Continuous Optimization
Introduction à l'optimisation continue
Contrôle
(14 janvier 2020)

Exercise I: conjugate function and "prox" operator

We consider the function $f:\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$, $d\geq 1$, defined by

$$f(x)=\begin{cases}-\ln(1-\|x\|) & \text{if } \|x\|<1\\ +\infty \text{ else.}\end{cases}$$

1. Show that f is convex, lower-semicontinuous (lsc).

First, f is continuous in $\{\|x\|<1\}$, and goes to $+\infty$ when $\|x\|\to 1$, so it is lower-semicontinuous. Then, given x,y and $t\in]0,1[$ and assuming $\|x\|,\|y\|<1$ (otherwise $tf(x)+(1-t)f(y)=+\infty$ and there is nothing to prove), one has:

$$\begin{aligned}&-\|tx+(1-t)y\|\geq -t\|x\|-(1-t)\|y\| \quad \text{so that}\\ &f(tx+(1-t)y)=-\ln(-\|tx+(1-t)y\|)\leq -\ln(-t\|x\|-(1-t)\|y\|)\end{aligned}$$

where we use that $\|\cdot\|$ is convex and $-\ln$ is decreasing. Then, as $-\ln$ is convex,

$$-\ln(-t\|x\|-(1-t)\|y\|)\leq -t\ln\|x\|-(1-t)\ln\|y\|=tf(x)+(1-t)f(y).$$

2. Show that $f(x)\geq\|x\|$. Deduce that $\partial f(0)\supseteq B(0,1)=\{y:\|y\|\leq 1\}$.

One has $-\ln(1+s)\geq -\ln 1-s=-s$ by convexity of $-\ln$. Hence $f(x)\geq\|x\|$. In particular $f(x)\geq y\cdot x$ for any y with $\|y\|\leq 1$, as $f(0)=0$ this shows that $\partial f(0)\supseteq B(0,1)$.

3. Show that for $x\neq 0$, $\|x\|<1$,

$$\nabla f(x)=\frac{x}{\|x\|(1-\|x\|)}.$$

Deduce that for any p with $\|p\|>1$, one can compute $x\in B(0,1)$ with $\nabla f(x)=p$. Deduce an expression for $f^*(p)$ (recall Legendre-Fenchel's identity). What is $f^*(p)$ for $\|p\|\leq 1$?

The formula is a simple differentiation, as f is C^∞ in $B(0,1)\setminus\{0\}$. Then, for $\|p\|>1$, if $p=\nabla f(x)$ one sees that x and p must be aligned and taking the norms, $\|p\|=1/(1-\|x\|)$ so that $\|x\|=1-1/\|p\|\in(0,1)$ and $x=(1-1/\|p\|)p/\|p\|$.

Using that $f(x)+f^*(p)=\langle p,x\rangle$ for $p=\nabla f(x)$, we deduce that

$$f^*(p)=x\cdot p-f(x)=\|p\|-1+\ln(1-(1-1/\|p\|))=\|p\|-\ln\|p\|-1.$$

For $\|p\|\leq 1$, on the other hand, we have $p\in\partial f(0)$ so that $f^*(p)=0\cdot p-f(0)=0$.

4. We want to compute $x = \mathrm{prox}_{\tau f}(\bar{x})$ for a given $\bar{x} \in \mathbb{R}^d$ and $\tau > 0$. Write the equation satisfied by x and show that (i) x and $\bar{x}$ are colinear; (ii) $x = 0$ if $\|\bar{x}\| \le \tau$; (iii) if $\|\bar{x}\| > \tau$, then $\rho = \|x\|$ satisfies a second order equation which has two positive solutions. Using then that $\rho < 1$ and $\rho \le \|\bar{x}\|$, give the right answer.

The problem to solve is

$$\min_x \frac{\|x - \bar{x}\|^2}{2\tau} + f(x)$$

and is solved by a unique point which satisfies

$$x - \bar{x} + \tau \partial f(x) \ni 0.$$

It means, either $\|\bar{x}\| \le \tau$ and $x = 0$ is a solution, or, for $\|\bar{x}\| > \tau$, $x \neq 0$, $\|x\| < 1$ (as $f(x)$ is finite) and

$$x + \frac{\tau x}{\|x\|(1 - \|x\|)} = \bar{x} \;\Leftrightarrow\; (\|x\|(1 - \|x\|) + \tau)x = \|x\|(1 - \|x\|)\bar{x}.$$

In particular we see that $x = \rho\bar{x}/\|\bar{x}\|$ for some $\rho > 0$, and taking the norm we have

$$\|x\|(1 - \|x\|) + \tau = (1 - \|x\|)\|\bar{x}\|.$$

Denoting $\bar{\rho} := \|\bar{x}\| > \tau$, $\rho \in (0, 1)$ must solve

$$\rho^2 - (1 + \bar{\rho})\rho + \rho - \tau = 0$$

which has two solutions

$$\rho^+ = \frac{1 + \bar{\rho} + \sqrt{(1 - \bar{\rho})^2 + 4\tau}}{2}, \quad \rho^- = \frac{1 + \bar{\rho} - \sqrt{(1 - \bar{\rho})^2 + 4\tau}}{2}$$

with $\rho^+ + \rho^- = 1 + \bar{\rho}$ and $\rho^+\rho^- = \bar{\rho} - \tau > 0$. Observe that $\rho^+ > (1 + \bar{\rho} + |1 - \bar{\rho}|)/2$ so that in case $\bar{\rho} \ge 1$, $\rho^+ > \bar{\rho} \ge 1$ (hence $\rho^- = 1 + \bar{\rho} - \rho^+ < 1$), and in case $\bar{\rho} < 1$, $\rho^+ > 1$ (hence $\rho^- < \bar{\rho} < 1$). Hence the right solution (giving the norm of x) is ρ^-, and one can write

$$x = \frac{1 + \|\bar{x}\| - \sqrt{(1 - \|\bar{x}\|)^2 + 4\tau}}{2}\bar{x}.$$

Exercise II: The "Extragradient" method

Let X be a Hilbert space and A a maximal-monotone operator. We assume that A is defined everywhere in X and L-Lipschitz[1] (in particular, it implies that Ax is just one point, denoted Ax, for any x).

One wants to find $x^* \in X$ with $Ax^* = 0$. Let $S = \{x \in X : Ax = 0\}$ and assume that $S \neq \emptyset$.

[1] A Lipschitz maximal monotone must be defined everywhere, thanks to Kirszbraun-Valentine's theorem.

1. We first consider the elementary algorithm $x^{k+1} = x^k - \tau A x^k$, for $\tau > 0$, and $x^0 \in X$ given. Recall from the lecture notes the standard conditions on A and τ which guarantee that the sequence $(x^k)_{k\geq 0}$ (weakly) converges to a point $x^* \in S$.

This is in the lecture notes. One needs A to be co-coercive: $\langle Ax - Ay, x - y\rangle \geq \mu\|Ax - Ay\|^2$, and that $0 < \tau < 2\mu$.

2. In general it is not clear that the algorithm in **1.** will converge. Consider for instance $X = \mathbb{R}^2$,

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Show that A satisfies the assumptions at the beginning of the exercise. Evaluate $\|x^{k+1}\|$ in function of x^k. Deduce that the algorithm always diverges if $x^0 \neq 0$.

A is Lipschitz (obvious), monotone since $\langle Ax - Ay, x - y\rangle = \langle A(x - y), x - y\rangle = 0$ for any x, y, maximal: if (z, p) are such that $\langle Ax - p, x - z\rangle \geq 0$ for all x, for $x = z + ty$, $y \in X$, $t > 0$, one finds that $0 \leq \langle Az + tAy - p, ty\rangle = t\,\langle Az - p, y\rangle$ so that $\langle Az - p, y\rangle \geq 0$ for all y (we have used $\langle tAy, ty\rangle = 0$). It implies that $p = Az$.

Then, given $(x, y) \in \mathbb{R}^2$, we compute

$$\left\|(I - \tau A)\begin{pmatrix} x \\ y \end{pmatrix}\right\|^2 = \left\|\begin{matrix} x - \tau y \\ y + \tau x \end{matrix}\right\|^2 = (x - \tau y)^2 + (y + \tau x)^2 = (1 + \tau^2)(x^2 + y^2).$$

Thus, $\|x^k\| = \sqrt{1 + \tau^2}^k\|x^0\|$ which shows the claim.

The extragradient algorithm (Korpelevich, 1976). One Considers the following two-steps algorithm, known as the "extra-gradient" method (as it computes an evaluation of A at an extrapolated point). Given $x^0 \in X$, $\tau > 0$, one lets for $k \geq 0$:

$$\begin{cases} y^k = x^k - \tau A x^k \\ x^{k+1} = x^k - \tau A y^k. \end{cases}$$

One wants to show that with this correction, the algorithm converges for τ well chosen.

3. We consider the setting of question **2.**. Compute the matrix $B := I - \tau A(I - \tau A)$. Evaluate again $\|x^{k+1}\|$, and show that if $\tau < 1$ the algorithm converges.

We have

$$B = I - \tau A(I - \tau A) = I - \begin{pmatrix} 0 & \tau \\ -\tau & 0 \end{pmatrix}\begin{pmatrix} 1 & -\tau \\ \tau & 1 \end{pmatrix} = I - \begin{pmatrix} \tau^2 & \tau \\ -\tau & \tau^2 \end{pmatrix} = \begin{pmatrix} 1 - \tau^2 & -\tau \\ \tau & 1 - \tau^2 \end{pmatrix}$$

One has $x^{k+1} = x^k - \tau A y^k = x^k - \tau A(x^k - \tau A x^k) = (I - \tau A(I - \tau A))x^k = Bx^k$. Given $(x, y) \in \mathbb{R}^2$, $\|B(x, y)^T\|^2 = ((1-\tau^2)x - \tau y)^2 + (\tau x + (1-\tau^2)y)^2 = ((1-\tau^2)^2 + \tau^2)\|(x, y)^T\|^2$. Using that $((1 - \tau^2)^2 + \tau^2) = 1 - \tau^2 + \tau^4 = 1 - \tau^2(1 - \tau^2)$ we find that if $0 < \tau < 1$,

$$\|x^k\| = \sqrt{1 - \tau^2(1 - \tau^2)}^k\|x^0\| \to 0$$

as $k \to \infty$ (with a linear rate) so that the algorithm converges.

4. We now return to the general case where A is a maximal monotone operator, L-Lipschitz, in a Hilbert space X. Show that if $0 < \tau L < 1$ and if x is a fixed point of the algorithm (meaning that if $x^k = x$, then $x^{k+1} = x$ as well), then $x \in S$.

We have, letting $y = x - \tau A x$, that $x - \tau A y = x$, that is $Ay = 0$ ($y \in S$). Then, $\|Ax\| = \|Ax - Ay\| \leq L\|x - y\|$ so that $\|\tau A x\| \leq \tau L \|x - y\| = \tau L \|\tau A x\|$. Since $0 < \tau L < 1$ this is possible only if $Ax = 0$, that is $x \in S$.

5. Let $x^* \in S$. Using that A is monotone, first show (using the second line of the algorithm) that

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - 2\tau \left\langle Ay^k, x^k - y^k \right\rangle + \tau^2 \|Ay^k\|^2 \\ &= \|x^k - x^*\|^2 - 2\left\langle x^k - x^{k+1}, x^k - y^k \right\rangle + \tau^2 \|x^k - x^{k+1}\|^2.\end{aligned}$$

One has

$$\|x^{k+1} - x^*\|^2 = \|x^k - x^*\|^2 - 2\tau \left\langle Ay^k, x^k - x^* \right\rangle + \tau^2 \|Ay^k\|^2.$$

We write (recall that $Ax^* = 0$)

$$\begin{aligned}\left\langle Ay^k, x^k - x^* \right\rangle &= \left\langle Ay^k, x^k - y^k \right\rangle + \left\langle Ay^k, y^k - x^* \right\rangle \\ &= \left\langle Ay^k, x^k - y^k \right\rangle + \left\langle Ay^k - Ax^*, y^k - x^* \right\rangle \geq \left\langle Ay^k, x^k - y^k \right\rangle.\end{aligned}$$

Hence,

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\tau \left\langle Ay^k, x^k - y^k \right\rangle + \tau^2 \|Ay^k\|^2.$$

We conclude using again that $\tau A y^k = x^k - x^{k+1}$.

6. Deduce, using now that A is L-Lipschitz, that

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - (1 - \tau^2 L^2)\|y^k - x^k\|^2.$$

One has

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &\leq \|x^k - x^*\|^2 - 2\left\langle x^k - x^{k+1}, x^k - y^k \right\rangle + \|x^{k+1} - x^k\|^2 \\ &= \|x^k - x^*\|^2 + \|x^{k+1} - y^k\|^2 - \|x^k - y^k\|^2 \\ &= \|x^k - x^*\|^2 + \tau^2 \|Ay^k - Ax^k\|^2 - \|x^k - y^k\|^2 \\ &\leq \|x^k - x^*\|^2 + \tau^2 L^2 \|y^k - x^k\|^2 - \|x^k - y^k\|^2\end{aligned}$$

7. We now assume $0 < \tau < 1/L$. What can we say of the sequence $(x^k)_{k\geq 0}$? Of the sequence $(\|x^k - x^*\|)_{k\geq 0}$ for $x^* \in S$? Of the sequence $(x^k - y^k)_{k\geq 0}$?

The first is bounded (and hence has weakly converging subsequences), the second is decreasing (and hence has a limit), the last must go to zero, as the series

$$(1-\tau^2L^2)\sum_{k=0}^{n}\|y^k - x^k\|^2 + \|x^{n+1} - x^*\|^2 \leq \|x^0 - x^*\|^2$$

is bounded.

8. As in the lecture notes, we denote $m(x^*) = \lim_{k\to\infty}\|x^k - x^*\|$, for $x^* \in S$. Let $\bar{x}$ be the (weak) limit of a subsequence $(x^{k_l})_l$. Show that $Ax^{k_l} \to 0$ (strongly). Using that A is maximal-monotone, deduce that $A\bar{x} = 0$, that is $\bar{x} \in S$. Deduce from Opial's lemma that x^k converges (weakly in X) to $\bar{x}$.

As we saw, (x^k) is bounded and therefore has weakly converging subsequences. One has here in addition that $Ax^{k_l} = (x^{k_l} - y^{k_l})/\tau \to 0$ in norm (strongly). If $z \in X$ by monotonicity one has $0 \leq \langle Az - Ax^{k_l}, z - x^{k_l}\rangle \to \langle Az, z - \bar{x}\rangle$. We deduce that for all $z \in X$, $\langle Az, z - \bar{x}\rangle \geq 0$ and this shows that $A\bar{x} = 0$ (precisely that $0 \in A\bar{x}$, but as we know already $A\bar{x}$ is a set with exactly one element). In particular $\|x^{k_l} - \bar{x}\| \to m(\bar{x})$.

Now, Opial's lemma shows that $m(\bar{x}) < m(x^*)$ for all $x^* \in S \setminus \{\bar{x}\}$. This means that if $\bar{x}' \in S$ is the limit of any other converging subsequence of (x^k), we must have $\bar{x}' = \bar{x}$. Hence $x^k \rightharpoonup \bar{x}$.

Exercise III: A nonlinear proximal point algorithm

We consider X a Hilbert space and a strictly convex lower-semicontinuous (lsc) function $\psi : X \to \mathbb{R}\cup\{+\infty\}$ such that the interior of $\operatorname{dom}\psi$, denoted D, is not empty, $\overline{D} = \operatorname{dom}\psi$, $\psi \in C^1(D)\cap C^0(\overline{D})$, and $\partial\psi(x) = \emptyset$ for all $x \notin D$. In other words, $\partial\psi(x)$ is either $\emptyset$ (if $x \notin D$), or a singleton $\{\nabla\psi(x)\}$ (if $x \in D$). We define the "Bregman distance associated to ψ", denoted $D_\psi(x,y)$, as, for $y \in D$ and $x \in X$,

$$D_\psi(x,y) := \psi(x) - \psi(y) - \langle\nabla\psi(y), x - y\rangle\,.$$

1. Show that $D_\psi(x,y) \geq 0$, and that $D_\psi(x,y) = 0 \Rightarrow y = x$. What further estimate can we write if in addition ψ is strongly convex? Why is D_ψ *not* a distance in the classical sense?

$D_\psi(x,y) \geq 0$ because ψ is convex. If $D_\psi(x,y) = 0$, then for $t \in [0,1]$, $\psi(tx+(1-t)y) \leq t\psi(x) + (1-t)\psi(y) = \psi(y) + t\langle\nabla\psi(y), x-y\rangle \leq \psi(y + t(x-y))$ Hence ψ is affine on $[x,y]$, which is a contradiction to ψ being strictly convex unless $x = y$. Finally, if ψ is γ-convex one has for x, y:

$$\psi(x) \geq \psi(y) + \langle\nabla\psi(y), x-y\rangle + \frac{\gamma}{2}\|x-y\|^2$$

so that $D_\psi(x,y) \geq (\gamma/2)\|x-y\|^2$.

On the other hand there is no reason to have $D_\psi(x,y) = D_\psi(y,x)$ in general.

2. Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be convex, lsc. Assume $\lim_{|x|\to\infty} f(x) = +\infty$. Let $\tau > 0$. Let $\bar{x} \in D$. Show that there exist a minimizer $\hat{x}$ of

$$\min_x \frac{1}{\tau} D_\psi(x, \bar{x}) + f(x).$$

Show that then $\hat{x} \in D$ and the "Euler-Lagrange" equation (or Fermat's rule),

$$\nabla\psi(\hat{x}) - \nabla\psi(\bar{x}) + \tau\partial f(\hat{x}) \ni 0 \tag{EL}$$

The function in the minimization problem is convex, lsc, and goes to infinity when $|x| \to \infty$. Hence it is also weakly convex and has a minimizer. (Unique as ψ is strictly convex.) Moreover, as ψ is C^1 in an open set, one has $\partial(D_\psi(\cdot, \bar{x})/\tau + f) = \partial D_\psi(\cdot, \bar{x})/\tau + \partial f = \nabla D_\psi(\cdot, \bar{x})/\tau + \partial f$. Hence one derives the equation by observing that one should have $0 \in \partial(D_\psi(\cdot, \bar{x})/\tau + f)(\hat{x})$. In particular $x \in D$ otherwise the subgradient would be empty.

3. Deduce from (EL) and the definition of a subgradient the "three point relationship": for any $x \in X$,

$$\frac{1}{\tau} D_\psi(x, \bar{x}) + f(x) \geq \frac{1}{\tau} D_\psi(\hat{x}, \bar{x}) + f(\hat{x}) + \frac{1}{\tau} D_\psi(x, \hat{x}). \tag{$3P$}$$

We have

$$-\nabla\psi(\hat{x}) + \nabla\psi(\bar{x}) \in \tau\partial f(\hat{x})$$

so that for any x,

$$f(x) \geq f(\hat{x}) + \frac{1}{\tau} \langle \nabla\psi(\bar{x}) - \nabla\psi(\hat{x}), x - \hat{x} \rangle .$$

Hence

$$\begin{aligned}
f(x) + \frac{1}{\tau} D_\psi(x, \bar{x}) &\geq f(\hat{x}) + \frac{1}{\tau} \left(\langle \nabla\psi(\bar{x}) - \nabla\psi(\hat{x}), x - \hat{x} \rangle + \psi(x) - \psi(\bar{x}) - \langle \nabla\psi(\bar{x}), x - \bar{x} \rangle \right) \\
&= f(\hat{x}) + \frac{1}{\tau} \left(- \langle \nabla\psi(\hat{x}), x - \hat{x} \rangle + \psi(x) - \psi(\bar{x}) - \langle \nabla\psi(\bar{x}), \hat{x} - \bar{x} \rangle \right) \\
&= f(\hat{x}) + \frac{1}{\tau} \left(\psi(x) - \psi(\hat{x}) - \langle \nabla\psi(\hat{x}), x - \hat{x} \rangle + \psi(\hat{x}) - \psi(\bar{x}) - \langle \nabla\psi(\bar{x}), \hat{x} - \bar{x} \rangle \right)
\end{aligned}$$

which shows $(3P)$.

4. We consider the "nonlinear proximal point" algorithm: $x^0 \in D$,

$$x^{k+1} = \arg\min_x \frac{1}{\tau} D_\psi(x, x^k) + f(x).$$

Using $(3P)$, show that $f(x^k)$ is non-increasing. Then, show that for any $x \in \overline{D}$,

$$f(x^k) - f(x) \leq \frac{1}{k\tau} D_\psi(x, x^0)$$

If we choose $x = \bar{x} = x^k$ and then $\hat{x} = x^{k+1}$ in $(3P)$ we find

$$f(x^{k+1}) + \frac{1}{\tau}\left(D_\psi(x^{k+1}, x^k) + D_\psi(x^k, x^{k+1})\right) \le f(x^k)$$

so that $f(x^k)$ must be nonincreasing. (Moreover if $f(x^k) = f(x^{k+1})$ then $x^{k+1} = x^k$, it is the minimizer of f in D, as then $0 \in \partial f(x^k)$).

If we choose $\bar{x} = x^k$, x arbitrary and then $\hat{x} = x^{k+1}$ in $(3P)$ we find

$$f(x^{k+1}) - f(x) + \frac{1}{\tau}\left(D_\psi(x^{k+1}, x^k) + D_\psi(x, x^{k+1})\right) \le \frac{1}{\tau} D_\psi(x, x^k)$$

Summing this for $k = 0, \dots, n-1$ and using that $f(x^k)$ is decreasing gives

$$n(f(x^n) - f(x)) + \frac{1}{\tau} D_\psi(x, x^n) \le \frac{1}{\tau} D_\psi(x, x^0).$$

which yields the requested inequality.

5. Assume $x^k \to x^*$ weakly: what is x^*?

One has $f(x^*) \le \liminf_k f(x^k)$ as f is convex, lsc (hence also weakly lsc). Hence $f(x^*) - f(x) \le 0$ for any $x \in \overline{D}$. Moreover as $x^k \in D$, $x^* \in \overline{D}$ (again we use that a closed convex set is weakly closed). Hence x^* is a minimizer of f in D.

6. We assume in addition that there exists $\gamma > 0$ such that $h = f - \gamma\psi$ is convex. Show (using h) that $(3P)$ can be improved into:

$$\frac{1}{\tau} D_\psi(x, \bar{x}) + f(x) \ge \frac{1}{\tau} D_\psi(\hat{x}, \bar{x}) + f(\hat{x}) + \frac{1 + \gamma\tau}{\tau} D_\psi(x, \hat{x}). \tag{$3P_\gamma$}$$

Hint: write that $f(x) = h(x) + \gamma\psi(x) = (h(x) + \gamma[\psi(\bar{x}) + \langle \nabla\psi(\bar{x}), x - \bar{x}\rangle]) + \gamma D_\psi(x, \bar{x})$ and use $(3P)$ after having added $(1/\tau) D_\psi(x, \bar{x})$; or write (EL) using that $\partial f(\hat{x}) = \partial h(\hat{x}) + \gamma \nabla\psi(\hat{x})$ and work as in the proof of $(3P)$ in **3.**.

Let us use that $f(x) = h(x) + \gamma\psi(x) = (h(x) + \gamma[\psi(\bar{x}) + \langle \nabla\psi(\bar{x}), x - \bar{x}\rangle])$ and denote $h'(x) := h(x) + \gamma[\psi(\bar{x}) + \langle \nabla\psi(\bar{x}), x - \bar{x}\rangle]$. Then from $(3P)$ we have

$$\begin{aligned}\frac{1}{\tau} D_\psi(x, \bar{x}) + f(x) = \frac{1 + \tau\gamma}{\tau} D_\psi(x, \bar{x}) + h'(x) &\ge \frac{1 + \tau\gamma}{\tau} D_\psi(\hat{x}, \bar{x}) + h'(\hat{x}) + \frac{1 + \tau\gamma}{\tau} D_\psi(x, \hat{x}) \\ &= \frac{1}{\tau} D_\psi(\hat{x}, \bar{x}) + f(\hat{x}) + \frac{1 + \tau\gamma}{\tau} D_\psi(x, \hat{x})\end{aligned}$$

7. Deduce the "linear" rate of convergence for the algorithm:

$$f(x^{k+1}) - f(x^*) \le \frac{1}{(1 + \gamma\tau)^k} D_\psi(x^*, x^0)$$

where x^* is a minimizer of f in $\overline{D}$.

Now we have

$$f(x^{k+1}) - f(x) + \frac{1}{\tau}D_\psi(x^{k+1}, x^k) + \frac{1+\tau\gamma}{\tau}D_\psi(x, x^{k+1}) \leq \frac{1}{\tau}D_\psi(x, x^k).$$

Choosing $x = x^*$ a minimizer of f over $\overline{D}$, we have that $f(x^{k+1}) - f(x^*) \geq 0$ so that

$$\frac{1+\tau\gamma}{\tau}D_\psi(x^*, x^{k+1}) \leq \frac{1}{\tau}D_\psi(x^*, x^k).$$

It follows that $D_\psi(x^*, x^k) \leq 1/(1+\tau\gamma)^k D_\psi(x^*, x^0)$ and therefore also $f(x^{k+1}) - f(x^*)$ (using the inequality once more).

Exercise IV: convex homogeneous functions

Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ ($d \geq 1$) be convex, lsc, and positively 2-homogeneous: for any $x \in \mathbb{R}^d$, $t > 0$, $f(tx) = t^2 f(x)$. We want to show that $\sqrt{f}$ is also convex (1-homogeneous).

1. Show that f^* (the convex conjugate) is positively 2-homogeneous. (Hint: evaluate $f^*(ty)/t^2$ for $t > 0$.)

$$\frac{1}{t^2}f^*(ty) = \sup_x \frac{1}{t}y \cdot x - \frac{1}{t^2}f(x) = \sup_x y \cdot \frac{x}{t} - f(\frac{x}{t}) = f^*(y).$$

2. Let $h(x) = \sup_{f^*(y)\leq 1} y \cdot x$ be the conjugate of the characteristic function $\delta_{\{f^*(\cdot)\leq 1\}}$. Show that h is convex, one-homogeneous, non-negative.

h is trivially convex lsc as a sup of affine functions (or as the conjugate of $\delta_{\{f^*(\cdot)\leq 1\}}$). Also, for $t > 0$, $h(tx) = \sup_{\dots} y \cdot tx = t \sup_{\dots} y \cdot x = th(x)$ is trivial. As $f^*(0) \leq 0$ (f^* is lsc, $f^*(0) \leq \liminf_{t\to 0} f^*(tx) = 0$ for any $x \in \operatorname{dom} f^*$), $0 \in \{f^*(\cdot) \leq 1\}$ and $h \geq 0$.

3. Show that $f = h^2/4$, conclude.

$$f(x) = \sup_y x \cdot y - f^*(y) = \sup_{f^*(\eta)=1, t>0, y=t\eta} tx \cdot \eta - t^2 = \sup_{t>0} th(x) - t^2 = \frac{h(x)^2}{4}.$$

Hence $\sqrt{f} = h/2$ is a convex function.