the ∞-norm). (A norm is any convex, 1-homogeneous, even, function with values in $[0, +\infty[$ and which is strictly positive except in 0.) We define the dual (or polar) norm $\|y\|_*$ by the formula:

$$\|y\|_* = \sup_{x : \|x\| \le 1} y \cdot x$$

where $y \cdot x$ is the standard dot product $y \cdot x = \sum_{i=1}^n y_i x_i$. In particular, one has $y \cdot x \le \|y\|_* \|x\|$ for all y, x. (The "right" point of view should be that y is in the dual E^* of $E = \mathbb{R}^n$ (which is also $E^* = \mathbb{R}^n$) and that $y \cdot x$ is the evaluation of the linear form y at x. Then, $\|\cdot\|$ is the norm on E while $\|\cdot\|_*$ is the norm on E^*.)

1. Show that if $\mathcal{F}(x) := \|x\|^2/2$, then its convex conjugate is $\mathcal{F}^*(y) = \|y\|_*^2/2$. Deduce that the dual norm of $\|\cdot\|_*$ is $\|\cdot\|$, that is, for all x,

$$\|x\| = \sup_{y:\|y\|_*\le 1} y\cdot x.$$

One has

$$\mathcal{F}^*(y) = \sup_x x\cdot y - \frac{\|x\|^2}{2} = \sup_{t\ge 0,\|x\|=t} x\cdot y - \frac{t^2}{2}$$
$$= \sup_{t\ge 0}\left(\sup_{x:\|x\|=t} x\cdot y\right) - \frac{t^2}{2} = \sup_{t\ge 0} t\|y\|_* - \frac{t^2}{2} = \frac{\|y\|_*^2}{2}.$$

Hence in particular if we introduce the dual norm $\|\cdot\|_{**}$, the same computation will show that $\mathcal{F}^{**}(x) = \|x\|_{**}^2/2$. Since $\mathcal{F}$ is obviously convex, lsc., then $\mathcal{F}^{**} = \mathcal{F}$ so that $\|x\|_{**} = \|x\|$.

2. Compute $\|\cdot\|_*$ in the following cases:

i. 1-norm: $\|x\| = \sum_{i=1}^n |x_i|$;

ii. 2-norm: $\|x\| = \left(\sum_{i=1}^n |x_i|^2\right)^{\frac{1}{2}} = \sqrt{x\cdot x}$.

(i) There are many ways to evaluate this dual norm, for instance one has:

$$\sum_{i=1}^n |x_i| = \sum_{i=1}^n \sup_{|y_i|\le 1} y_i x_i = \sup_{|y_i|\le 1\,\forall i} \sum_{i=1}^n y_i x_i$$

which shows that $\|\cdot\|$ is the dual norm of the norm $y\mapsto \max_{i=1,\dots n}|y_i|$. We deduce that this is also its dual norm.

(ii) For the 2-norm, we know that $x\cdot y \le \|x\|\|y\| \le \|x\|$ if $\|y\|\le 1$, and choosing $y = x/\|x\|$, we have equality, showing that $\|\cdot\|_* = \|\cdot\|$.

Now, we consider a function f whose differential is L-Lipschitz in the normed space $E = (\mathbb{R}^n, \|\cdot\|)$, which means precisely that for any $x, x' \in \mathbb{R}^n$,

$$\|\nabla f(x) - \nabla f(x')\|_* \le L\|x - x'\|$$

where $\nabla f(x) \in E^*$ is the vector of partial derivatives $(\partial f/\partial x_i)_{i=1}^n$.

3. Show that, as in the Euclidean case, one has for $x, x' \in E$,

$$f(x') \le f(x) + \nabla f(x)\cdot(x' - x) + \frac{L}{2}\|x - x'\|^2.$$

This follows as usual from

$$
\begin{aligned}
f(x') = f(x) + \int_0^1 \nabla f(x + t(x'-x)) \cdot (x'-x)dt \\
= f(x) + \nabla f(x)\cdot(x'-x) + \int_0^1 \left(\nabla f(x+t(x'-x)) - \nabla f(x)\right)\cdot(x'-x)dt \\
\le f(x) + \nabla f(x)\cdot(x'-x) + \int_0^1 \left\|\nabla f(x+t(x'-x)) - \nabla f(x)\right\|_* \|x'-x\|dt \\
\le f(x) + \nabla f(x)\cdot(x'-x) + L\int_0^1 \|t(x'-x)\|\,\|x'-x\|dt \\
= f(x) + \nabla f(x)\cdot(x'-x) + \frac{L}{2}\|x'-x\|^2
\end{aligned}
$$

We want to define a “gradient descent” method in the norms $\|\cdot\|, \|\cdot\|_*$. We choose $x^0 \in E$. Given x^k, $k \ge 0$, we define $x^{k+1} = x^k - p^k$ and we find the descent direction p^k as follows: we observe that

$$
f(x^{k+1}) \le f(x^k) - \nabla f(x^k)\cdot p^k + \frac{L}{2}\|p^k\|^2.
$$

Then, we choose a p^k which minimizes the expression in the right-hand side of this equation.

4. Show that one has to choose $p^k \in \partial\mathcal{F}^*(\frac{1}{L}\nabla f^*)$, and that one obtains, for such a choice:

$$
f(x^{k+1}) \le f(x^k) - \frac{1}{2L}\|\nabla f(x^k)\|_*^2.
$$

One has to find p^k which minimizes

$$
\min_{p^k} -\nabla f(x^k)\cdot p^k + \frac{L}{2}\|p^k\|^2 = -L\max_{p^k} \tfrac{1}{L}\nabla f(x^k)\cdot p^k - \mathcal{F}(p^k) = -L\mathcal{F}^*(\tfrac{1}{L}\nabla f(x^k)) = -\tfrac{1}{2L}\|\nabla f(x^k)\|_*^2.
$$

A minimizer satisfies $\frac{1}{L}\nabla f(x^k) - \partial\mathcal{F}(p^k) \ni 0$, that is, $\frac{1}{L}\nabla f(x^k) \in \partial\mathcal{F}(p^k)$, or equivalently, $p^k \in \partial\mathcal{F}^*(\frac{1}{L}\nabla f(x^k))$.

5. We assume the set $X = \{f \le f(x^0)\}$ is bounded, and observe that $x^k \in X$ for any $k \ge 1$. We also assume that f has a minimizer x^* (obviously, $x^* \in X$). As in the Lecture notes, show that for all $k \ge 0$:

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

and:

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

where $C = \max_{x\in X}\|x - x^*\|^2$.

See Lemma 2.6, Thm. 2.7 in the Lecture notes.

4. Optimization: Polyak's subgradient descent method

In his book from 1987, Boris T. Polyak suggests the following variant of the subgradient descent method, which can be used whenever the optimal value of a problem is known. One consider a convex function $f : \mathbb{R}^n \to \mathbb{R}$ ($\mathrm{dom}\, f = \mathbb{R}^n$), which has a non empty set of minimizer(s) X^*, and we assume that the minimal value f^* is known. For instance:

$$f(x) = \max_{1\le i\le p} |a^i \cdot x - b_i|$$

where $a^i \in \mathbb{R}^n, b \in \mathbb{R}^p$ are such that $a^i \cdot x = b_i$, $i = 1, \dots, p$ has a solution: in that case $f^* = 0$.

Then, one chooses $x^0 \in \mathbb{R}^n$ and computes a subgradient descent method by picking for all $k \ge 0$, $p^k \in \partial f(x^k)$ and

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

1. Show that, if $x^* \in X^*$ is any minimizer,

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

What do we deduce for the sequence $(x^k)_{k\ge 0}$?

We write:

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^*\|^2 - 2\frac{(f(x^k) - f^*)p^k \cdot (x^k - x^*)}{\|p^k\|^2} + \frac{(f(x^k) - f^*)^2}{\|p^k\|^2}$$
$$\le \|x^k - x^*\|^2 - \frac{(f(x^k) - f^*)^2}{\|p^k\|^2}.$$

because simply, $p^k \cdot (x^k - x^*) \ge f(x^k) - f(x^*)$ since $p^k \in \partial f(x^k)$. In particular, the sequence of iterates is bounded (and has converging subsequences).

2. Why is it true that $C := \sup_k \|p^k\| < +\infty$? Deduce that $\sum_{k=0}^{\infty} (f(x^k) - f^*)^2 < +\infty$.

The first question shows that x^k is a bounded sequence. By assumption, f is convex, defined on $\mathbb{R}^n$, so we know that it is locally Lipschitz, and its subgradients are locally bounded. This shows that the $\sup_k \|p^k\|$ must be finite.

Hence, letting $C \ge \sup_k \|p^k\|$, one finds

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

and summing from $k = 0$ to n we get

$$\sum_{k=0}^{n} (f(x^k) - f^*)^2 + C^2\|x^{n+1} - x^*\|^2 \le C^2\|x^0 - x^*\|^2.$$

for any $n \ge 0$, and in particular one can let $n \to \infty$.

3. We deduce that $f(x^k) \to f^*$. Show that there is one minimizer $x^* \in X^*$, such that $x^k \to x^*$.

We have seen that (x^k) is bounded so there exists x^* such that a subsequence (x^{k_l}) converges to x^*. Then, $f(x^{k_l}) \to f(x^*) = f^*$, so that x^* is a minimizer. Now, from the first question, we deduce that $\|x^k - x^*\|^2$ is a non-increasing sequence. (In particular it has a limit.) Since it goes to zero along the subsequence (k_l), then it must go to zero so that $x^k \to x^*$.

4. We now assume that the function is "α-sharp", $\alpha \geq 1$, meaning that for some $\gamma > 0$,

$$f(x) - f^* \geq \gamma \mathrm{dist}\,(x, X^*)^\alpha.$$

Show that

$$\mathrm{dist}\,(x^{k+1}, X^*)^2 \leq \mathrm{dist}\,(x^k, X^*)^2 - \frac{\gamma^2 \mathrm{dist}\,(x^k, X^*)^{2\alpha}}{C^2}.$$

In case $\alpha = 1$ (which is the situation in the example mentioned in the introduction of this exercise), what do we deduce?

Let $x \in X^*$ be the projection of x^k on the solution set X^* (which is closed, convex) so that $\|x^k - x\| = \mathrm{dist}\,(x^k, X^*)$. Then:

$$\mathrm{dist}\,(x^{k+1}, X^*)^2 \leq \|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 - \frac{(f(x^k) - f^*)^2}{C^2} \leq \mathrm{dist}\,(x^k, X^*)^2 - \frac{\gamma^2 \mathrm{dist}\,(x^k, X^*)^{2\alpha}}{C^2}.$$

When $\alpha = 1$ this reduces to $\mathrm{dist}\,(x^{k+1}, X^*)^2 \leq \mathrm{dist}\,(x^k, X^*)^2(1 - \frac{\gamma^2}{C^2})$, one has a geometric (linear) convergence to the solution set (note that the actual convergence to x^* could be slower).

5. We consider a sequence a_k, $k \geq 0$, with for all $k \geq 0$, $a_k \geq 0$ and $a_{k+1} \leq a_k - c^{-1} a_k^{1+\beta}$, $c > 0$, $\beta > 0$. Show that:

$$a_k \leq \left(\frac{c}{\max\{\beta, 1\}(k+1)} \right)^{1/\beta}.$$

Hint: introduce $b_k := a_k^\beta$, and depending on whether $\beta \geq 1$ or $\beta \leq 1$, try to show that $b_k \leq b_k - c'^{-1} b_k^2$ for some c' (depending on c, β). Use then the Lecture notes.

Since $a_k \leq a_k(1 - c^{-1} a_k^\beta)$, we can write that $b_{k+1} \leq b_k(1 - c^{-1} b_k)^\beta$. If $\beta \leq 1$, we use (concavity) that $(1 - c^{-1} b_k)^\beta \leq 1 - \beta c^{-1} b_k$. If $\beta \geq 1$, we use rather than $(1 - c^{-1} b_k)^\beta \leq 1 - c^{-1} b_k$. In both cases, we get

$$b_{k+1} \leq b_k - \frac{\max\{\beta, 1\}}{c} b_k^2$$

and from Lemma 2.6 in the notes, we deduce $b_k \leq c/(\max\{\beta, 1\}(k+1))$. The conclusion follows.

6. Deduce the rate of convergence for the distance from x^k to the set X^* in case $\alpha > 1$.

From the two previous questions, we deduce that

$$\mathrm{dist}\,(x^k, X^*)^2 \leq \left(\frac{C^2}{\gamma^2 \max\{\alpha - 1, 1\}} \frac{1}{k+1} \right)^{\frac{1}{\alpha - 1}}.$$