Continuous (convex) optimisation

M2 - PSL / Dauphine / S.U.

Antonin Chambolle, CNRS, CEREMADE

Université Paris Dauphine PSL

Oct.-Dec. 2023

Lecture 2: fixed point iterations; convexity.

Contents

Krasnoselskii-Mann's convergence theorem

Let $T : \mathcal{X} \to \mathcal{X}$ ($\mathcal{X}$ is a Banach or Hilbert space) which is "nonexpansive" or 1 Lipschitz:

$$|Tx - Ty| \leq |x - y| \quad \forall\, x, y \in \mathcal{X}.$$

We consider the problem of finding a fixed point $Tx = x$.
If in addition it is ρ-Lipschitz with $\rho < 1$, then Picard's classical fixed point theorem shows that the iterates $x^k = T^k x^0$, $k \geq 1$, form a Cauchy sequence and therefore converge to a fixed point, necessarily unique. (This relies on the fact that the space is complete.)
If $\rho = 1$, then the iterations may, or may not, converge. For instance, if $Tx = -x$, then $T0 = 0$ but $x^k = (-1)^k x_0$ will never converge unless $x_0 = 0$.

Definition: Averaged operator

A nonexpansive operator is *averaged* if it is of the form

$$(1-\theta)I + \theta T$$

where $\theta \in [0,1)$ and T is 1-Lipschitz.

Remark: for $\theta = 0$, it is I, which is averaged but not very interesting.
Example: $x \mapsto x - \tau\nabla f(x)$ where f is convex, with L-Lipschitz gradient, and $0 \leq \tau < 2/L$.

Krasnoselskii-Mann's convergence theorem

Setting: T a nonexpansive operator in a *Hilbert* space $\mathcal{X}$.
$F = \{x \in \mathcal{X} : Tx = x\} \neq \emptyset$. $T_\theta := (1-\theta)I + \theta T$.

Theorem

Let $x_0 \in \mathcal{X}$, $0 < \theta < 1$, *and for* $k \geq 1$, $x^k = T_\theta^k x_0$. *Then there exists* $x \in F$, $x^k \rightharpoonup x$.

Krasnoselskii-Mann's convergence theorem

Proof: Step 1.: We remark that $x^* \in F$ if and only if $T_\theta x^* = x^*$. In particular, $|x^{k+1} - x^*| = |T_\theta x^k - T_\theta x^*| \le |x^k - x^*|$. We say that the sequence (x^k) is "Fejér"-monotone with respect to the set F. Then there exists, for any $x^* \in F$, $m(x^*) = \inf_k |x^k - x^*| = \lim_k |x^k - x^*|$. If $m(x^*) = 0$ for some x^* we are done, hence we may assume $m(x^*) > 0$ for all $x^* \in F$.

Step 2: We show that $x^{k+1} - x^k \to 0$: the operator T_θ is said to be "asymptotically regular". It follows from the "parallelogram identity". One has

$$x^{k+1} - x^* = (1-\theta)(x^k - x^*) + \theta(T_1 x^k - x^*).$$

Hence for all k:

$$\begin{aligned} |x^{k+1} - x^*|^2 = (1-\theta)|x^k - x^*|^2 + \theta|T_1 x^k - x^*|^2 - \theta(1-\theta)|T_1 x^k - x^k|^2 \\ \le |x^k - x^*|^2 - \frac{1-\theta}{\theta}|x^{k+1} - x^k|^2 \end{aligned}$$

from which one deduces that $\sum_k |x^{k+1} - x^k|^2 < \infty$, hence the result.

Krasnoselskii-Mann's convergence theorem

Remark: rate. In addition, one observes that the sequence $\frac{1-\theta}{\theta}|x^{k+1}-x^k|^2$ (which is nonincreasing) is controlled in the following way:

$$\frac{1-\theta}{\theta}(k+1)|x^{k+1}-x^k|^2 \le \frac{1-\theta}{\theta}\sum_{i=0}^{k}|x^{i+1}-x^i|^2 \le |x^0-x^*|^2 - |x^{k+1}-x^*|^2.$$

Using again $x^{k+1}-x^k = \theta(T_1x^k - x^k)$ we get:

$$|T_1x^k - x^k| \le \frac{|x^0-x^*|}{\sqrt{\theta(1-\theta)}\sqrt{k+1}}.$$

In fact, Cominetti-Soto-Vaisman showed recently (2014) that in **any normed space**:

$$|x^k - Tx^k| \le \frac{1}{\sqrt{\pi}}\frac{|x^0-x^*|}{\sqrt{\sum_{i=1}^k \theta_i(1-\theta_i)}}$$

(θ is allowed to vary at each step, assuming there exists a fixed point x^*, optimal constant $1/\sqrt{\pi}$.)

Krasnoselskii-Mann's convergence theorem

Step 3 relies on:

Opial's lemma

If in a Hilbert space $\mathcal{X}$ the sequence $(x_n)_n$ is weakly convergent to x_0 then for any $x \neq x_0$,

$$\liminf_n |x_n - x| > \liminf_n |x_n - x_0|$$

Proof:

$$|x_n - x|^2 = |x_n - x_0|^2 + 2\langle x_n - x_0, x_0 - x\rangle + |x_0 - x|^2.$$

Since $\langle x_n - x_0, x_0 - x\rangle \to 0$ by weak convergence, we deduce

$$\liminf_n |x_n - x|^2 = \liminf_n (|x_n - x_0|^2 + |x_0 - x|^2) = |x_0 - x|^2 + \liminf_n |x_n - x_0|^2$$

and the claim follows. $\square$

Krasnoselskii-Mann's convergence theorem

Step 3: since T_θ is nonexpansive, for any $\bar{x}$ and each k,

$$|x^k - \bar{x}| \geq |T_\theta x^k - T_\theta \bar{x}|$$
$$= |x^{k+1} - x^k + x^k - T_\theta \bar{x}| \geq |x^k - T_\theta \bar{x}| - |x^{k+1} - x^k|.$$

Let x^{k_l} be a weakly converging subsequence of x^k (which is bounded by Step 1), to some $\bar{x}$. Since $x^{k+1} - x^k \to 0$ (Step 2), we get:

$$\liminf_l |x^{k_l} - \bar{x}| \geq \liminf_l |x^{k_l} - T_\theta \bar{x}|.$$

Opial's lemma implies that $T_\theta \bar{x} = \bar{x}$, that is, $\bar{x} \in F$.

Step 4: To conclude, assume that a subsequence $(x^{m_l})_l$ of $(x^k)_k$ converges weakly to another fixed point $\bar{y} \neq \bar{x}$.
Then by Opial's lemma we deduce both that $m(\bar{x}) < m(\bar{y})$ and $m(\bar{y}) < m(\bar{x})$:

$$m(\bar{y}) = \liminf_l |x^{m_l} - \bar{y}| < \liminf_l |x^{m_l} - \bar{x}| = m(\bar{x}),$$

$$m(\bar{x}) = \liminf_l |x^{k_l} - \bar{x}| < \liminf_l |x^{k_l} - \bar{y}| = m(\bar{y}),$$

a contradiction. It follows that the whole sequence (x^k) must weakly converge to $\bar{x}$. $\square$

Krasnoselskii-Mann's convergence theorem

Variants

Varying steps: One can consider more generally iterations of the form

$$x^{k+1} = x^k + \tau_k (T_1 x^k - x^k)$$

with varying steps τ_k. Then, if $0 < \underline{\tau} \leq \tau_k \leq \overline{\tau} < 1$, the convergence still holds, with almost the same proof.

Remark: A sufficient condition is that $\sum_k \tau_k (1 - \tau_k) = \infty$, see (Reich, 1979).

Assume now the sequence (x_k) is an "inexact" iteration of T_θ:

$$|x^{k+1} - T_\theta x^k| \leq \varepsilon_k.$$

Then one has the following result:

Theorem

If $\sum_k \varepsilon_k < \infty$, then $x^k \rightharpoonup \bar{x}$ a fixed point of T (if one exists).

(The condition is quite strong. [but clearly necessary!])

Krasnoselskii-Mann's convergence theorem

Errors

Proof: now, x^k is "quasi-Fejér monotone": denoting $e_k = x^{k+1} - T_\theta x^k$ so that $|e_k| \le \varepsilon_k$,

$$|x^{k+1} - x^*| = |T_\theta x^k - T_\theta x^* + e_k| \le |x^k - x^*| + \varepsilon_k$$

for all k, and any $x^* \in F$. Hence, $|x^{k+1} - x^*| \le |x^0 - x^*| + \sum_{i=0}^k \varepsilon_i$ is bounded. Letting $a_k = \sum_{i=k}^\infty \varepsilon_i$ which is finite and goes to 0 as $k \to \infty$, this can be rewritten

$$|x^{k+1} - x^*| + a_{k+1} \le |x^k - x^*| + a_k$$

so that once more one can define

$$m(x^*) := \lim_{k\to\infty} |x^k - x^*| = \inf_{k\ge 0} |x^k - x^*| + a_k$$

Again, if $m(x^*) = 0$ the theorem is proved. If not, the rest of the proof follows steps 2, 3, 4 with very little changes.

It follows from this Theorem the convergence for the explicit and implicit gradient descent for convex functions.
Consider $T_\tau(x) = x - \tau\nabla f(x)$ for f convex with L-Lipschitz gradient. We have that

$$T_{2/L}(x) = x - \frac{2}{L}\nabla f(x)$$

is nonexpansive. Hence if $0 < \tau < 2/L$, one has

$$T_\tau(x) = x - \frac{\tau L}{2}\frac{2}{L}\nabla f(x) = \frac{\tau L}{2}T_{2/L}(x) + \left(1 - \frac{\tau L}{2}\right)x$$

is an averaged operator with $\theta = L\tau/2 \in]0,1[$.
Hence if $x^k = T_\tau^k x_0$, and T_τ has a fixed point, x^k weakly converges to a fixed point. Now, $T_\tau x = x$ if and only if $\nabla f(x) = 0$. We have $|\nabla f(x^k)| \lesssim 1/\sqrt{k}$.
(For the implicit descent, if f is convex, we will see that the associated operator is always $1/2$-averaged.)

Important remark: Let T_θ, S_λ be averaged operators: $T_\theta = (1-\theta)I + \theta T_1$, $S_\lambda = (1-\lambda)I + \lambda S_1$. Then $T_\theta \circ S_\lambda$ is also averaged: letting $\mu = \theta + \lambda(1-\theta) \in]0,1[$, one has

$$T_\theta \circ S_\lambda = (1-\mu)I + \mu \frac{(1-\theta)\lambda S_1 + \theta T_1 \circ ((1-\lambda)I + \lambda S_1)}{\theta + (1-\theta)\lambda}.$$

Application: "Forward-Backward splitting". Show that if, for g convex and f convex with L-Lipschitz gradient

$$\min_x f(x) + g(x)$$

has a solution x^* then

$$x^{k+1} = \arg\min_x \frac{|x - (x - \tau\nabla f(x))|^2}{2\tau} + g(x)$$

weakly converges to a solution if $0 < \tau < 2/L$.

Convex analysis (in Hilbert spaces)

A *convex* set C in a linear space is such that $x, y \in C \Rightarrow [x, y] \in C$ where $[x, y] = \{tx + (1-t)y : t \in [0, 1]\}$.
A *convex function* is a function $f : C \to \mathbb{R}$ such that for any $x, y \in C$, $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Convex functions

A *"generalized" convex function* is a function $f : \mathcal{X} \to \mathbb{R} \cup \{-\infty, +\infty\}$ such that its *epigraph*

$$\operatorname{epi} f := \{(x, \lambda) \in \mathcal{X} \times \mathbb{R} : \lambda \geq f(x)\}$$

is a convex set.

We define the *domain* of f as $\operatorname{dom} f = \{x : f(x) < +\infty\}$. It is (clearly) a convex set.

Then, f is *proper* if it is not identically $+\infty$ and never $-\infty$. In this case, again, one can say it is convex if and only if for any $x, y \in \mathcal{X}$, $t \in [0, 1]$,

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y).$$

where now some values here may be $+\infty$.

Convex functions

Remark: If there exists $\check{x}$ with $f(\check{x}) = -\infty$, then for any $x \in \operatorname{dom} f$, as soon as there exists $z \in \operatorname{dom} f$ with $x \in [\check{x}, z]$, one has for some $t \in (0,1)$:

$$f(x) \leq tf(\check{x}) + (1-t)f(z) = -\infty.$$

In particular, f is identically $-\infty$ in the relative interior $\operatorname{ri} \operatorname{dom} f$ of its domain, where for a convex C, the *relative interior of C* is defined as:

$$\operatorname{ri} C := \{x \in C : \forall y \in C \setminus \{x\},\ \exists z \in C \setminus \{x\}, x \in [y,z]\}\,.$$

(In finite dimension, this is nothing but the interior of C in the subspace spanned by C.)
Hence, convex functions which take the value $-\infty$ are not particularly interesting...

f is *strictly convex* if the inequality is strict whenever $x \neq y$ and $0 < t < 1$. If $\mathcal{X}$ is a normed space, f is *strongly convex* (or μ-convex) if in addition, there exists $\mu > 0$ such that for all $x, y \in \mathcal{X}$ and $t \in [0,1]$,

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

[This can be used also with non-Hilbertian norms!]

Thanks to the parallelogram identity, *in the Hilbertian setting*, one easily checks that this is equivalent to require that $x \mapsto f(x) - \mu/2|x|^2$ is still convex. The archetypical example of a μ-convex function is a quadratic plus affine function $\mu|x|^2/2 + \langle b, x\rangle + c$.

Convex functions

More definitions and examples

The function f is *lower semi-continuous* (l.s.c.) if for all $x \in \mathcal{X}$, if $x_n \to x$, then

$$f(x) \leq \liminf_{n\to\infty} f(x_n).$$

It is easy to see that f is l.s.c. if and only if $\text{epi}\, f$ is closed.

Example: An important example of a convex function the *characteristic function* or *indicator function* of a set C (often denoted ι_C, χ_C, or δ_C):

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ +\infty & \text{else,} \end{cases}$$

which is convex and proper as soon as C is convex and non-empty, and is l.s.c. as soon as C is closed.

Regularity of convex functions

Lemma

Let f be proper, convex.
If there exists $B \subset \operatorname{dom} f$ a ball with $\sup_B f < +\infty$, then f is locally Lipschitz in the interior of $\operatorname{dom} f$.
In finite dimension, a proper convex function f is locally Lipschitz in $\operatorname{ri} \operatorname{dom} f$.

Proof: Wlog assume $B = B(0,\delta)$, $\delta > 0$, let $M = \sup_B f < \infty$.

For $x \in B$, by convexity $f(x) \geq 2f(0) - f(-x) \geq 2f(0) - M$, hence $|f| \leq M + 2|f(0)|$.

We prove that f is Lipschitz in $B(0,\delta/2)$: given $x, y \in B(0,\delta/2)$, there is $z \in B(0,\delta)$ such that $y = (1-t)x + tz$ for some $t \in [0,1]$, and $|z - x| \geq \delta/2$.

By convexity, $f(y) - f(x) \leq t(f(z) - f(x)) \leq t2(M - f(0))$. We have $t(z - x) = y - x$ so that $t \leq |y - x|/|z - x| \leq 2|y - x|/\delta$, so that

$$f(y) - f(x) \leq \left(\frac{4(M - f(0))}{\delta}\right)|y - x|.$$

Regularity of convex functions

Now let $\bar{x}$ in the interior of $\operatorname{dom} f$. Observe that for some $\lambda > 1$, $\lambda\bar{x} \in \operatorname{dom} f$ hence $B' = 1/\lambda(\lambda\bar{x}) + (1 - 1/\lambda)B(0, \delta) = B(\bar{x}, \delta(1 - 1/\lambda)) \subset \operatorname{dom} f$; moreover, if $x \in B'$,$x = 1/\lambda(\lambda\bar{x}) + (1 - 1/\lambda)z$ for some z with $f(z) \leq M$ hence $f(x) \leq 1/\lambda f(\lambda\bar{x}) + (1 - 1/\lambda)M$, so that $\sup_{B'} f < \infty$. Hence as before f is Lipschitz in a smaller ball. This shows that f is locally Lipschitz in the interior of $\operatorname{dom} f$.

In finite dimension, assume $0 \in \operatorname{dom} f$ and let d be the dimension of $\operatorname{vect}\operatorname{dom} f$. It means there exist $x_1, \dots, x_d$ independent points in $\operatorname{dom} f$. Now, the d-dimensional set $\{\sum_i t_i x_i : t_i > 0, \sum_i t_i \leq 1\}$ (the interior of the convex envelope of $\{0, x_1, \dots, x_d\}$) is an open set in $\operatorname{vect}\operatorname{dom} f$, moreover if $x = \sum_i t_i x_i$, $f(x) \leq \sum_i t_i f(x_i) + (1 - \sum_i t_i)f(0) \leq M := \max\{f(0), f(x_1), \dots, f(x_d)\}$. Hence we can apply the first part of the theorem, and f is locally Lipschitz in the relative interior of the domain.

Regularity, remark

In infinite dimension one can possibly find noncontinuous linear forms hence noncontinuous convex functions. (the typical example is a linear function defined by $f(e_n) = n$ where $(e_n)_{n\geq 1}$ is an independent family, which is then completed into a basis $\mathcal{B}$, then, one lets $f(e) = 0$ if $e \in \mathcal{B} \setminus \{e_n : n \geq 1\}$.) [Hence: such a function cannot be bounded on an open set!]

A convex proper *lower semi-continuous* function is always locally bounded in the interior of its domain, and therefore locally Lipschitz.
Indeed if 0 is an interior point and one considers the convex closed set $C = \{x : f(x) \leq 1 + f(0)\}$, one can check that $\cup_{n\geq 1} nC = \mathcal{X}$, as if $x \in \mathcal{X}$, $t \mapsto f(tx)$ is locally Lipschitz near $t = 0$. Hence $\mathring{C} \neq \emptyset$ by Baire's property: it follows that there is an open ball where f is bounded, as requested (see Ekeland-Temam, Cor. 2.5).
But in infinite dimension there are also many interesting lsc. convex functions whose domain has empty interior. [ex: $u \mapsto \int |\nabla u|^2 dx$ for $u \in L^2$.]

Separation of convex sets

Convex sets are easily "separated" by hyperplanes. This is called a geometric version of Hahn-Banach's theorem. In the Hilbert settings, the proofs are quite easy, and constructive (why in more general settings, the proofs usually rely on Zorn's lemma).

Separation of convex sets

Convex sets are easily "separated" by hyperplanes. This is called a geometric version of Hahn-Banach's theorem. In the Hilbert settings, the proofs are quite easy, and constructive (why in more general settings, the proofs usually rely on Zorn's lemma).

Separation Theorem (1st version)

Let $\mathcal{X}$ be a (real) Hilbert space, $C \subset \mathcal{X}$ a ***closed****, convex set and $x \notin C$. Then there exists a closed hyperplane which "separates" strictly x and C:*
$\exists v \in \mathcal{X}, \alpha \in \mathbb{R}$ s.t.:

$$\langle v, x\rangle > \alpha \geq \langle v, y\rangle \ \forall\, y \in C$$

Separation of convex sets

Proof: We introduce the projection $z = \Pi_C(x)$ defined by $|x - z| = \min_{y \in C} |x - y|$.

The first order optimality condition for z is found by writing that for any $y \in C$,

$$|x - z|^2 \leq |x - (z + t(y - z))|^2 \quad \forall t \in (0, 1].$$

Sending $t \to 0$, we find:

$$\langle x - z, y - z \rangle \leq 0 \;\forall\; y \in C.$$

Hence if $v = x - z \neq 0$, $y \in C$,

$$\langle v, x \rangle = \langle x - z, x \rangle = |x - z|^2 + \langle x - z, z \rangle \geq |x - z|^2 + \langle x - z, y \rangle = |v|^2 + \langle v, y \rangle .$$

The result follows (letting for instance $\alpha = |v|^2/2 + \sup_{y \in C} \langle v, y \rangle$).

The proof can easily be extended to the situation where $\{x\}$ is replaced with a compact convex set K not intersecting C: consider $C' = C - K$ and $0 \notin C'$

Separation of convex sets

Corollary

In a real Hilbert space $\mathcal{X}$, a closed convex set C is weakly closed.

Indeed, if $x \notin C$, one finds v, α with $\langle v, x\rangle > \alpha \geq \langle v, y\rangle$ for all $y \in C$ and this defines a neighborhood $\{\langle v, \cdot\rangle > \alpha\}$ of x for the weak topology which does not intersect C: the complement of C is therefore open.

Separation of convex sets

Separation Theorem (2nd version)

Let $\mathcal{X}$ be a real Hilbert space, $C \subset \mathcal{X}$ an ***open*** *convex set and $C' \subset \mathcal{X}$ a convex set with $C' \cap C = \emptyset$. Then there exists a closed hyperplane which "separates" C and C':*
$\exists v \in X, \alpha \in \mathbb{R}$, $v \neq 0$, *s.t.:*

$$\langle v, x \rangle \geq \alpha \geq \langle v, y \rangle \ \forall\, x \in C, y \in C'$$

Separation of convex sets

Proof: first case: $C' = \{\bar{x}\}$ is a singleton.
The difficult case is whenever $\bar{x} \in \overline{C} \setminus C$, otherwise we can apply the previous theorem to separate (strictly) $\bar{x}$ and $\overline{C}$.

By assumption, there exists a ball $B = B(y, \delta) \subset C$.
Let $x_n = y + (1 + \frac{1}{n})(\bar{x} - y)$, which is such that $x_n \to \bar{x}$ as $n \to \infty$. Since

$$\bar{x} = \frac{n}{n+1} x_n + \frac{1}{n+1} y,$$

one has $x_n \notin \overline{C}$. Indeed, if $x_n \in \overline{C}$, there is $x' \in C$ with $|x' - x_n| < \frac{\delta}{2n}$, and by convexity, $C \supset \frac{n}{n+1} x' + \frac{1}{n+1} B(y, \delta) = B(\frac{n}{n+1}(x' - x_n) + \bar{x}, \frac{\delta}{n+1}) \ni \bar{x}$, a contradiction.
By the previous separation Theorem, there exists v_n such that for all $x \in \overline{C}$,

$$\langle v_n, x_n \rangle \leq \langle v_n, x \rangle$$

and we can assume $|v_n| = 1$. Up to a subsequence, we may even assume that $v_n \rightharpoonup v$ weakly in $\mathcal{X}$.
In the limit, (using that $x_n \to \bar{x}$ strongly) we obtain $\langle v, \bar{x} \rangle \leq \langle v, x \rangle \; \forall \; x \in C$, which is our claim if $v \neq 0$ (for instance, in finite dimension).

Separation of convex sets

Using again the ball $B(y,\delta) \subset C$, one has for any $|z| \le 1$

$$\langle v_n, x_n \rangle \le \langle v_n, y - \delta z \rangle$$

so that $\langle v_n, y - x_n \rangle \ge \delta \langle v_n, z \rangle$.
We consider the sup over all possible z: we find $\langle v_n, y - x_n \rangle \ge \delta$. In the limit we deduce $\langle v, y - \bar{x} \rangle \ge \delta$ which shows that $v \ne 0$.

Using again the ball $B(y,\delta) \subset C$, one has for any $|z| \leq 1$

$$\langle v_n, x_n \rangle \leq \langle v_n, y - \delta z \rangle$$

so that $\langle v_n, y - x_n \rangle \geq \delta \langle v_n, z \rangle$.
We consider the sup over all possible z: we find $\langle v_n, y - x_n \rangle \geq \delta$. In the limit we deduce $\langle v, y - \bar{x} \rangle \geq \delta$ which shows that $v \neq 0$.

Now, to show the general case, one lets $A = C' - C = \{y - x : y \in C', x \in C\}$: this is an open convex set and by assumption, $0 \notin A$. Hence by the previous part, there exists $v \neq 0$ such that $\langle v, y - x \rangle \leq \langle v, 0 \rangle = 0$ for all $y \in C'$, $x \in C$, which is what we wanted to show. □

General separation Theorem

For instance: Thm 1.11 of [Convex Optimization in Normed Spaces, Peypouquet, 2015, pp. 7-8]. Here, X^* denotes the set of continuous linear functionals on X (the *topological* dual of X).

Theorem

Let C be a nonempty convex subset of a normed space $(X, \|\cdot\|)$ not containing the origin.

i) if C is open, there exists $L \in X^* \setminus \{0\}$ such that $Lx < 0$ for each $x \in C$;

ii) if C is closed, there exist $L \in X^* \setminus \{0\}$ and $\varepsilon > 0$ such that $Lx + \varepsilon \leq 0$ for each $x \in C$.

[Proof relatively simple, based on Zorn's Lemma: every non-empty partially ordered set in which any chain has an upper bound has a maximal element.]

Convex duality: Legendre-Fenchel conjugate

(Or "convex conjugate")

Consider a function $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$.

Definition: Legendre-Fenchel conjugate

We define, for any $y \in \mathcal{X}$,

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

We also define the *bi-conjugate* of f^* as $f^{**} = (f^*)^*$.

Remark: in a general vector space, f^* is defined for $y \in \mathcal{X}^*$ (a dual space).
Remark: f^* is defined as a sup of continuous, linear forms. It is therefore lower-semicontinuous (as a sup of continuous functions — including for the weak topology) and convex (as a sup of convex functions).

Convex duality: Legendre-Fenchel conjugate

Proposition

Let $f : \mathcal{X} \to [-\infty, +\infty]$. Then $f^{**} \leq f$.

This is because, by definition, for any x, y,

$$f^*(y) \geq \langle y, x \rangle - f(x) \Leftrightarrow f(x) + f^*(y) \geq \langle y, x \rangle \Leftrightarrow f(x) \geq \langle y, x \rangle - f^*(y).$$

Taking the supremum with respect to y, we find that $f \geq f^{**}$. □

Remark: Given $y \in \mathcal{X}$, if for any $c \in \mathbb{R}$, there is $x \in \mathcal{X}$ with $f(x) \leq \langle y, x \rangle - c$, then $f^*(y) = +\infty$. On the other hand, if there exists c such that $f(x) \geq \langle y, x \rangle - c$ for all x, then $f^*(y) \leq c < +\infty$.
So $f^* \not\equiv +\infty$ if and only if f is larger than at least an affine function. If not, $f^* \equiv +\infty$ and $f^{**} \equiv -\infty$.

If f is larger than at least one affine function, and $f \not\equiv +\infty$, then $f^* \not\equiv +\infty$ and $f^*(y) > -\infty$ for all y: f^* is proper. In this case, also f^{**} is proper and one has:

Theorem - convex bi-conjugate

Let $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$, and assume there is an affine function below f. Then f^{**} is the largest convex, lsc. function below f.

Convex duality: Legendre-Fenchel conjugate

The result will be a consequence of the following:

Theorem

Let $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be convex, proper, lower-semicontinuous. Then $f^{**} = f$.

(Actually, both results are corollaries of the other.)

Convex duality: Legendre-Fenchel conjugate

The result will be a consequence of the following:

Theorem

Let $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be convex, proper, lower-semicontinuous. Then $f^{**} = f$.

(Actually, both results are corollaries of the other.)

Proof: If not, there is $(x, t) \in \operatorname{epi} f^{**}$ with $f^{**}(x) \leq t < f(x)$, that is $(x, t) \notin \operatorname{epi} f$. Since $\operatorname{epi} f$ is convex, closed, thanks to the first separation theorem, $\exists (v, \lambda) \in \mathcal{X} \times \mathbb{R}, \alpha \in \mathbb{R}$, s.t.:

$$\langle v, x \rangle - \lambda t > \alpha \geq \left\langle v, x' \right\rangle - \lambda t' \quad \forall\, (x', t') \in \operatorname{epi} f.$$

Since one can send $t' \to +\infty$, we see that λ must be non-negative. If $\lambda \neq 0$, one has, letting $y = v/\lambda$,

$$\langle y, x \rangle - t > \frac{\alpha}{\lambda} \geq \left\langle y, x' \right\rangle - f(x')$$

for any $x' \in \operatorname{dom} f$, so that $\alpha/\lambda \geq f^*(y)$. We deduce $f^{**}(x) \leq t < \langle y, x \rangle - f^*(y)$, which is a contradiction. So the difficult case is $\lambda = 0$.

Convex duality: Legendre-Fenchel conjugate

If $\lambda = 0$, one has $\langle v, x\rangle > \alpha \geq \langle v, x'\rangle \quad \forall\, x' \in \operatorname{dom} f$.

We first remark that f being convex, lsc, proper, is above an affine function: indeed given a point $\bar{x} \in \operatorname{dom} f$, the separation theorem applied to $(\bar{x}, f(\bar{x}) - 1) \notin \operatorname{epi} f$ shows that there exists (w, μ) such that

$$\langle w, \bar{x}\rangle - \mu(f(\bar{x}) - 1) > \left\langle w, x'\right\rangle - \mu f(x')$$

for all $x' \in \operatorname{dom} f$. Taking $x' = \bar{x}$ we see that $\mu > 0$, hence letting $p = w/\mu$, we obtain $f(x') > \langle p, x'\rangle - c$ for $c = f(\bar{x}) - 1 - \langle p, \bar{x}\rangle$.

Convex duality: Legendre-Fenchel conjugate

If $\lambda = 0$, one has $\langle v, x\rangle > \alpha \geq \langle v, x'\rangle \quad \forall\, x' \in \operatorname{dom} f$.

We first remark that f being convex, lsc, proper, is above an affine function: indeed given a point $\bar{x} \in \operatorname{dom} f$, the separation theorem applied to $(\bar{x}, f(\bar{x}) - 1) \notin \operatorname{epi} f$ shows that there exists (w, μ) such that

$$\langle w, \bar{x}\rangle - \mu(f(\bar{x}) - 1) > \left\langle w, x'\right\rangle - \mu f(x')$$

for all $x' \in \operatorname{dom} f$. Taking $x' = \bar{x}$ we see that $\mu > 0$, hence letting $p = w/\mu$, we obtain $f(x') > \langle p, x'\rangle - c$ for $c = f(\bar{x}) - 1 - \langle p, \bar{x}\rangle$.

Hence for $t > 0$, one has:

$$f^*(tv + p) = \sup_{x'} t\left\langle v, x'\right\rangle + \left\langle p, x'\right\rangle - f(x') \leq t \sup_{x' \in \operatorname{dom} f} \left\langle v, x'\right\rangle + c \leq \alpha t + c,$$

so that

$$f^{**}(x) = \sup_q \langle q, x\rangle - f^*(q) \geq \sup_{t>0} \langle tv + p, x\rangle - f^*(tv + p) \geq \sup_{t>0} t(\langle v, x\rangle - \alpha) + \langle p, x\rangle - c = +\infty$$

which is again a contradiction since we assumed $f^{**}(x) < f(x)$. $\square$

Convex duality: Legendre-Fenchel conjugate

Now we prove the first theorem, that is the case where f is not necessarily convex, lower-semicontinuous. In this case, we use that as before $f^{**} \leq f$. If $f \equiv +\infty$ then $f^* \equiv -\infty$ and $f^{**} \equiv +\infty$: the result is trivial. Otherwise $f^{**} \leq f$ is proper.

Consider $g \leq f$ convex, lower-semicontinuous. We want to show that $g \leq f^{**}$. We may assume g is proper, otherwise $g \leq g' = \max\{g, a\} \leq f$, where a is some affine function below f, g' is proper, and $g' \leq f^{**} \Rightarrow g \leq f^{**}$.
One has that $g^* \geq f^*$ and $g^{**} \leq f^{**}$. By the last theorem, $g^{**} = g$, so that $g \leq f^{**}$.
Hence f^{**} is the largest convex, semi-continuous function below f. □

Convex duality: Legendre-Fenchel conjugate

Remark: the following "local" version of the last result can also be useful.

Corollary

Let f be convex, proper and assume it is lower-semicontinuous at x. Then $f^{**}(x) = f(x)$.

Proof: $f(x) \leq \liminf_{y\to x} f(y)$, equivalently for any $t < f(x)$, there is a ball $B = B(x,\delta)$ with $f > t$ in B. Hence $\operatorname{epi} f \cap (B(x,\delta)\times(-\infty,t)) = \emptyset$, so that also $\overline{\operatorname{epi} f} \cap (B(x,\delta)\times(-\infty,t)) = \emptyset$. Being f convex, $\overline{\operatorname{epi} f} = \operatorname{epi} f^{**}$ hence $f^{**}(x) \geq t$. Letting $t \to f(x)$ we deduce $f^{**}(x) \geq f(x)$ □

Conjugate: Examples

1. $f(x) = |x|^2/(2\alpha)$, $\alpha > 0$: $f^*(y) = \alpha|y|^2/2$;
2. $f(x) = |x|^p/p$: $f^*(y) = |y|^{p'}/p'$, $1/p + 1/p' = 1$;
3. $F(f) = \|f\|_{L^p}^p/p$: $F^*(g) = \|g\|_{L^{p'}}^{p'}/p'$ (the duality is in L^2, however this is also true in the $(L^p, L^{p'})$ duality, *cf* Ekeland-Temam's book);
4. $f(x) = \delta_{B(0,1)}(x) = 0$ if $x \in B(0,1)$, $+\infty$ else: $f^*(p) = |p|$.

Conjugate: 1-homogeneous functions

If f is positively 1-homogeneous, then

$$f^*(y) = \sup_x \langle y, x\rangle - f(x) = \sup_{t>0} \sup_x \langle y, tx\rangle - f(tx) = \sup_{t>0} t f^*(y) \in \{0, +\infty\}$$

and precisely

$$f^*(y) = \begin{cases} 0 & \text{if } \langle y, x\rangle \le f(x)\ \forall x \in \mathcal{X}, \\ +\infty & \text{if } \exists x \in \mathcal{X},\ \langle y, x\rangle > f(x). \end{cases}$$

Conjugate: 1-homogeneous functions

If f is positively 1-homogeneous, then

$$f^*(y) = \sup_x \langle y, x \rangle - f(x) = \sup_{t>0} \sup_x \langle y, tx \rangle - f(tx) = \sup_{t>0} t f^*(y) \in \{0, +\infty\}$$

and precisely

$$f^*(y) = \begin{cases} 0 & \text{if } \langle y, x \rangle \leq f(x) \; \forall x \in \mathcal{X}, \\ +\infty & \text{if } \exists x \in \mathcal{X}, \; \langle y, x \rangle > f(x). \end{cases}$$

Letting $C = \{y : \langle y, x \rangle \leq f(x) \; \forall x \in \mathcal{X}\}$ one has $f^* = \delta_C$ (C is clearly closed and convex, and f^* convex lsc). Eventually, observe that if f is convex and lsc, then $f^{**} = f$ which shows that in this case:

$$f(x) = \sup_{y \in C} \langle y, x \rangle .$$

Conjugate: "∞-homogeneous" functions / β-homogeneous functions.

Conversely if $f = \delta_C$ for some set C (f is the *support function* of C), one easily sees that $f^*(y) = \sup_{y \in C} \langle y, x \rangle$ is convex, 1-homogeneous, and that $f^{**} = \delta_{\overline{\text{co}\, C}}$ the characteristic of the closed convex envelope of C, that is the smallest closed convex set containing C.

Conjugate: "∞-homogeneous" functions / β-homogeneous functions.

Conversely if $f = \delta_C$ for some set C (f is the *support function* of C), one easily sees that $f^*(y) = \sup_{y \in C} \langle y, x \rangle$ is convex, 1-homogeneous, and that $f^{**} = \delta_{\overline{\text{co}\, C}}$ the characteristic of the closed convex envelope of C, that is the smallest closed convex set containing C.

In general for f positively β-homogeneous, $1 < \beta < \infty$, then

$$f^*(ty) = \sup_x \langle ty, x \rangle - f(x) = t^\alpha \sup_x \left\langle y, t^{1-\alpha} x \right\rangle - f(t^{-\alpha/\beta} x) = t^\alpha f^*(y)$$

if $1 - \alpha = -\alpha/\beta$, hence if $1/\alpha + 1/\beta = 1$.

Conjugate of a sum

Inf-convolutions

We consider f, g convex lsc. functions. We define their *inf-convolution* as

$$f \Box g(x) = \inf_y f(x-y) + g(y).$$

This defines a convex function (more generally, given $G(x,y)$ convex in (x,y), $x \mapsto \inf_y G(x,y)$ is also convex).

Lemma

If there is $p \in \mathcal{X}$ where f^* is continuous and g^* is finite, then the inf is a "min" and $f \Box g$ is convex, lsc. In finite dimension, it is enough to have $p \in \operatorname{ri} \operatorname{dom} f^* \cap \operatorname{ri} \operatorname{dom} g^*$.

Inf-convolutions: proof

Proof: consider indeed $x_n \to x$ and y_n such that

$$f\Box g(x_n) \geq f(x_n - y_n) + g(y_n) - \frac{1}{n}.$$

Consider a subsequence with

$$\lim_k f(x_{n_k} - y_{n_k}) + g(y_{n_k}) = \liminf_n f(x_n - y_n) + g(y_n) \leq \liminf_n f\Box g(x_n)$$

Observe that if f^* is continuous at p, then it means that there is a constant c and $\varepsilon > 0$ such that

$$f^*(q) \leq c + \delta_{B(0,\varepsilon)}(q - p)$$

(where δ_C is the characteristic function of C which is zero in C and $+\infty$ elsewhere) while $g^*(p) < +\infty$: so that for all z

$$f(z) = f^{**}(z) \geq \langle p, z\rangle - c + \varepsilon|z|, \quad g(z) \geq \langle p, z\rangle - g^*(p).$$

Hence,

$$f(x_{n_k} - y_{n_k}) + g(y_{n_k}) \geq \langle p, x_{n_k} - y_{n_k}\rangle - c + \varepsilon|x_{n_k} - y_{n_k}| + \langle p, y_{n_k}\rangle - g^*(p)$$
$$= \langle p, x_{n_k}\rangle + \varepsilon|x_{n_k} - y_{n_k}| - (c + g^*(p))$$

so that $(x_{n_k} - y_{n_k})_k$ is a bounded sequence, hence there exists y and a subsequence of (y_{n_k}) (not relabelled) with $y_{n_k} \rightharpoonup y$. In the limit (as, f, g are weakly lsc),

$$f\Box g(x) \leq f(x - y) + g(y) \leq \liminf_k f(x_{n_k} - y_{n_k}) + g(y_{n_k}) \leq \liminf_n f\Box g(x_n).$$

If the sequence $x_n \equiv x$, we find that there is a minimizer y in the definition of the inf-convolution.

Conjugate of a sum

Corollary

Let f, g be convex, lsc: if there exists $x \in \operatorname{dom} f \cap \operatorname{dom} g$ such that f is continuous at x (in finite dimension, $x \in \operatorname{ri}\operatorname{dom} f \cap \operatorname{ri}\operatorname{dom} g$), then $(f+g)^* = f^* \Box g^*$,

Proof: By our assumption and the previous result, $f^* \Box g^*$ is lsc, and:

$$(f^* \Box g^*)^*(x) = \sup_{p,q} \langle x, p\rangle - f^*(q) - g^*(p-q)$$
$$= \sup_{p,q} \langle x, q\rangle - f^*(q) + \langle x, p-q\rangle - g^*(p-q) = f(x) + g(x).$$

Hence $(f+g)^* = (f^* \Box g^*)^{**} = f^* \Box g^*$.

Fenchel-Rockafellar duality

Now we consider the problem

$$\min_{x\in\mathcal{X}} f(Kx) + g(x)$$

with $K : \mathcal{X} \to \mathcal{Y}$ is continuous linear map and f, g convex, lsc. Then:

$$\begin{aligned}(\mathcal{P}) = \min_x f(Kx) + g(x) &= \min_x \sup_y \langle y, Kx\rangle - f^*(y) + g(x) \\ &\geq \sup_y \inf_x \langle K^*y, x\rangle + g(x) - f^*(y) = \sup_y -(g^*(-K^*y) + f^*(y)) = (\mathcal{D})\end{aligned}$$

If there is equality one says that there is "Strong duality". This is "often" true. The problem "$(\mathcal{P})$" is usually called the *primal problem* and "$(\mathcal{D})$" the *dual problem*.

Fenchel-Rockafellar duality

The *primal-dual gap*

$$\mathcal{G}(x,y) = f(Kx) + g(x) + g^*(-K^*y) + f^*(y)$$

is a measure of optimality. It vanishes at (x^*,y^*), if and only if $(\mathcal{P}) = (\mathcal{D})$, x^* is optimal in $(\mathcal{P})$ and y^* in $(\mathcal{D})$, if and only if (x^*,y^*) is a *saddle point* of the *Lagrangian*

$$\mathcal{L}(x,y) = \langle y, Kx\rangle - f^*(y) + g(x).$$

Indeed $\mathcal{G}(x^*,y^*) = 0 \Rightarrow \forall\, y, x$,
$\mathcal{L}(x^*,y) \le f(Kx^*) + g(x^*) = -f^*(y^*) - g^*(-K^*y^*) \le \mathcal{L}(x,y^*)$ and one deduces

$$(\forall x \in \mathcal{X}, y \in \mathcal{Y}) \quad \mathcal{L}(x^*,y) \le \mathcal{L}(x^*,y^*) \le \mathcal{L}(x,y^*). \qquad (\mathcal{S})$$

[The converse is not harder.]

Fenchel-Rockafellar duality

The following is an example of Strong duality theorem

Theorem

If there exists $\bar{x} \in \operatorname{dom} g$ with f continuous at $K\bar{x}$, then $(\mathcal{P}) = (\mathcal{D})$. (Moreover under these assumptions, $(\mathcal{D})$ has a solution.)

In finite dimension, it is shown in Rockafellar, (Cor 31.2.1) that Strong duality if there exists $x \in \operatorname{ri} \operatorname{dom} g$ with $Kx \in \operatorname{ri} \operatorname{dom} f$, or even more generally that $0 \in \operatorname{ri} (\operatorname{dom} f - K \operatorname{dom} g)$ (the proof is almost the same as below).

Fenchel-Rockafellar duality

Proof: We use the so-called "perturbation method": We introduce, for $z \in \mathcal{Y}$,

$$\Phi(z) := \inf_{x\in\mathcal{X}} f(Kx+z) + g(x).$$

Assume $\Phi(0) > -\infty$ (otherwise there is nothing to prove), then by assumption, one can find M and ε such that for $|z| < \varepsilon$, $\Phi(z) \le f(K\bar{x}+z) + g(\bar{x}) \le M < +\infty$. Being Φ convex, we deduce that it is locally Lipschitz near 0 and in particular thanks to a previous theorem, $\Phi(0) = \Phi^{**}(0) = \sup_y -\Phi^*(y)$. We compute:

$$\Phi^*(y) = \sup_{z\in\mathcal{Y}} \langle y, z\rangle - \inf_{x\in\mathcal{X}} (f(Kx+z) + g(x))$$
$$= \sup_{x,z} \langle y, z+Kx\rangle - \langle K^*y, x\rangle - f(Kx+z) - g(x) = f^*(y) + g^*(-K^*y).$$

The claim follows. The subdifferentiability theory (next lecture) will also show that under these assumptions, there always is a solution to the dual problem. (One can also follow the steps of the proof of minimality of inf-convolutions, for a minimizing sequence of the dual problem.)

Remark: In finite dimension, if $0 \in \operatorname{ri}(\operatorname{dom} f - K\operatorname{dom} g)$, one can show again that Φ is lsc at 0 and proceed in the same way.

We consider the problem

$$\min_x \lambda |Dx|_1 + \frac{1}{2}|x - x^0|^2 \qquad (\mathcal{P})$$

where $D : \mathbb{R}^n \to \mathbb{R}^m$ is a continuous operator, $x^0 \in \mathbb{R}^n$, $|\cdot|_1$ is the ℓ^1-norm. One has

$$f = \lambda|\cdot|_1, \quad K = D, \quad g = \frac{1}{2}|\cdot - x^0|^2.$$

Then the Lagrangian is

$$\mathcal{L}(x, y) = \langle y, Dx \rangle - f^*(y) + g(x)$$

where $f^*(y) = 0$ if $|y_i| \leq \lambda$ for $i = 1, \dots, n$, and $+\infty$ else. To find the dual problem, we compute $g^*(z) = \langle z, x^0 \rangle + |z|^2/2$, and we obtain

$$\max \left\{ \langle D^* y, x^0 \rangle - \frac{1}{2}|D^* y|^2 : |y_i| \leq \lambda, i = 1, \dots, n \right\}. \qquad (\mathcal{D})$$

This can be rewritten as a projection problem:

$$\min_{|y_i| \leq \lambda} |D^* y - x^0|^2.$$

and can be solved for instance by *implicit* gradient descent with the metric $I/\tau - DD^*$, for $\tau < 1/\|D\|^2$.