Convex Analysis

Lecturer: Michael Renger
Weierstraß Institut
renger@wias-berlin.de

Course Material: several books (will be more specific for each lecture)
mostly:
Juan Peypouquet - Convex Optimization in
Normed Spaces: Theory, Methods & examples
(Downloadable from his website)

Language: English or Deutsch

Lecture 1

A Introduction to convexity

Def. X Banach space
F : X → IR∪{∞} convex : ⇔ F(λx + (1−λ)y) ≤ (1−λ)F(x) + λF(y)
∀x, y ∈ X, λ ∈ [0, 1]
strictly convex : ⇔ ... < ... ∀x ≠ y, λ ∈ (0, 1)

→ why Banach? Vector space
→ why IR∪{∞}? Connection with topology (later on)

Def. DomF = {x ∈ X: F(x) < ∞} Domain
F is proper : ⇐ DomF ≠ ∅

Typically, X will be an L^2-space, Sobolev space or "Orlicz" space

Examples:
• F : L^2(Ω) → IR∪{∞}, F(x) = ∫_Ω |∇x|^2 dx, if x ∈ W^{1,2}(Ω),
   otherwise.
• F(x) = ∫_Ω f(x, ∇x) dx convex if f convex in x
• F(x) = ∫_Ω f(x, ∇x) dx convex if f convex in ∇x
• F(x) = ∫_Ω f(x, ∇x) dx convex if f convex in x

No! F only convex if f is convex in (x, ∇x) jointly convex
For example f(x, x, x) = x^2 + 4xy + y^2 is convex in x and in y
but on the diagonal f(x, x, x) = −2x^2.
Motivating and important application:

The direct method in the calculus of variations

\[
\inf_{x \in X} F(x) = \inf_{x \in X} \int_{\Omega} f(q, x(q), \nabla x(q)) dq
\]

If a minimiser exists, then it must solve the Euler-Lagrange eq:

\[
\frac{\partial f}{\partial x}(q, x(q), \nabla x(q)) = \text{div} \left( \frac{\partial f}{\partial \nabla x}(q, x(q), \nabla x(q)) \right).
\]

So this would prove existence of a solution to this (possibly nonlinear) equation!

How to prove that a minimiser exists?

**Strategy**

(i) Assume \( F \) is bounded from below.

Then there exists a minimising sequence \((x_n)_n \subset X\), i.e.

\[
\lim_{n \to \infty} F(x_n) = \inf F,
\]

\[
F(x_{n+1}) \leq F(x_n) \quad \text{for all } n \geq 1.
\]

(ii) Show that the sequence converges (somehow)

(iii) Show that the limit must be a minimiser of \( F \)

(ii) \((x_n)_n \subset \{ x \in X: F(x) \leq F(x_1) =: C \} = \{ F \leq C \} \) (sub) level set of \( F \)

1-d: \( \mathbb{R}^d \) has no minimiser

but if \( \lim_{x \to \infty} f(x) = \infty \) then \( \{ F \leq C \} \) is bounded

oo-d: Assume \( F(x) \geq \varphi(\|x\|) - \alpha \) for some \( \alpha \in \mathbb{R} \)

and non-negative increasing \( \varphi: \mathbb{R}^d \to \mathbb{R} \) coercivity/ growth condition

Then \( \varphi(\|x\|) \leq F(x) + \alpha \leq C + \alpha \) on \( \{ F \leq C \} \)

\( \|x\| \leq \varphi^{-1}(C + \alpha) \Rightarrow \) sublevel set \( \{ F \leq C \} \)

is bounded in \( X \).

Compactness? \( \Rightarrow \) Intermediary weak topologies

Assume \( X \) has a predual, i.e. \( X = Y^* \) and \( \|x\| = \|y^*\| \)

\((x_n)_n \subset \{ F \leq C \} \) bounded \( \Rightarrow \) By Banach-Alaoglu:

\((x_n)_n\) is relatively weak-* compact,

\( \Rightarrow \) there exists a weakly-* convergent subsequence

\( x_{n_m} \overset{*}{\to} x_0 \) candidate for minimiser!
\((iii) \lim_{m \to \infty} F(x_m) = \lim_{n \to \infty} F(x_n) = \inf F\)

If \(F\) would be weakly-* continuous, then
\[\inf F = \lim_{m \to \infty} F(x_m) = F(x_\infty),\]
and hence \(x_\infty\) would indeed be a minimiser!

Generally not true \(\Box\). In fact,

**Prop** ~ \(F\) is weakly or weakly-* continuous
\[\Rightarrow \quad F\text{ is }"\text{strongly}"\ (in the norm-topology) continuous

Claim: Lower semi-continuity is enough:

Assume \(F\) is lower semicontinuous (l.s.c.) i.e.
\[\liminf_{n \to \infty} F(x_n) \geq F(x) \quad \forall x \in X\]

Then:
\[\inf F \geq \liminf_{m \to \infty} F(x_m) \geq F(x) \geq \inf F\]

Hence \(x_\infty\) is a minimiser!

**Th.** Assume hold from below (⇒ minimizing sequence)
- \(F\) is lower semi-continuous (⇒ bounded level sets)
- Coercivity (⇒ bounded level sets)
- \(X\) has a predual (⇒ weakly-* compact level sets)
- \(F\) is weakly-* l.s.c. (⇒ cluster points are minimisers)

Then \(\exists\) there exists a minimiser \(\inf F = F(x)\)

*Remarks:*
- the existence of a minimiser has nothing to do with topology
  We chose a topology that was helpful to prove this.
- By choosing a weaker (= coarser) topology, we
  made it easier to prove compactness, but more difficult
  to prove (lower semi-) continuity
- I haven't told you how to prove l.s.c. & coercivity
  Here convexity will play an important role
- Next week more convexity.
Recap: weak topologies

\[ X^* = \{ x^* : X \to \mathbb{R} \text{ linear, } \| x^* \| \text{ bounded} \} \]

\[ (X^*, \| \cdot \|_{X^*}) \text{ is a Banach space} \]

**Definition**

A sequence \((x_n)\) converges weakly to \(x\) if \(\langle x_n, x^* \rangle \to \langle x^*, x \rangle\) for all \(x^* \in X^*\).

*The weak topology \(\sigma(X, X^*)\) is the topology generated by the subbase \(\{ \{ x : \| x^*, x - x_0 \| < \varepsilon \} : x \in X^*, x_0 \in X, \varepsilon > 0 \}\).*

**Remarks:**
- Weak topologies are "Hausdorff," which implies that limits are unique.
- Weak topologies are not metrisable, which implies that convergent sequences are not enough to fully characterise the topology.
- Weak topologies are so-called "locally convex vector spaces (LCS)."
- This is beyond the scope of this course.

**Definition**

A sequence \((x^*_n)\) converges weak-* to \(x^*\) if \(\langle x^*_n, x^* \rangle \to \langle x^*, x^* \rangle\) for all \(x^* \in X^*\).

*The weak-* topology is the topology on \(X^*\) generated by the subbase \(\{ \{ x^* : \| x^*_n - x^* \| < \varepsilon \} : x \in X, \varepsilon > 0 \}\).*

**Theorem (Banach-Alaoglu)** The balls \(\{ \| x^* \|_{X^*} \leq 1 \}\) in \(X^*\) are weakly-* compact.

**Corollary** Any bounded set \(A \subseteq X^*, \sup_{x^*} \| x^* \|_{X^*} \leq C\) is relatively weakly-* compact.
Recall \( x^* \in X^* \iff x^*: X \to \mathbb{R} \) linear & bounded (\( |x^*(x)| \leq \|x^*\| \|x\| \))

\( x^*: X \to \mathbb{R} \) linear & \( \|x^*\| = \sup_{\|x\| \leq 1} \langle x^*, x \rangle < \infty \)

\( \text{in fact} \quad x^*: X \to \mathbb{R} \) linear & continuous

In literature sometimes written as \( X' \); however \( X' \) is also sometimes used for the space of all linear functionals.

We write \( \langle x^*, x \rangle = x^*(x) \) to stress "bilinearity", i.e.,

\[
\langle x^* ax_1 + bx_2, x \rangle = a \langle x^*, x \rangle + b \langle x^*, x \rangle
\]

and

\[
\langle ax^*, bx_2, x \rangle = a \langle x^*, x \rangle + b \langle x^*, x \rangle
\]

Three topologies:

\[
x_n \to x \quad \text{("strong" in norm"}) \quad : \iff \|x_n - x\| \to 0
\]

\[
x_n \to x \quad \text{("weak")} \quad \iff \langle x^*, x_n \rangle \to \langle x^*, x \rangle \forall x^* \in X^*
\]

\[
x_n^* \to x^* \quad \text{("weak-*")} \quad \iff \langle x^*, x_n \rangle \to \langle x^*, x \rangle \forall x \in X
\]

From exercise 1:

\[
x_n^* \to x^* \Rightarrow x_n^* \to x^* \Rightarrow x_n \to x \Rightarrow x_n \to x
\]

A. Examples and properties of weak/weak-* topologies

\( X = L^2(0, 2\pi) \) (Hilbert space!), \( x_n(t) = \sin(nt) \)

Th (Riesz representation theorem for Hilbert spaces):

\( L^2(0, 2\pi)^* = L^2(0, 2\pi) \), i.e.,

any \( x^* \in L^2(0, 2\pi)^* \) is of the form \( x^*(x) = \langle x^*, x \rangle = \int_0^{2\pi} x^*(t)x(t) dt \)

for some \( \tilde{x}^* \in L^2(0, 2\pi) \), and \( \|x^*\| = \|\tilde{x}^*\| \)

\( \|x\|_{L^2} \quad \text{is customary to identify} \quad x^* \quad \text{with} \quad x^* \)

\( \text{Since } L^2 \text{ is Hilbert, it is its own dual and predual.} \)

Hence weak and weak-* convergence are the same!

\( x_n = \sin(nt) \) does not converge in norm.

Since \( L^2 \) is Hilbert, it is its own dual and predual.

Hence weak and weak-* convergence are the same!

\( x_n = \sin(nt) \) does not converge in norm.

However, \( \|x\|_{L^2} = \int_0^{2\pi} \sin^2(nt) dt \leq 2\pi \) hence by Banach-Alaoglu there exists at least a weak-* (weakly) conv. subseq.
Prop $X_n \to 0$

Uses the following lemma:

Lem $C_c^\infty (0, 2\pi) \subset L^2(0, 2\pi)$ in the norm topology $\nabla$ 

smooth functions with compact support $\nabla$ 

(actually true for any $L^p$, $1 \leq p < \infty$)

proof of prop:

Take an arbitrary "test function" $\varphi \in L^2(0, 2\pi)$, and approximate $C_c^\infty \ni \varphi_m \to \varphi$. For each such $\varphi_m$:

$$\langle \varphi_m, X_n \rangle = \int_0^{2\pi} \varphi_m(t) \sin nt \, dt = \frac{1}{n} \int_0^{2\pi} \varphi_m(t) \cos nt \, dt$$

$$\leq \frac{1}{n} \| \varphi_m \|_{L^1} \to 0.$$ 

Then

$$\| \langle \varphi, X_n \rangle - \langle \varphi, 0 \rangle \| = \| \langle \varphi_m, X_n \rangle - \langle \varphi_m - \varphi, X_n \rangle \|$$

(approximation)

$$\leq \| \langle \varphi_m, X_n \rangle \| + \| \varphi_m - \varphi \|_L^2 \| X_n \|_L^2$$

$$\leq \| \langle \varphi_m, X_n \rangle \| + 2\pi \| \varphi_m - \varphi \|_L^2$$

$$\to 0 \quad \text{as} \quad m \to \infty.$$ 

(Exercise 2)

II

Prop Assume $X$ has a predual (and the norms coincide).

Then $F(x) = \| x \|_X = \| x \|_{X^*}$ is weakly-* l.s.c.

Proof: Take any sequence $X_n \to x$. Then:

$$\liminf_{n \to \infty} \| X_n \|_{X^*} = \liminf_{n \to \infty} \sup_{\| y \|_{X^*} \leq 1} \langle X_n, y \rangle$$

(for any $\| y \|_{X^*} \leq 1$)

$$\geq \liminf_{n \to \infty} \langle X_n, y \rangle = \langle x, y \rangle.$$

Now take the supremum over $\| y \|_{X^*} \leq 1$ on both sides:

$$\liminf_{n \to \infty} \| X_n \|_{X^*} \geq \| x \|_{X^*}.$$ 

$\to$ More general principle: a supremum over lsc functions is always l.s.c.

$\to$ We give a more precise definition of l.s.c. later...

$\to$ Will see: any lsc (cvx functional) is $\sup$ (affine, cont. functions)
Exercise 3: Can you construct a convex discontinuous function $x : \mathbb{R} \to \mathbb{R}$?
- Impossible in finite dimensions! (We come back to this)

B. Topological properties of convex sets

Definition: A set $A \subseteq X$ is convex if $\forall x, x_2 \in A$ and $\sigma \in [0, 1]$, $(1-\sigma)x + \sigma x_2 \in A$.

Theorem (Hahn-Banach geometric/separation Theorem)

A, B $\subseteq X$ disjoint convex sets ($A \cap B = \emptyset$)
A compact (in norm topology)
B closed

Then there exists a "separating hyperplane"

$$H = \{x : \langle x^*, x \rangle = \gamma \}$$
for some $x^* \in X^*$ and $\gamma \in \mathbb{R}$, and $\varepsilon > 0$ s.t.

$$\langle x^*, a \rangle \leq \gamma - \varepsilon \ \forall a \in A$$
$$\langle x^*, b \rangle \geq \gamma + \varepsilon \ \forall b \in B$$

Brezis Th. 1.7

The proof is beyond the scope of this course, however...

... Keep in mind that the proof uses the axiom of choice!

Hence the proof is not constructive!
Recall the subbase of the weak topology:
\[ \{ \{ \langle x^*, x - x_0 \rangle : 1 \leq \varepsilon \} : x^*, x_0, \varepsilon \}. \]
In particular, sets of the form \( \{ \langle x^*, x \rangle < \varepsilon \} \) are weakly open.
Moreover, \( \sigma(X, X^*) < \sigma(1, \|x\|) \), i.e. the weak topology is coarser/weaker than the norm topology.

Hence, \( C \subseteq X \) weakly closed
\[ C \in \sigma(X, X^*) < \sigma(1, \|x\|) \]
\[ \Rightarrow \]
\[ C \text{ also norm closed.} \]

Remarkably:

**Theorem:** If \( C \subseteq X \) is convex, then:
\[ C \text{ weakly closed} \iff C \text{ norm closed} \]

**Proof:**

\( \Rightarrow \) trivial (just proven above)

\( \Leftarrow \) Let \( C \) be norm-closed, and take an arbitrary \( x_0 \in C \).
We will show that there is a "weakly open ball" \( x_0 \in V \subseteq C \), and hence \( C \) must be weakly open.
Observe that \( \{x_0\} \) is convex and compact.

**Hahn-Banach:** \( \exists x^* \in X^*, \gamma \in \mathbb{R} \) s.t.
\[ \langle x^*, x_0 \rangle < \gamma < \langle x^*, x \rangle \quad \forall x \in C. \]

Let \( V := \{ x \in X : \langle x^*, x \rangle < \gamma \} \). Then:

1. \( x_0 \in V \)
2. \( \forall x \in C \Rightarrow x \in V \)
3. \( V \) is weakly open.
\[ \sigma(X, X^\ast) \text{ is not metrisable: need to distinguish between topological and sequential properties} \]

**Def:** \( C \subseteq X \) closed (in some topology) \( \iff C \subseteq C^\circ \)

\( C \subseteq X \) sequentially closed \( \iff \forall C \text{ is closed under } \sigma \text{-convergent sequences, i.e.} \forall \ x_n \rightarrow x, (x_n)_{n=1}^\infty \subseteq C \implies x \in C. \)

**Cor:** For \( C \subseteq X \):

1. \( C \) is weakly closed
   \( \Downarrow \) ("sequence lemma"")
2. \( C \) is weakly sequentially closed
   \( \Downarrow \) (trivial; last week's exercise)
3. \( C \) is (norm) sequentially closed
   \( \Downarrow \) ("sequence lemma for metrisable spaces"")
4. \( C \) is (norm) closed
5. All equivalent if \( C \) is convex
A Lower semi-continuity

**Definition.** \( F: X \to [\mathbb{R} \cup \{\infty\}] \) is \( \text{Lsc} \) (in some topology \( \sigma \)) if:

- \( \Rightarrow \) All level sets \( \{F \leq C\} \) are closed.
- \( \Leftarrow \) For all convergent sequences \( x_n \to x \):
  \[
  \liminf_{n \to \infty} F(x_n) \geq F(x)
  \]

**Lemma (in some topology).**

\( F \) is seq. lsc. \( \iff \) All level sets \( \{F \leq C\} \) are seq. closed.

\( \Rightarrow \) Take a convergent sequence \( (x_n)_n \subseteq \{F \leq C\} \), \( x_n \to x \).

Then \( F(x) \leq \liminf F(x_n) \leq C \), hence \( x \in \{F \leq C\} \).

\( \Leftarrow \) Take a convergent sequence \( (x_n)_n \subseteq X \), \( x_n \to x \).

Pick an arbitrary \( \epsilon > 0 \) and a (convergent) subsequence for which \( F(x_{n_m}) \leq \left( -\frac{1}{\epsilon} \right) \liminf_{n \to \infty} F(x_n) + \epsilon =: C \).

Since \( \{F \leq C\} \) is seq.-closed: \( (\nu \text{ denotes maximum}) \)

\[
F(x) \leq C = \left( -\frac{1}{\epsilon} \right) \liminf_{n \to \infty} F(x_n) + \epsilon.
\]

As \( \epsilon \) was chosen arbitrarily:

\[
F(x) \leq \liminf_{n \to \infty} F(x_n). \quad \square
\]

**Lemma.** \( F: X \to [\mathbb{R} \cup \{\infty\}] \) convex \( \Rightarrow \) All level sets \( \{F \leq C\} \) convex.

\( \Rightarrow \) Beware: not necessarily the other way around!

**Corollary.**

For \( F: X \to [\mathbb{R} \cup \{\infty\}] \):

- \( \text{a) } F \text{ weakly lsc} \)
- \( \text{b) } F \text{ weakly seq. lsc} \)
- \( \text{c) } F \text{ (norm) seq. lsc} \)
- \( \text{d) } F \text{ (norm) lsc} \)

\( \text{b) equivalent if } F \text{ is convex (by Hahn-Banach)} \)
Recall: in the direct method one needs to prove
(1) compact level sets,
(2) lower semicontinuity (sequential)
in some topology. But if level sets are compact, then they are also closed, hence the l.s.c. is trivial!

**Def** (Epigraph). For an \( F : X \to IR \cup \{ \infty \} \)
\[ \text{epi}(F) := \{(x, c) \in X \times IR : F(x) \leq c \} \]
sometimes useful in proofs, for example:

\[ \text{epi}(F) \]

(should have been in Peypenquet, but it's very easy !)

**Prop** For an \( F : X \to IR \cup \{ \infty \} \):
- \( F \) convex
  \[ \iff \]
  \[ \text{epi}(F) \] convex
- \( \Rightarrow \) (last week exercise 1.8)
- All level sets \( \{ F \leq c \} \) are convex

**Prop** \( F : X \to IR \cup \{ \infty \} \) is lsc
\[ \iff \]
\[ \text{epi}(F) \] is closed in \( X \times IR \) (in product topology)

Peypenquet uses a different, equivalent definition of lsc.

**Proof** "\( \Rightarrow \)" Take an \( (x_0, c) \in \text{epi}(F)^c \), i.e. \( F(x_0) > c_0 \).
Let \( \mu := F(x_0) + c_0. \) By lsc, the level set
\[ \{ F \leq \mu \} \] is closed, hence
\[ \{ F > \mu \} \] is open, hence
\[ (x_0, c) \in \{ F > \mu \} \times (-\infty, \mu) \]
is open, and disjoint with \( \text{epi}(F) \).
\[ \{ (x, c) : F(x) > \mu > c \} \]
Hence \( \text{epi}(F)^c \) is open.

"\( \Leftarrow \)" If \( \text{epi}(F) \) is closed, then also (for any \( C \))
\[ \text{epi}(F) \cap (X \times \{ c \} ) = \{ F \leq C \} \times \{ c \} \]
closed.
It follows that \( \{ F \leq C \} \) is closed.
Lemma \( F: X \to IR \cup \{0\} \) convex, \( x_0 \in X \) (Pymanquert Prop 3.2)

i) \( F \) bdd from above on a nbhd of \( x_0 \)

ii) \( F \) is Lipschitz cont. on a nbhd of \( x_0 \)

iii) \( F \) is cont. in \( x_0 \)

iv) \( F(x_0) < C \Rightarrow (x_0, C) \in \text{int}(\text{epi}(F)) \)

i) \( \Rightarrow \) ii) There exists a nbhd \( B(x_0, 2r) = \{ ||x_0-x|| < 2r \} \) on which \( F(z) \leq K \) for all \( z \in B(x_0, 2r) \).

Will prove Lipschitz cont. on \( B(x_0, r) \).

Take any \( x, y \in B(x_0, r) \) and construct:

\[
\bar{y} = y + r \frac{y-x}{||y-x||} \quad \Rightarrow \quad y = \lambda \bar{y} + (1-\lambda)x, \quad \lambda = \frac{||y-x||}{||y-x|| + r} \\
\tilde{x} = 2x_0 - x \quad \Rightarrow \quad x_0 = \frac{1}{2} \bar{x} + \frac{1}{2} \tilde{x}.
\]

Then \( \bar{x}, \tilde{x} \in B(x_0, 2r) \) and so \( F(x), F(y) \leq K \). Convexity:

\[
F(y) \leq \lambda F(\bar{y}) + (1-\lambda)F(x) \leq \lambda K + (1-\lambda)F(x) \tag{1}
\]

\[
F(x_0) \leq \frac{1}{2} F(\bar{x}) + \frac{1}{2} F(x) \leq \frac{1}{2} K + \frac{1}{2} F(x) \tag{2}
\]

\[
F(y) - F(x) \leq \lambda (K - F(x_0)) \leq 2 \lambda (K - F(x_0)) \leq 2 ||y-x|| (K - F(x_0)) \tag{3}
\]

ii) \( \Rightarrow \) iii) Trivial.

Continuity (in norm topology)

We first need a technical result; its implications will be super cool...
(iii) \( \Rightarrow \) (iv) Take a \( C > F(x_0) \) and an arbitrary \( \eta \in (F(x_0), C). \) By continuity, there exists a ball such that \( F(z) < \eta \) \( \forall z \in B(x_0, \delta). \) Then \( (x_0, \eta) \in B(x_0, \delta) \times (\eta, \infty) \subset \text{epi}(F) \) open nbhd of \((x_0, C)\) and so \((x_0, \eta) \in \text{int}(\text{epi}(F)).\)

(iv) \( \Rightarrow \) (i) Take any \( C > F(x_0); \) since \((x_0, C) \in \text{int}(\text{epi}(F))\) there is again a ball and \( \eta \in (F(x_0), C) \) s.t.
\[
(x_0, \eta) \in B(x_0, \delta) \times (\eta, \infty) \subset \text{epi}(F).
\]
Therefore \( F(z) < \eta \) \( \forall z \in B(x_0, \delta). \)

**Proposition**

If \( F: X \to \mathbb{R} \cup \{\infty\} \) is convex, (Peypenquet and cont. at some \( x_0 \in \text{dom}(F), \) then Prop 3.3)
\( x_0 \in \text{int}(\text{dom}(F)) \) and \( F \) is cont. on whole \( \text{int}(\text{dom}(F)). \)

\( \Rightarrow \) Note that \( \text{dom}(F) = \bigcup_{C \subset \mathbb{R}} \text{int}(F \leq C) \) is always convex!

**proof.** Assume \( F \) is cvx, and cont. at some \( x_0 \in \text{dom}(F). \)
By the previous lemma \((\text{iii}) \Rightarrow (\text{i})\) \( x_0 \in \text{int}(\text{dom}(F)). \)
By the same lemma, \( F(x) \leq K \) on some \( B(x_0, r) \supset X. \)
Take an arbitrary \( y_0 \in \text{int}(\text{dom}(F)); \) we shall prove that \( y_0 \) is also held on a small neighborhood, so that \( F \) is continuous in \( y_0. \)
Since \( y_0 \) lies in the interior, one can find a \( \rho > 0 \) for which
\[
Z_0 := y_0 + \rho(y_0 - x_0) \in \text{dom}(F).
\]
\[
(y_0 = \frac{\rho}{1 + \rho} x_0 + \frac{1}{1 + \rho} Z_0)
\]
Take any \( y \in B(y_0, \frac{\rho}{2 + \rho}) \), and define \( w \) such that
\[
y = \frac{\rho}{1 + \rho} w + \frac{1}{1 + \rho} Z_0.
\]
Then \( w \in B(x_0, r) \) so \( F(w) \leq K. \) By convexity
\[
F(y) \leq \frac{\rho}{1 + \rho} F(w) + \frac{1}{1 + \rho} F(Z_0) \leq K \vee F(Z_0). \]
\( \Box \)
So far we haven't used completeness of the space $X$ in finite dimensions:

**Prop** Let $X$ be finite-dimensional and $F : X \to \mathbb{R} \{\infty\}$ convex. Then $F$ is continuous on $\text{int}(\text{dom}(F))$. (Peypouquet prop 3.5)

→ First week exercise 3.
→ Recall: all norms on a finite-dim. space are equivalent ⇒ we may take $X = \mathbb{R}^d$.

**proof** For any $x \in \text{int}(\text{dom}(F))$, take a ball $B(x, \varepsilon)$.

By convexity, take a $p > 0$ such that $x + p e_i \in \text{dom}(F)$, $i = 1, \ldots, d$, and let $V$ be the convex hull of these points, i.e.:

$$V := \text{co}\{x + p e_i : i = 1, \ldots, d\}$$

By convexity, $V \subseteq \text{dom}(F)$ and $V \subseteq \text{int}(\text{dom}(F))$. Hence $x$ is held above on a neighborhood $\Rightarrow$ cont. in $x$. □

In infinite dimensions:

**Prop** Let $X$ be a (complete!) Banach space and $F : X \to \mathbb{R} \{\infty\}$ convex, and l.s.c.. Then $F$ is continuous on $\text{int}(\text{dom}(F))$. (Peypouquet prop 3.6)

→ Proof beyond the scope of this course (based on choice axiom, although it may be circumvented if the space is separable).

→ Beware that these continuity results are all in the norm topology.

→ However, we saw before that norm (lower-semi-)continuity implies weak lower semicontinuity if the functional is convex.
L2, ex. 1) in $\mathbb{R}$ (or in any finite dimensions), the norm, weak and weak-* topologies coincide!

L2, ex. 2) $C_c(\mathbb{R})^* \cong \{\text{regular, signed measures on } \mathbb{R}\}$

$C_b(\mathbb{R})^* \cong \{\text{regular, finitely additive set functions on } \mathbb{R}\}$

$(S_n)_{n \in \mathbb{N}} \subset C_c(\mathbb{R})^* \subset C_b(\mathbb{R})^*$

a) weak-* compact in $C_c(\mathbb{R})^*$? "vague topology"

$$\|S_n\|_{C_c^*} = \sup_{\|\varphi\|_{C_c} \leq 1} \varphi(n) = 1.$$  

Hence $(S_n)_n \subset B(c, 1); \text{ weak-* compactness follows by Banach- Alaoglu.}$  

(Rudin, Funct. Analysis Th 2.17)

**Theorem.** If $X$ is a separable Banach space, then any weakly-* compact $K \subset X$ is metrisable and hence weakly-* sequentially compact.

$C_c(\mathbb{R})$ is separable, so $(S_n)_n$ has a convergent subsequence (against $C_c(\mathbb{R})$).

b) Limit? For any $\varphi \in C_c(\mathbb{R})$:

$$<S_n, \varphi> = \varphi(n) \frac{\text{ultimately}}{n \to \infty} \to 0 \text{ and so } S_n \xrightarrow{w*} 0.$$  

But not in norm: $\|S_n - 0\|_{C_c^*} \equiv 1 \not\to 0$.

(Recall: strong convergence $\Rightarrow$ weak * convergence $\&$ limits coincide)

c) weak-* compact in $C_b(\mathbb{R})^*$ "weak or narrow topology"

$$\|S_n\|_{C_b^*} = \sup_{\|\varphi\|_{C_b} \leq 1} \varphi(n) = 1 \Rightarrow \text{weak* compact by Banach- Alaoglu.}$$

However, $C_b(\mathbb{R})$ is not separable, and so we can not deduce weak-* sequential compactness!
L3, ex 1. \( F : X \to \mathbb{R} \cup \{0\} \) convex & (norm) lsc
\[ \Rightarrow \text{weakly lsc} \]

short proof using the epigraph:

\[ F \text{ convex & lsc} \iff \text{epi}(F) \text{ convex & closed} \]
\[ \Rightarrow \text{epi}(F) \text{ convex & weakly closed} \]
\[ \iff F \text{ convex & weakly lsc} \]
A1 Differentiation in Banach spaces

First consider finite dimensions: \( F: \mathbb{R}^2 \to \mathbb{R} \).
\[
\nabla F(x) := \left[ \frac{\partial F(x)}{\partial x_1} \frac{\partial F(x)}{\partial x_2} \right].
\]
Why is it meaningful to know only the derivatives in the directions \( x_1 \) and \( x_2 \)? Well, if \( F \) is differentiable, then
\[
dF(x; h) := \lim_{\epsilon \to 0} \frac{F(x+\epsilon h) - F(x)}{\epsilon} \text{ is a linear operator acting on directions } h.
\]

Similarly in an (infinite-dim.) Banach space: [Peyman P.14]

Def \( F: X \to \mathbb{R} \) (or \( \text{dom} F : \to \mathbb{R} \)) is Gâteaux differentiable if
\[
dF(x; h) := \lim_{\epsilon \to 0} \frac{F(x+\epsilon h) - F(x)}{\epsilon}
\]
is a linear & bounded functional.

We write \( \langle x, dF(x)h \rangle \), and \( dF(x)^e \) is called the Gâteaux derivative of \( F \) in \( x \).

→ Peyman writes Gâteaux derivatives as \( \nabla \). There are good reasons not to do this...

Def \( F: X \to \mathbb{R} \) is \( 2 \) twice Gâteaux differentiable if
\[
d^2F(x; h, h) := \lim_{\epsilon \to 0} \langle dF(x+\epsilon h), h \rangle - \langle dF(x), h \rangle
\]
is a linear & bounded functional on \( h, h_2 \in X \times X \).

We write \( \langle x, d^2F(x)[h, h] \rangle \) or \( \langle x, \frac{\partial^2 F(x)}{\partial h_1 \partial h_2} \rangle \).

Example (Euler-Lagrange, formula)

(Peyman Example 1.2.8)

\[ F(x) := \int_0^T L(x(t), \dot{x}(t)) \, dt, x \in L^2(0, T) \]
\[
\lim_{\epsilon \to 0} \frac{F(x+\epsilon h) - F(x)}{\epsilon} = \frac{d}{d\epsilon} \int_0^T L(x(t)+\epsilon h(t), \dot{x}(t)+\epsilon \dot{h}(t)) \, dt \bigg|_{\epsilon \to 0}
\]
\[
= \int_0^T \dot{x}(t) \frac{\partial L(x(t), \dot{x}(t))}{\partial \dot{x}} \, dt + \int_0^T h(t) \frac{\partial L(x(t), \dot{x}(t))}{\partial x} \, dt + \int_0^T \dot{h}(t) \frac{\partial L(x(t), \dot{x}(t))}{\partial \dot{x}} \, dt
\]
\[
= \langle \frac{\partial L(x, \dot{x})}{\partial \dot{x}}, h \rangle - \langle \frac{\partial^2 L(x, \dot{x})}{\partial x \partial \dot{x}}, \dot{h} \rangle
\]
\[
= :DF(x)[h] \in L^2(0, T) \text{ (differentiable if this!)}
\]
Proposition. Let \( A = \text{dom} F \) be open and convex, \( F: A \to \mathbb{R} \) be Gâteaux differentiable.\(^{\text{Prop 3.10}}\)

1) \( F \) is convex

\[ F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y) \]

\[ F(x + \lambda(y-x)) - F(x) \leq \frac{\lambda}{\lambda} F(y) - F(x) \]

\[ \frac{\langle DF(x), y-x \rangle}{\lambda} \leq \frac{F(y) - F(x)}{\lambda} \]

\[ \lambda \to 0 \]

\[ \langle DF(x), y-x \rangle \leq F(y) - F(x) \]

\[ \langle DF(y), x-y \rangle \leq F(x) - F(y) + \langle DF(x) - DF(y), y-x \rangle \]

ii) \( \Rightarrow \) iii)

(iii) \( \Rightarrow \) i)

(not using twice differentiability! We don't really need this)

For any \( x, y \in A \), define

\[ \phi(\lambda) = F(\lambda x + (1-\lambda)y) - \langle DF(x), x-y \rangle - \lambda F(y) \quad \lambda \in [0, 1] \]

\[ \phi(\lambda) = \langle DF(\lambda x + (1-\lambda)y), y-x \rangle + F(x) - F(y) \]

If \( 0 < \lambda_1 < \lambda_2 < 1 \) and \( x_1 = (1-\lambda_1)x_1 + \lambda_1 y, \quad x_2 = (1-\lambda_2)x_2 + \lambda_2 y \), then

\[ \phi'(\lambda_2) - \phi'(\lambda_1) = \langle DF(x_2) - DF(x_1), y-x \rangle + F(x) - F(y) + F(y) - F(x) \]

\[ = \langle DF(x_2) - DF(x_1), x_2-x_1 \rangle \geq 0 \text{ by assumption} \]

(\( \phi' \) is non-decreasing. If \( \phi' \geq 0 \) then clearly \( \phi' \) is non-increasing.

Since \( \phi(0) = 0 = \phi(1) \), \( \exists \lambda \in (0, 1) \) s.t. \( \phi'(\lambda) = 0 \).

But \( \phi' \) is non-decreasing, so \( \phi' [0, \lambda] \leq 0 \) and \( \phi' [\lambda, 1] \geq 0 \).

Hence \( \phi(1) \leq 0 \)

\[ \langle DF(x+\epsilon h) - DF(x), eh \rangle \geq 0 \]

\[ DF(\epsilon h, h) \geq 0 \]
Proposition  \[ \text{strictly convex} \iff \text{strict inequalities} \]

Example (From lecture 4)

\[ X = \mathbb{R}^2, \quad F(x, y) = x^2 - \alpha xy + y^2 \]

\[ \nabla F(x, y) = \begin{bmatrix} 2x - \alpha y \\ -\alpha x + 2y \end{bmatrix} \]

\[ \nabla^2 F(x, y) = \begin{bmatrix} 2 & -\alpha \\ -\alpha & 2 \end{bmatrix} \]

eigenvalues:

\[ \det \begin{bmatrix} 2 - \lambda & -\alpha \\ -\alpha & 2 - \lambda \end{bmatrix} = (2 - \lambda)^2 - \alpha^2 \Rightarrow \lambda = 2 \pm \alpha \]

pos. semidef \[ \iff \alpha \in [-2, 2] \]

\[ \nabla F \text{ is always convex in } x \text{ and in } y \]

\[ \nabla F \text{ only convex in } (x, y) \text{ ("jointly convex") } \iff \alpha \in [-2, 2] \].
Cor: The hyperplane \( V_x = \{ (y, z) \in \mathbb{R}^2 : F(x) + \langle g(x), y - x \rangle = z \} \) lies below \( \text{epi}(F) \).

**Subdifferentials**

**Def.** For \( F : X \to (\mathbb{R} \cup \{\infty\}) \) (proper), the subdifferential \( \partial F(x) \) at \( x \in X \) is the set:

\[
\partial F(x) = \{ x^* \in X^* : F(y) \geq F(x) + \langle x^*, y - x \rangle \ \forall y \in X \}
\]

The set \( \partial F(x) \) is the set of **slopes** of the hyperplane below the epigraph of \( F \).

**Example:**

\[
F : \mathbb{R} \to \mathbb{R}, \quad F(x) = x
\]

\[
\partial F(x) = \begin{cases} 
-1, & x < 0, \\
\{0\}, & x = 0, \\
\{1\}, & x > 0.
\end{cases}
\]

**Prop.** If \( F \) is Gâteaux differentiable in \( x \in X \), then \( \nabla F(x) = x^* \)

\[
\partial F(x) = \{ x^* \} \quad \text{(i.e. the subdifferential is a singleton)}
\]

**Proof.** "\( \Rightarrow \)" Clearly \( \nabla F(x) \in \partial F(x) \); need to prove that this is the *only* one.

Take any \( x^* \in \partial F(x) \), so that \( \forall t > 0, \ h \in X \):

\[
F(x + th) \geq F(x) + \langle x^*, th \rangle
\]

\[
\frac{\langle \nabla F(x), h \rangle}{t} \xrightarrow{t \to 0} \frac{F(x + th) - F(x)}{t} \geq \langle x^*, h \rangle \quad \forall h \in X.
\]

Hence \( \nabla F(x) = x^* \).

"\( \Leftarrow \)" Will be proven later... □
Note that for the example \( F(x) = 1 \times 1 \), the subdifferential is always a closed interval. In higher (possibly infinite) dimensions, convex sets can be seen as generalisations of intervals... (Peyton (Prop. 3.21))

**Prop**

\[ F : X \rightarrow \mathbb{R} \cup \{ \infty \} \] convex. For any \( x \in X \), the subdifferential \( \partial F(x) \) is (norm) closed and convex.

**Proof**

**Convex:** Take \( x_1^*, x_2^* \in \partial F(x) \) and \( \lambda \in (0, 1) \). For all \( y \in X \):

\[
\begin{align*}
F(y) &\geq F(x) + \langle x_1^*, y - x \rangle + \lambda (1 - \lambda) \\
F(y) &\geq F(x) + \langle x_2^*, y - x \rangle \\
F(y) &\geq F(x) + \langle \lambda x_1^* + (1 - \lambda) x_2^*, y - x \rangle, \text{ hence} \\
\lambda x_1^* + (1 - \lambda) x_2^* &\in \partial F(x).
\end{align*}
\]

**Closed:** The norm topology is metric, so we only need to prove sequential closedness. Take \( x_n \rightarrow x \in \partial F(x) \).

\[
\begin{align*}
F(y) &\geq F(x) + \langle x_n^*, y - x \rangle \\
\downarrow &\quad n \rightarrow \infty \\
F(y) &\geq F(x) + \langle x^*, y - x \rangle \quad \square \quad (Peyton (Prop. 3.22))
\end{align*}
\]

**Prop** (monotonicity) \( F : X \rightarrow \mathbb{R} \cup \{ \infty \} \). If \( x^* \in \partial F(x) \), then \( \langle x^*, y - x \rangle \geq 0 \).

**Proof:** exercise \( \circ \)

**Th** (Fermat's Rule) \( F : X \rightarrow \mathbb{R} \cup \{ \infty \} \) (proper and) convex. \( x \) is a global minimiser of \( F \) iff \( 0 \in \partial F(x) \).

**Proof:** exercise \( \circ \) (USC sufficient) (Peyton 3.25)

**Prop** \( F : X \rightarrow \mathbb{R} \cup \{ \infty \} \) convex and \( F \) cont. in \( x \in \text{dom}(F) \). Then \( \partial F(x) \) is non-empty and bounded (and closed and convex).
We need an alternative Hahn-Banach Theorem:

**Theorem (Hahn-Banach geometric / separation Theorem)**

A, B ⊂ X disjoint, empty, convex & A is open.

Then ∃ x* ∈ X* \ {0} s.t.

\[ \langle x^*, a \rangle < \langle x^*, b \rangle \quad \forall a \in A, b \in B \]

(Penotuge Th. 1.10)

(Brezis Th. 1.6)

Proof of the prop. Take A, B ⊂ X x IR:

@ non-empty

(open) A := int (epiF) ≠ ∅ since F is cont. at at least one point

B := \{(x, F(x))\} \quad \text{(and (x, F(x)) ∉ int(epiF) so \(A \cap B = \emptyset\))}

Hahn-Banach: ∃ \((x^*, s) \in X \times R \setminus \{0, 0\}\) s.t.

\[ \langle (x^*, s), (x, C) \rangle < \langle (x^*, s), (x, F(x)) \rangle \quad \forall \ y, C \in \text{int(epi(F))} \]

\[ \langle x^*, y \rangle + sc < \langle x^*, x \rangle + sF(x) \]

For \( y = x \),

\[ \langle x^*, y \rangle + sc < \langle x^*, x \rangle + sF(x) \quad \text{so} \ s \leq 0. \]

\[ sF(x) \]

C > \( F(x) + \langle -x^*, y - x \rangle \)

\[ \text{C} \uparrow \text{C} \uparrow F(y) \]

\[ F(y) \geq F(x) + \langle -\frac{x^*}{s}, y - x \rangle \Rightarrow -x^* \in \partial F(x) \]

bounded. Take any \( x^* \in \partial F(x) \). Since F cont in x, F Lipschitz on nbd(x):

\[ F(x) + \langle x^*, y - x \rangle \leq F(y) \leq F(x) + M\|y - x\| \quad \forall \ y \in \text{nbd}(x) \]

\[ x^* \in \partial F(x) \quad \text{Lipschitz} \]

Hence \( \|x\| = \sup_{y \neq x} \frac{\langle x^*, y - x \rangle}{\|y - x\|} = \sup_{y \neq x} \frac{\langle x^*, y - x \rangle}{\|y - x\|} \leq M. \)
Lecture 6  

A envelopes

\[ F : X \rightarrow \mathbb{R} \cup \{\infty\} \]

If \( F \) is not convex, can it be "convexified"?
If \( F \) is not lsc, can it be "lsc-ified"?

For a set \( A \subset X \):
\[ \overline{A} := \text{closure of } A := \text{smallest closed set containing } A \]
\[ \text{co } A := \text{convex hull of } A := \text{smallest convex set containing } A \]
\[ = \{ \sum_{i=1}^{n} \sigma_i a_i : \sigma \in \mathbb{P}(n) \text{ and } (a_i)_i \subseteq \text{co } A, n \geq 1 \} \]

For a functional \( F : X \rightarrow \mathbb{R} \cup \{\infty\} \)
\[ \overline{F} = \text{lsc envelope of } F := \text{largest lsc functional below } F \]

well-defined since \( F(x) = \sup_{G \leq F} G(x) \) is lsc (see below)
\[ \text{co } F = \text{convex envelope of } F := \text{largest cvx functional below } F \]

well-defined since \( (\text{co } F)(x) = \sup_{G \leq F} G(x) \) is convex (see below)

Recall \( \text{epi } F = \{(x, c) : F(x) \leq c\} \). Then (we already proved this)

Prop: \( \sup_{G \leq F} G(x) \) is lsc
Proof: \( \text{epi } (\sup_{G \leq F} G) = \bigcap_{G \leq F} \text{epi } G \text{ is closed/cvx } \)

General principle: \( \sup \) over lsc/cvx functions is lsc/cvx.

Special role played by affine functions:

Def: A continuous affine functional \( F : X \rightarrow \mathbb{R} \) is a functional of the form
\[ F(x) = \langle x^*, x \rangle + \alpha \quad \text{for some } x^* \in X^*, \alpha \in \mathbb{R} \]
Proposition: \( F: X \to \mathbb{R} \cup \{0\} \) (proper).
(Peypouquet Prop 3.1)

\( F \) is convex & lsc \( \iff \exists \) family \( \{F_i\}_{i \in I} \) of continuous affine functions such that
\[
F(x) = \sup_{i \in I} F_i(x).
\]

\[ \begin{align*}
\text{proof } & \Rightarrow \ " \text{we already proved this (epigraph)} \\
\Rightarrow & " \text{ Let } F \text{ be convex & lsc; Need to construct a family of continuous affine functions. We shall take } I = X \times \mathbb{R}_+. \text{ Pick } x_0 \in X, \ \delta > 0. \\
& \text{ We shall prove that for any } x \in X \text{ there exist a cont. affine } \\
& F_{x_0, \delta}: X \to \mathbb{R} \text{ such that } \\
& \quad (i) \quad F_{x_0, \delta}(x) \leq F(x) \quad \forall x \in X \\
& \quad (ii) \quad \frac{1}{\delta} \wedge (F(x_0) - \delta) \leq F_{x_0, \delta}(x_0) \leq F(x_0),
\end{align*} \]

hence
\[
F(x) = \sup_{x_0, \delta} F_{x_0, \delta}(x). \text{ How to construct } F_{x_0, \delta} \ldots?
\]

\[ \begin{align*}
\text{Hahn-Banach:} & \quad \text{epi}(F) \\
& \quad \text{Cvx} & \text{Closed} & \text{compact} \\
& \quad \{x_0, \frac{1}{\delta} \wedge (F(x_0) - \delta)\} & \text{compact}
\end{align*} \]

\[ \exists (x^*, s) \in X^* \times \mathbb{R} \setminus \{0, 0\} \text{ and } \epsilon > 0 \text{ such that } \forall (x, c) \in \text{epi} F:
\]
\[
\langle x^*, x \rangle + s \left( \frac{1}{\delta} \wedge (F(x_0) - \delta) \right) = \langle x^*, x_0 \rangle + \frac{1}{\delta} \wedge (F(x_0) - \delta) + \epsilon \leq \langle x^*, x \rangle, (x, c) = (x^*, x) + sC(x^*)
\]

since \( C \) is arbitrarily large, \( s > 0. \) Two cases:

\( s > 0 \) (wlog assume \( s = 1 \))

\[ F_{x_0, \delta}(x) := \langle -x^*, x \rangle + \langle x^*, x_0 \rangle + \frac{1}{\delta} \wedge (F(x_0) - \delta) + \epsilon. \]

From (x) with \( x \in \text{dom} F \) and \( C = F(x) : i) F_{x_0, \delta}(x) \leq F(x) \\
(ii) F_{x_0, \delta}(x_0) = F(x_0) - \delta + \epsilon \geq F(x) - \frac{1}{\delta} \wedge (F(x_0) - \delta).
\]

\( s = 0 \) (then \( x_0 \in \text{dom} F \))

\[ G_{x_0, \delta}(x) := \langle -x^*, x \rangle + \langle x^*, x_0 \rangle + \epsilon \text{ and take } F_{x, \delta}(x) \text{ for some } \delta \in \text{dom} F. \]

\[ F_{x_0, \delta}(x) := F(x_0) + \delta G_{x_0, \delta}(x) \text{ for some } \delta \to \infty. \]
From (4): \( \forall x \in \text{dom} F \) \( G_{x_0} (x) = \langle x^*_0, x \rangle \geq 0 \)

Hence (i) \( \forall x \in \text{dom} F \) \( F_{x_0} \delta (x) \leq F_{x_0} g (x) \leq F(x) \)
\( \forall x \in \text{dom} F \) \( F_{x_0} g (x) \leq \infty = F(x) \).

Moreover ii) \( F_{x_0} g (x_0) = F_{x_0} g (x_0) + n \delta \geq \delta \) for \( n \delta \) sufficiently large.

Of particular interest will be \( F(x) = \sup_{x^*_0} \frac{\langle x^*_0, x \rangle - G(x)}{\text{affine function in } x} \).

Recall the definition of the subdifferential:
\( \partial F(x) := \{ x^* \in X^*: F(y) \geq F(x) + \langle x^*_0, y-x \rangle \quad \forall y \in X \} \)

this means that
\( \langle x^*_0, x \rangle - F(x) \geq \sup_y \langle x^*_0, y \rangle - F(y) \)

**Def** Convex/Moreau/Fenchel dual /conjugate
a.k.a. Legendre transform (Peypounet eq (3.17))
\( F^* : X^* \to \mathbb{R} \cup \{\infty\} \)
\( F^*(x^*_0) := \sup_{x \in X} \langle x^*_0, x \rangle - F(x) = \sup_{x \in \text{dom} F} \langle x^*_0, x \rangle - F(x) \)

**Examples**

(a) \( F(x) = \frac{1}{p} \| x \|_{L_p}^p \), \( 1 < p < \infty \). \( (X^* = L^{p_*}, \frac{1}{p} + \frac{1}{p_*} = 1) \)
\( F^*(x^*_0) = \sup_{x \in L^p} \langle x^*_0, x \rangle - \frac{1}{p} \| x \|_{L_p}^p \)

Does a maximizer exist? Superlevel set:
\( \{ \langle x^*_0, x \rangle - \frac{1}{p} \| x \|_{L_p}^p \geq C \} \)
\( C \leq \langle x^*_0, x \rangle - \frac{1}{p} \| x \|_{L_p}^p \leq \| x^*_0 \| \| x \|_{L_1} - \frac{1}{p} \| x \|_{L_p}^p \leq \| x^*_0 \| \| x \|_{L_1} (\| x \|_{L_1} - \frac{1}{p} \| x \|_{L_p})^\alpha \)

if \( \| x \|_{L_p} \leq 1 \) then \( \| x \|_{L_p} \leq 1 \) (duh)
if \( \| x \|_{L_p} \geq 1 \) then
\( \frac{1}{p} \| x \|_{L_p}^{p-1} - \| x^*_0 \|_{L_p}\| x \|_{L_p} \leq -\frac{C}{\| x \|_{L_1}} \leq 1 C \)
\( \| x \|_{L_1} \leq \sqrt[p]{(1C + \| x^*_0 \|_{L_p})} \)

Banach-Alaoglu: weak* cpt super level sets.
hence also weak-* closed super-level sets, i.e. upper semicontinuity.

Direct method: there exists a maximiser.
(even unique by strict concavity)

Maximiser is a critical point. Gâteaux derivative:
\[ 0 = \lim_{\varepsilon \to 0} \frac{\langle x^*, x + \varepsilon h \rangle - \frac{1}{p} \| x + \varepsilon h \|_p^p - \langle x^*, x \rangle - \frac{1}{p} \| x \|_p^p}{\varepsilon} \]
\[ = \frac{d}{d \varepsilon} \langle x^*, x + \varepsilon h \rangle \bigg|_{\varepsilon = 0} = \frac{1}{p} \int (x + \varepsilon h)^{p^* - 1} h \, \text{d} \varepsilon \bigg|_{\varepsilon = 0} \]
\[ = \int (x^* - x^{p^*}) h \]
\[ \Rightarrow x = \sqrt[p^*]{x^*} \text{ (assuming } x^* \geq 0 \text{ a.e.)} \]

\[ F^*(x^*) = \sup_x \langle x^*, x \rangle - \frac{1}{p} \| x \|_p^p = \int x^*^{p^*} + \frac{p}{p^* - 1} - \frac{1}{p} \int x^*^{p^*} \]
\[ = \frac{1}{p^*} \int |x|^p^{p^*} = \frac{1}{p^*} \| x^* \|_{L^{p^*}}^{p^*} \]

b) \( F(x) = \langle x^*, x \rangle + \alpha \quad \text{(continuous affine)} \) (Peypouquet example 3.46)

\[ F^*(x^*) = \sup_x \langle x^*, x \rangle - \langle x^*, x^* \rangle - \alpha \]
\[ = \sup_x \langle x^* - y^*, x \rangle - \alpha = \begin{cases} \infty, & x^* \neq y^* \\ -\alpha, & x^* = y^* \end{cases} \]

C) \( F(x) = \chi_C(x) := \begin{cases} \infty, & x \in C \\ 0, & x \not\in C \end{cases} \quad \text{"characteristic function"} \)

\[ F^*(x^*) = \sup_{x \in C} \langle x^*, x \rangle - \chi_C(x) \]
\[ = \sup_{x \in C} \langle x^*, x \rangle \quad \text{"support function"} \]
Lecture 7
Answers to exercises

3. \textbf{Prop} \quad F : X \to (\mathbb{R} \cup \{\infty\}) \quad (\text{proper})

\begin{align*}
F \text{ convex \& lsc} & \iff \exists \text{ family } (F_i)_{i \in I} \text{ of con. affine functions such that } F(x) = \sup_{i \in I} F_i(x) \\
\text{We proved } \Rightarrow \text{ by Hahn-Banach,} \\
\text{Can also be proven more directly!} \\
\text{We already know that convex \& lsc implies nonempty subdifferential.} \\
\text{(also based on a Hahn-Banach theorem)} \\
\text{Hence } F(x) = \sup_{x \in X} \left( F(x) + \langle x^*, x-x_0 \rangle \right). \\
\end{align*}

4. \quad F : \mathbb{R} \to \mathbb{R} \\
F(x) = b(e^x - 1) \iff F^*(x^*) = s(x^*|b) = \\
(b > 0) \\
= b \lambda_b(\frac{x^*}{b}) = \begin{cases} 
 b, & x = 0 \\
 x \log \frac{x^*}{b} - x + b, & x > 0 \\
 \infty, & x < 0 \\
\end{cases} \\
\text{"Boltzmann function,"} \\
\Rightarrow F^{**} : \mathbb{R} \to \mathbb{R} \\
F^{**}(x^{**}) = b(e^{x^{**}} - 1).
Prop (properties of convex duals)

\[ F^* \] is convex & lsc (in norm and even in weak-* topology)

1. \( F \leq G \Rightarrow F^* \geq G^* \) (pointwise)

2. \( F^*(x^*) = \sup_{x \in X} \langle x^*, x \rangle - F(x) = \sup_{x \in \text{dom} F} \langle x^*, x \rangle - F(x) \)

3. If \( A : Z \to X \) is a linear bounded operator

   \[ \|A\| = \sup_{\|z\| \leq 1} \|Az\| < \infty \]

   \[ \langle Ax^*, z \rangle_z = \langle x^*, Az \rangle_x \]

   \( A^T : X^* \to Z^* \) adjoint operator

   (corresponds to transpose for matrices)

   (more commonly denoted by \( A^* \), but this gets confusing in convex analysis)

   and \( F(x) = \inf_{z \in Z} \langle H(z), z \rangle \), \( H : Z \to \mathbb{R} \cup \{\infty\} \)

\[ \Rightarrow F^*(x^*) = H^*(A^T x^*) \]

4. \( F(x) = G(\alpha x) \Rightarrow F^*(x^*) = G^*(\frac{x^*}{\alpha}) \), \( \alpha \in \mathbb{R} \setminus \{0\} \)

5. \( F(x) = x^* G(x) \Rightarrow F^*(x^*) = x^* G^*(\frac{x}{x^*}) \), \( x^* \in \mathbb{R} \setminus \{0\} \)

6. \( F(x) = G(x + x_0) \Rightarrow F^*(x^*) = G^*(x^*) - \langle x^*, x_0 \rangle \)

7. \( F(x) = G(x) - \langle y^*, x \rangle \Rightarrow F^*(x^*) = G^*(x^* + y^*) \)

8. \( F(x) = G(x) + H(x) \Rightarrow F^*(x^*) = (G^* \ast H^*)(x^*) \)

   \[ = \inf_{\alpha^*, b^* \in X} G^*(\alpha^*) + H^*(b^*) \]

   "inf-convolution" (sometimes denoted by \( *_{\inf} \))

9. \( F(x) + F^*(x^*) \geq \langle x^*, x \rangle \) Fenchel-Moreau-Young inequality

Proofs: exercise (all but VIII)

Remark: IX is trivial but very important none-the-less.

It generalises a "completing-the-squares" argument for quadratic functions (Young inequality)

\[ \frac{1}{2}x^2 + \frac{1}{2}y^2 = \frac{1}{2}(x-y)^2 + xy \geq xy \]
**Corollary**

\[ F(x) + F^*(x) = \langle x^*, x \rangle \]

\[ \iff \quad x \in \partial F^*(x^*) \]

**Proof**

\[ F(x) + F^*(y^*) - \langle y^*, x \rangle \geq 0 = F(x) + F^*(x^*) - \langle x^*, x \rangle \]

(Alternately, \( F(x) + F^*(\cdot) - \langle \cdot, x \rangle \) is minimized by \( x^* \to \text{Fenchel's rule} \))

**B The convex bidual**

**Def** For \( F : X \to \mathbb{R} \cup \{ \infty \} \) the bidual is

\[ F^{**} : X \to \mathbb{R} \cup \{ \infty \} \]

\[ F^{**}(x) := \sup_{x^* \in X^*} \langle x^*, x \rangle - F(x^*) \]

Why not defined on \( X^{**} \)? (Peyseronnet remark 3.54)

\[ (F^{**})^*(x^{**}) := \sup_{x^* \in X^*} \langle x^{**}, x^* \rangle - F(x^*) \]

restricted to \( X \subset X^{**} \):

\[ (F^{**})^*|_X = F^{**} \]

**Def (Canonical embedding)**

For \( x \in X \), let \( S_x : X^* \to \mathbb{R} \) be the linear mapping

\[ \langle x^*, S_x \rangle = S_x(x^*) := \langle x^*, x \rangle \]

Then \( 1 \leq \|x\| \|x^*\| \Rightarrow S_x \text{ hold/cont.} \)

Hence, identifying \( x \) with \( S_x \):

\[ X \subset X^{**} \]

**Prop** \( F : X \to \mathbb{R} \cup \{ \infty \} \) proper.

\( F \) is convex & lsc \( \iff F = F^{**} \)

**Proof**

\( \Rightarrow \) \( F^{**} \) is a supremum over cont. affine functions.

\( \Leftarrow \) \( F^{**}(x) = \sup_{x^*} \langle x^*, x \rangle - F^*(x^*) \leq F(x) \)

On the other hand, \( F(x) = \sup_{i \in I} F_i(x) \) for family of cont. affine functions.

\[ F^*(x^*) = \sup_{x^*} \inf_{i \in I} \langle x^*, x \rangle - F_i(x) \leq \inf_{i \in I} \sup_{x^*} \langle x^*, x \rangle - F_i(x) = F^*(x^*) \]

\[ F^{**}(x) \geq \sup_{x^* \in X^*} \inf_{i \in I} F_i^*(x^*) = \sup_{x^* \in X^*} \langle x^*, x \rangle - F^*(x^*) = \sup_i F_i^*(x^*) = \sup_i F_i(x) = F(x) \]
**Cor.** $F : X \to IR^v$ proper.

$$F^{**} = \text{co } F$$

*Proof.* If $G \leq F$ (pointwise) and $G$ is convex & lsc, then

$$G = G^{**} \leq F^{**}.$$

On the other hand, $F^{**} \leq F$ and $F^* = \text{convex } \text{ & lsc}.$

$$F^{**} \leq \sup \text{ (pointwise) } G \leq \text{co } F \leq F^{**} \quad \square$$

**Prop.** $F : X \to IR^v \{\infty\}$ convex.

$F$ is Gâteaux differentiable in $x \in X$, $\nabla F(x) = x^*$

$$\implies \nabla F(x) = \{x^*\} \Rightarrow \text{F cont. in } x \in \text{int(dom F)}$$

"$\Rightarrow$" proven in Lecture 5. "Still need to prove the other direction.

"$\Leftarrow$" $\Phi_x : x \to IR^v \{\infty, 0\}$

$$\Phi_x(h) := \inf_{\varepsilon > 0} \frac{F(x + \varepsilon h) - F(x)}{\varepsilon}, \quad \varepsilon \to 0$$

non-decreasing in $\varepsilon$

by convexity

$\Phi_x$ is convex in $h$ by convexity of $F$.

$F$ cont. in $x$ $\Rightarrow$ Lipschitz on nblh($x$) (by the technical lemma):

$$\text{Lip} \leq \frac{F(x + \varepsilon h) - F(x)}{\varepsilon} \leq \text{Lip} h$$

$\Rightarrow \Phi_x(h) < \infty \forall h$ ("dom $\Phi_x" = X)

$\Phi_x$ bdd in a nblh($\bullet$)

$\Rightarrow \Phi_x$ cont. ( & convex) on int(dom $\Phi_x$) = $X$.

Now we know that

$$\Phi_x(h) = \Phi_x^{**}(h) = \sup_{y^*} \langle y^*, h \rangle = \Phi_x^{**}(h)$$

$$\Phi_x^{**}(h) = \sup_{h} \sup_{\varepsilon > 0} \langle y^*, h \rangle - \frac{F(x + \varepsilon h) - F(x)}{\varepsilon}$$

$$= \sup_{\varepsilon > 0} \frac{F(x + \varepsilon h) - F(x)}{\varepsilon} \geq \text{Fenchel-Moreau-Yosida}$$

$$= 0 \text{ if } y^* \in \nabla F(x)$$

$$> 0 \text{ otherwise}$$

$$= X^* \cdot \nabla F(x) = X^* \cdot \Phi_x^{**}(h)$$

$$= \sup_{y^* \in X^*} \langle y^*, h \rangle - \langle x^*, h \rangle \quad \square$$
Recall that a convex $F: X \to \mathbb{R}$:
- is Gateaux differentiable at $x$ if $\mathcal{D}F(x) = x^*$
- $\partial F(x) = \{ x^* \}$ and $F$ is cont. at $x$.
- is cont. in $x$ if $\partial F(x) \neq \emptyset$ and $x \in \text{int}(\text{dom } F)$
- is lsc $\Rightarrow$ cont. on $\text{int}(\text{dom } F)$

**Theorem:** Let $F: X \to \mathbb{R}$ be lsc and convex (so $\text{dom } F \neq \emptyset$ and $\partial F$ nowhere empty)

$F$ is strictly convex $\iff F^*$ is (everywhere) Gateaux differentiable.

(Rockafellar, Convex Analysis, 1970, Th. 26.3)

→ A bit more complicated if $\text{dom } F \neq X$. Has to do with the possibility that $\partial F$ may not be convex.

**Proof**

"$\Leftarrow$" Suppose $F$ is not strictly convex:

$\exists x_1 \neq x_2, \lambda \in (0,1)$ s.t. $F((1-\lambda)x_1 + \lambda x_2) < (1-\lambda)F(x_1) + \lambda F(x_2)$.

Take $x^* \in \partial F(x)$, hence:

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

In particular:

$F(x_1) \leq (1-\lambda)F(x_1) + \lambda F(x_2) + \langle x^*, x_1-x_2 \rangle$  \hspace{1cm} (**) 

and so

$F(x_1) \geq F(x_2) + \langle x_1, x_2-x_2 \rangle \geq F(x_1)$ \hspace{1cm} (***) 

$F(x_2) \geq F(x_1) + \langle x_2, x_2-x_1 \rangle \geq F(x_2)$ \hspace{1cm} (*****)

(This means that the supporting hyperplane of $\text{epi } F$ at $(x_1, F(x_1))$ passes through $(x_2, F(x_2))$.)

We can then rewrite, for any $y \in X$:

$$F(y) \geq F(x) + \langle x^*, y-x \rangle \stackrel{(\dagger)}{=} (1-\lambda)F(x) + \lambda F(x_2) + \langle x^*, y-(1-\lambda)x_1 - \lambda x_2 \rangle$$

$$\begin{cases} 
\stackrel{(\dagger\dagger)}{=} F(x_2) + \langle x^*, y-x_2 \rangle \\
\stackrel{(***)}{=} F(x_1) + \langle x^*, y-x_1 \rangle.
\end{cases}$$

It follows that $x^* \in \partial F(x_1)$ and $x^* \in \partial F(x_2)$.

But then $x_1, x_2 \in \partial F^*(x^*)$, and so $F^*$ can not be differentiable!
Suppose $F^*$ is not differentiable at some $x^* \in X^*$; then $x^* \in \partial F(x_1) \cap \partial F(x_2)$ for some $x_1, x_2 \in X$.

For arbitrary $x \in (0, 1)$:

$$
F(x) = F(x_1) + \langle x^*, y-x_1 \rangle \quad \times (1-x)
$$

$$
F(x) = F(x_2) + \langle x^*, y-x_2 \rangle \quad \times x
$$

$$
F(x) = (1-x)F(x_1) + xF(x_2)
$$

$(x = (1-x)x_1 + x x_2)$

Hence by convexity $F((1-x)x_1 + x x_2) = (1-x)F(x_1) + xF(x_2)$

and so $F$ is not strictly convex.

---

**Inf-convolutions and Moreau–Yosida regularisation**

**"(Usual) convolution"**

- $(F*G)(x) = \int G(x-z)G(z)dz = \int F(z)G(x-z)dz$

- $(F*\delta_0)(x) = F(x)$

**Typical smoothing kernel:**

- $\Theta_\varepsilon(x) = \frac{1}{V(\varepsilon(x))} e^{-\frac{1}{\varepsilon^2}|x|^2}$

- $\Theta_\varepsilon \rightharpoonup \delta_0$ (as measures, weakly-$*$ against $C_c^\infty(\mathbb{R}^d)$)

**Regularising/smoothing effect:**

- $(F*\Theta_\varepsilon) \in C_0^\infty(\mathbb{R}^d)$

**Approximation:**

- $F*\Theta_\varepsilon \rightarrow F$ (e.g. in $L^1$ if $F \in F'$)

**"Inf-convolution"**

- $(F \square G)(x) = \inf \int F(x-z) + G(z)$

- $(F \square \delta_0)(x) = F(x)$

**Typical smoothing kernel:**

- $\Theta_\varepsilon(x) = \frac{1}{2\varepsilon} ||x||^2$

- $\Theta_\varepsilon \rightharpoonup \delta_0$

**Regularising/smoothing effect:**

- $F \square \Theta_\varepsilon$ convex & differentiable

- $\varepsilon$-Lipschitz-cont., derivative

**Approximation:**

- $F \square \Theta_\varepsilon \rightarrow F$ (pointwise)
\[ F_\varepsilon (x) := \inf_{z \in \mathcal{X}} F(z) + \frac{1}{2\varepsilon} \|x - z\|^2 \]

Moreau-Yosida Regularization

In the following we shall take \( X = H \) Hilbert space.
(Some results are generalizable to Banach, or even metric space) (Peyroutenet prop 3.35)

Prop. \( F : H \to \mathbb{R} \cup \{\infty\} \) proper, cvx, lsc. For any \( x \in H \), \( \varepsilon > 0 \), \((\ast)\) has a unique minimiser \( J_\varepsilon (x) \). Moreover,

\[ J_\varepsilon (x) - x \in \partial F(J_\varepsilon (x)) \]

Proof. \( F \) is proper, cvx, lsc \( \Rightarrow \) \( F(x) = \sup_{\alpha \in \mathcal{F}} F_\alpha (x) \) for some family of cont. affine functions. Hence for the level sets of \( F + \frac{1}{2\varepsilon} \| \cdot \|^2 \),

\[ -\|x\|^2 + \|x\|^2 \leq \sup_{\alpha \in \mathcal{F}} F_\alpha (x) + \frac{1}{2\varepsilon} \|x\|^2 = F(x) + \frac{1}{2\varepsilon} \|x\|^2 \leq C (\ast \ast) \]

(if \( F_\alpha (x) = \langle x, \alpha \rangle + \alpha ) \).

Bdd level sets \( \Rightarrow \) weak-\( \ast \) compact level sets.

(Read that \( H \) has a predual, namely \( H_{**} \)).

Norms are always weak-\( \ast \) lsc
\( F \) lsc \& convex \( \Rightarrow \) \( F \) weak-\( \ast \) seg. lsc \( \Leftrightarrow \) weak-\( \ast \) seg. lsc (Hilbert space).

\( F + \frac{1}{2\varepsilon} \| \cdot \|^2 \) has weak-\( \ast \) compact level sets, is weak-\( \ast \) seg. lsc, and bounded from below by \((\ast \ast)\), and so the minimiser exists.

By strict convexity, the minimiser must be unique.
The optimality equation follows from Fermat’s rule \( \Box \)

Remarks

D. Since there exists a unique solution we can write

\[ J_\varepsilon (x) = (I + \varepsilon \partial F)^{-1}(x) \]

This is similar to the ‘resolvent’ of an operator \( Q = \partial F \) which is studied in the Hille-Yoseida Theorem to prove existence of a semigroup s.t. \( P_t = Q P_t \).

ii) Clearly

\[ \inf F \leq F_\varepsilon (x) \leq F(x), \]

And so:

- \( \inf F = \inf F_\varepsilon \)
- \( \inf F = F(x) \iff \inf F = F_\varepsilon (x) \forall \varepsilon > 0. \)
Lemma \( F: H \to (\mathbb{R} \cup \{0\}) \) proper, lsc & cvx, then \( J_\varepsilon: H \to H \) is a contraction, i.e. \( 1 \)-Lipschitz continuous.

**Proof** For any \( x, y \in H \)
\[
- \frac{J_\varepsilon(x) - x}{\varepsilon} \in \partial F(J_\varepsilon(x)) \quad \text{and} \quad - \frac{J_\varepsilon(y) - y}{\varepsilon} \in \partial F(J_\varepsilon(y)).
\]

By monotonicity of \( \partial F \):
\[
\left< - \frac{J_\varepsilon(x) - x}{\varepsilon} + \frac{J_\varepsilon(y) - y}{\varepsilon}, x - y \right> \geq 0.
\]

Hence
\[
0 \leq \|J_\varepsilon(x) - J_\varepsilon(y)\|^2 \leq \left< J_\varepsilon(x) - J_\varepsilon(y), x - y \right> \leq \|J_\varepsilon(x) - J_\varepsilon(y)\| \|x - y\|
\]
and so
\[
\|J_\varepsilon(x) - J_\varepsilon(y)\| \leq \|x - y\|.
\]

**Prop** \( F: H \to (\mathbb{R} \cup \{0\}) \) proper, lsc & cvx, \( \varepsilon > 0 \), then \( F_\varepsilon \) is Gateaux (even Fréchet) differentiable, and convex, and
\[
DF_\varepsilon(x) = \frac{1}{\varepsilon} (x - J_\varepsilon(x)),
\]
(and \( DF_\varepsilon \) is Lipschitz continuous with constant \( \frac{1}{\varepsilon} \)).

**Proof** For any \( x, y \in H \)
\[
F_\varepsilon(y) - F_\varepsilon(x) = F_\varepsilon(J_\varepsilon(y)) - F_\varepsilon(J_\varepsilon(x)) + \frac{1}{2\varepsilon} \left( \|J_\varepsilon(y) - y\|^2 - \|J_\varepsilon(x) - x\|^2 \right)
\]
\[
\geq \left< - \frac{J_\varepsilon(y) - y}{\varepsilon}, J_\varepsilon(y) - J_\varepsilon(x) \right> + \frac{1}{2\varepsilon} \left( \|J_\varepsilon(y) - y\|^2 \right)
\]
\[
+ 2 \left< J_\varepsilon(x) - y, J_\varepsilon(y) - J_\varepsilon(x) \right>.
\]

Similarly
\[
F_\varepsilon(x) - F_\varepsilon(y) \geq \frac{1}{\varepsilon} \left< y - J_\varepsilon(y), x - y \right>.
\]

Setting \( y = x + \varepsilon h \) for \( \varepsilon > 0 \) and an arbitrary direction \( h \in H \):
\[
\frac{1}{\varepsilon} \left< x - J_\varepsilon(x), \varepsilon h \right> \leq \frac{F_\varepsilon(x + \varepsilon h) - F_\varepsilon(x)}{\varepsilon} \leq \frac{1}{\varepsilon} \left< x + \varepsilon h - J_\varepsilon(x + \varepsilon h), \varepsilon h \right>.
\]

Hence, by the Lipschitz continuity of \( J_\varepsilon \):
\[
\lim_{\varepsilon \to 0} \frac{F_\varepsilon(x + \varepsilon h) - F_\varepsilon(x)}{\varepsilon} = \frac{1}{\varepsilon} \left< x - J_\varepsilon(x), h \right> \Rightarrow DF_\varepsilon(x) = \frac{1}{\varepsilon} (x - J_\varepsilon(x)).
Note that \( DF_\epsilon(x) \in \partial F(\epsilon x) \). Therefore
\[
\langle DF_\epsilon(y) - DF_\epsilon(x), y - x \rangle = \langle y - J_\epsilon(y) - x + J_\epsilon(x), y - J_\epsilon(y) - x + J_\epsilon(x) \rangle 
+ \langle y - J_\epsilon(y) - x + J_\epsilon(x), J_\epsilon(y) - J_\epsilon(x) \rangle \geq 0,
\]
and so \( F_\epsilon \) is convex.

\( \Box \)

\[ \text{(Prop. lsc & cvx)} \]

**Prop** \( \lim_{\epsilon \to 0} F_\epsilon(x) = F(x) \quad \forall x \in H \) \quad (Proposition 3.41)

**Proof** (For fixed \( x \in \text{dom} F \))
\[
F_\epsilon(x) = F(J_\epsilon(x)) + \frac{1}{\epsilon} \| J_\epsilon(x) - x \|^2 \leq F(x)
\]
\[
F_i(J_\epsilon(x)) + \frac{1}{\epsilon} \| J_\epsilon(x) - x \|^2 \quad \text{for some cont. affine } F_i : H \to \mathbb{R}.
\]

Hence \( J_\epsilon(x) \) is \( \epsilon \)-bounded as \( \epsilon \to 0 \).

Therefore \( F_i(J_\epsilon(x)) \) is bounded from below, and so is \( F(J_\epsilon(x)) \). Then:
\[
\frac{F(J_\epsilon(x))}{\epsilon} \quad \text{is bounded below,}
\]

and so \( \| J_\epsilon(x) - x \| \to 0. \)

By (norm-sequential) lower semicontinuity:
\[
F(x) \leq \liminf_{\epsilon \to 0} F(J_\epsilon(x)) \leq \limsup_{\epsilon \to 0} F(J_\epsilon(x)) \leq F(x).
\]
If $F: X \to \mathbb{R} \cup \{\infty\}$ is convex then also (for any finite convex combination)

$$F(\sum_{i=1}^{n} \lambda_i x_i) \leq \sum_{i=1}^{n} \lambda_i F(x_i) \quad \forall (x_i)_{i=1}^{n} \in X, \lambda_i \in P([0,1]), \sum_{i=1}^{n} \lambda_i = 1.$$ 

How far can we push this? For example, if we use a "Schauder basis" and assume $F$ is lsc, then

$$\sum_{i=1}^{n} \lambda_i x_i \xrightarrow{\mathbb{P}} x$$

$$\Rightarrow$$

$$F(x) \leq \liminf_{n \to \infty} F(\sum_{i=1}^{n} \lambda_i x_i) \leq \liminf_{n \to \infty} \sum_{i=1}^{n} \lambda_i F(x_i).$$

This again shows that lsc will be important.

How about integrals against a probability measure? We will now restrict to real-valued (one-dimensional) convex functions. These are continuous on the interior of their (convex) domain. Lsc is needed only to control behaviour at the boundary:

---

**Theorem (Jensen's inequality).** Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $F: \Omega \to \mathbb{R} \cup \{\infty\}$ be convex and lsc, and $X \in L^1_{\mathbb{P}}(\Omega)$. Then:

i) $F(\int X \, d\mathbb{P}) \leq \int F \circ X \, d\mathbb{P}$

ii) Moreover, if $F$ is strictly convex, then

$$F(\int X \, d\mathbb{P}) = \int F \circ X \, d\mathbb{P} \iff \mathbb{P}(\{\omega \in \Omega : X(\omega) = q\}) = 1$$

(for some $q \in \mathbb{R}$)
> Can be found anywhere in the literature (not in Peyronnet), e.g. 
  Bogachev - Measure Theory (2006) Th. I.2.12.1g 
  Lieb & Loss - Analysis (2001) Th. 2.2 
  Evans - PDE's (1997) Th. B.1.2 

However, not always very precise about lsc, or point (ii), ... 

> A generalisation to Banach spaces would require 
  "Bochner-integrals", i.e. Banach-valued integrals.

> It is essential that \( \sigma(\Omega) = 1 \). One often needs to rescale, 
  for example \( \int_{\Omega} \frac{dx}{1+|u|} \) is impossible if \( \sigma(\Omega) = \infty \), for example \( \int_{\mathbb{R}^d} \).

**Proof** (Fox is measurable)

(i) \[
\begin{align*}
F(\{x\in\Omega\}) &= \sup_{i\in I} F_i(\{x\in\Omega\}) \\
&= \sup_{i\in I} \int_{\Omega} F_{i0} x \, dx \\
&\leq \sum_{i\in I} \int_{\Omega} F_{i0} x \, dx
\end{align*}
\]

for some family of cont. affine functions, since 
\( F \) is convex & lsc. 

since integrals are linear \( \sigma(\Omega) = 1 \)

(ii) Assume \( \sigma \) is not "deterministic", i.e. \( \sigma(\{x\in\Omega: x(\omega) = x\}) > 1 \) \( \forall \omega \in \Omega \).

Wlog. we may consider \( \rho \in P(\Omega), \rho(A) = \sigma(x'(A)) \).

The assumption means that \( \rho \neq \delta \) \( \forall \omega \in \Omega \).

Hence there are two disjoint sets \( A_1 \cup A_2 = \Omega \) such that 
\( \rho(A_1) > 0 \) and \( \rho(A_2) > 0 \). Then \( \Omega_1 = x'(A_1), \Omega_2 = x'(A_2) \) are also disjoint and \( \Omega_1 \cap \Omega_2 = \emptyset \), now:

\[
\begin{align*}
F(\{x\in\Omega\}) &= F(\sigma(\Omega_1) \int_{\Omega_1} x \, dx + \sigma(\Omega_2) \int_{\Omega_2} x \, dx) \\
&< F(\sigma(\Omega_1) \int_{\Omega_1} x \, dx) + \sigma(\Omega_2) F(\int_{\Omega_2} x \, dx) \\
&\leq \sigma(\Omega_1) F(x) \, dx + \sigma(\Omega_2) F(x) \, dx
\end{align*}
\]

**Point (iii)** is very important! It shows that Jensei's ineq. 

*can be very powerful when considering a sequence of 
probability measures that converge to a delta measure!
Jensen and coercivity

**Cor.** If \( F(x) = \int f(1x) \mu(dx) \), \( f: \mathbb{R} \to [0, \infty) \)
convex and \((\Omega, A, \mu)\) is a probability space, then \( F \) has \( L^p(\Omega) \)-uniformly bounded level sets.

Of course, a uniform \( L^p \)-bound is usually not very helpful since \( L^p \) doesn’t have a predual, and so we can not deduce weak-* compactness. However, one can often do better:

**Cor.** Let \((\Omega, A, \mu)\) be a probability space and \( p \geq 1 \) and \( x \in L^p(\Omega) \). Then for any \( r \in (0, p] \):

\[
(\int |x| \mu(dx))^r \leq \left( \int |1x| \mu(dx) \right)^r
\]

(Bogachev–Measure Theory 2007,
Cor I.2.12.21)

(Bogachev Th. I.2.12.24)

**Th.** (Pinsker–Kullback–Csiszár). Let \( \mu, \nu \) be two probability measures on a measurable space \((\Omega, A)\), and assume \( \|\mu\| = 1 \).

\[
\|\mu - \nu\|_{TV}^2 = \left( \int \frac{d\mu}{\nu} - 1 \right) \mu(dx) \leq 2 \int (\log \frac{d\mu}{\nu}) d\mu
\]

\[
= \int \lambda_B \left( \frac{d\mu}{\nu} \right) d\mu \quad \text{(relative entropy)}
\]

→ In fact, we will see that \( \int f(1x) \mu(dx) \) is related to the norm of so-called Orlicz–spaces. When estimating the \( L^p \)-norm we throw away a lot of information that would otherwise be useful for analysis in an Orlicz space!
Jensen and convergence: an example

Let \((\Omega, \mathcal{A}, \sigma) = (\mathbb{R}^d, \text{Borel}, L^1(\sigma))\).

\[
F(x) := \int \mathbb{A}_E \left( \frac{dx}{dx} \right) d\mathbb{x}^+= \int (x(q) \log x(q) - x(q) + 1) d\mathbb{q}
\]

if not \(x \ll L^1(\sigma, \mathcal{A})\), so we may assume \(x \in L^1(\sigma, \mathcal{A})\).

Smoothing: \(\theta_{E}(q) = \frac{1}{V(\sigma)E} \frac{1}{\exp \left( \frac{x(q)}{E} \right)}\), \(x_{E} := x + \theta_{E} \in C_{c}^{\infty}(\mathbb{R}^d)\).

Can we prove that \(F(x_{E}) \rightarrow F(x)\)?
Very difficult to find a majorant for dominated convergence... However we can exploit convexity!

**Lemma** If \(x \in L^p \Rightarrow \theta x + \theta_{E} \overset{L^p}{\rightarrow} x\), \(1 \leq p < \infty\)

**Prop** If \(x \in L^1(\sigma, \mathcal{A})\) then \(F(x_{E}) \rightarrow F(x)\)

**Proof** On the one hand,

\[
F(x_{E}) = \int \mathbb{A}_E \left( \frac{dx}{dx} \right) d\mathbb{x}^+ \leq \int \mathbb{A}_E \left( \frac{dx}{dx} \right) \theta_{E} d\mathbb{x}^+ \rightarrow \int \mathbb{A}_E (x(q)) d\mathbb{q} = F(x), \text{ since } \mathbb{A} \text{ is } L^1(\sigma, \mathcal{A})
\]

On the other hand,

\[
F(x_{E}) = \int \sup_{x \in L^1(\sigma, \mathcal{A})} (x(q) - \frac{e^{x(q)}}{x(q)} - 1) d\mathbb{q}
\]

Hence \(F\) is \(L^1\)-lsc (topologically sequentially) and so

\[
F(x) \leq \lim_{E \to 0} F(x_{E}) \leq \limsup_{E \to 0} F(x_{E}) \leq F(x) \quad \square
\]
\[ F(x) = S(x) = \int \lambda_b \left( \frac{dx}{dx_t} \bigg|_{t=0} \right) (g) \mathcal{L}_0 \mathcal{L}_1 (dg), \quad x \in \mathbb{R}^d, \quad \|x\|_1 \leq \lambda, \quad \text{other-wise.} \]

"Relative entropy" of measure \( x \) w.r.t. measure \( \mathcal{L}_0 \mathcal{L}_1 \).

Last week we proved by Jensen that

\[ F(x + \theta_t) \leq F(x), \quad \text{for } x \in L'(\mathbb{R}^d). \]

There's another reason why this is true...

Instead of the heat kernel, let \( \theta_t \) be the solution of

\[
\begin{cases}
\dot{\theta}_t = \Delta \theta_t, & \text{on } \Omega \times (0, \infty), \ t > 0, \\
\frac{\partial \theta}{\partial n} = 0, & \text{on } \partial \Omega \times (0, \infty), \ t > 0, \quad (\text{Neumann BC}) \\
\theta = \delta_0, & \text{t = 0.}
\end{cases}
\]

Then for \( x_t := x + \theta_t \) also:

\[
\begin{cases}
\dot{x}_t = \Delta x_t, & \text{on } \Omega \times (0, \infty), \ t > 0, \\
\frac{\partial x_t}{\partial n} = 0, & \text{on } \partial \Omega \times (0, \infty), \ t > 0, \\
x_0 = x, & \text{t = 0.}
\end{cases}
\]

At least formally,

\[ \langle DF(x), h \rangle = \lim_{h \to 0} \frac{F(x + h) - F(x)}{h} = \int \log x \log_+ x. \]

Therefore the evolution of \( x_t \) can also be written as

\[ \dot{x}_t = \text{div}(x_t \nabla \log x_t) = \text{div}(x_t \nabla \log x_t) = \text{div}(x_t \nabla x_t) = \Delta x_t. \]

Hence \( x_t \) is the gradient flow of \( F \) on some strange manifold.

Therefore (this is always true for a gradient flow):

\[ \frac{\partial}{\partial t} F(x_t) = \langle DF(x_t), \dot{x}_t \rangle = \langle DF(x_t), \text{div}(x_t \nabla DF(x_t)) \rangle \]

\[ = - \langle \nabla F(x_t), x_t \nabla DF(x_t) \rangle + \int_{\partial \Omega} \frac{\partial}{\partial n} DF(x_t) x_t \nabla DF(x_t) \cdot n \]

\[ = - \int_{\Omega} \nabla DF(x_t)^2 (x_t) \leq 0 \]

\[ = \int_{\Omega} |\nabla x_t|^2 (x_t) \leq 0 \quad \text{"Fisher information"} \]
Lower semicontinuity of Lagrangian actions

Previous lecture: strategy how to prove lsc of \( F(x) = \int f(x(t)) dt \), for some convex & lsc \( f \).

Now: how to prove lsc of \( F(x) = \int L(x(t), x(t), \dot{x}(t)) dt \)?

(Evans - PDE's Th 8.2.1)

**Th:** \( \Omega \subset \mathbb{R}^n \) open bounded with smooth boundary, \( 1 < p < \infty \).
Assume that \( L: \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R} \) is bdd from below and convex in the first argument. Then \( F \) is sequentially weak lsc in \( W^{1,p}(\Omega) \).

\[ \rightarrow \| x \|_{W^{1,p}} = \| x \|_p + \| \nabla x \|_p, \]
\( W^{1,p} \) reflexive, so weak = weak*.

We first need a number of results:

(Brezis Th II.1)

**Th:** (Banach-Steinhaus / Uniform Boundedness Principle)
Any weakly (or weakly *) convergent sequence \( x_k \rightarrow x \)
in a Banach space \( X \) is bounded:
\[ \sup_k \| x_k \| < \infty \rightarrow x_k \text{ rel. cpt. in } L^p. \]

(Brezis Th IV.9)

**Th:** (Rellich-Kondrakov)
For all \( 1 \leq p \leq \infty \),
\( W^{1,p}(\Omega) \hookrightarrow C^0(\Omega) \) i.e. i) \( \| x \|_p \leq C \| x \|_{W^{1,p}} \) (dual)
"compact embedding"

ii) \( \sup_k \| x_k \|_{W^{1,p}} < \infty \rightarrow x_k \text{ rel. cpt. in } L^p \)

(Brezis Th IV.9)

**Th:** (Converse dominated convergence) If \( x_k \overset{L^p}{\rightarrow} x \), \( 1 \leq p \leq \infty \), then \( \exists \) subsequence \( x_{k_n} \) such that i) \( x_{k_n} \overset{a.e.}{\rightarrow} x(t) \)
ii) \( \exists h \in L^p \) with \( |x_{k_n}(t)| \leq h(t) \) a.e.

(Evans Th E.2.2)

**Th:** (Egorov) \( \Omega \subset \mathbb{R}^n \) measurable and \( f_k, f \in M(\mathbb{R}) \) with \( f_k(x) \overset{a.e.}{\rightarrow} f(x) \) for almost every \( x \in \Omega \). Then for each \( \epsilon > 0 \)
there exist a measurable \( E \subset \Omega \) such that:

i) \( m(E) \leq \epsilon \) (in Lebesgue measure),

ii) \( f_k \overset{\text{uniformly}}{\rightarrow} f \) uniformly on \( E \).
Proof (that $F$ is strong weakly lsc)

1. Take a sequence $x_k \overset{w^*}{\to} x$. By uniform bounded principle,
   \[ \sup_k \|x_k\|_{L^p} < \infty. \]

   Rellich-Kondrachov: $\exists L^p$-convergent subsequence.

   Converse dominated convergence: $\exists$ a.e.-convergent subsequence.

   Taking a further sub-subsequence, we may assume that
   \[ \begin{align*}
   &i) x_{k_\ell}(q) \to x(q) \text{ for a.e. } q \in \Omega, \\
   &ii) \liminf_F(x_{k_\ell}) = \liminf_F(x_k)
   \end{align*} \]

2. Egorov: $\exists E_\varepsilon$ such that $x_{k_\ell} \to x$ uniformly on $E_\varepsilon$ and
   \[ \lim_{\varepsilon \to 0} |E_{2\varepsilon}| = 0. \]

   Let $H_\varepsilon := \{ q \in \Omega : |x(q) + 10\varepsilon X(q)| \leq \frac{1}{\varepsilon}, \text{ so } |x(q) + 10\varepsilon X(q)| \to 0 \}.

   Let $G_\varepsilon := E_\varepsilon \cap H_\varepsilon$, and so
   \[ \lim_{\varepsilon \to 0} |G_\varepsilon| = \lim_{\varepsilon \to 0} |E_{2\varepsilon}| + \lim_{\varepsilon \to 0} |H_\varepsilon| = 0. \]

3. Assume w.l.o.g. that $L \geq 0$ (recall $L$ is bounded from below). Then
   \[ F(x_{k_\ell}) = \int_{G_\varepsilon} L(\nabla x(q), x_{k_\ell}(q), q) \, dq \]
   \[ \geq \int_{G_\varepsilon} L(\nabla x(q), x(q), q) \, dq. \]

   Since $L$ is smooth, $L(\nabla x(q), x(q))$ is uniformly continuous on $E_\varepsilon$, and thus
   \[ \lim_{\varepsilon \to 0} \int_{G_\varepsilon} L(\nabla x(q), x(q), q) \, dq = 0. \]

4. Similarly, $\nabla x \to \nabla x$ uniformly in $G_\varepsilon$, and so this convergence is in $L^1$, and since $G_\varepsilon \subset \Omega$ is bounded,
   \[ \lim_{\varepsilon \to 0} \int_{G_\varepsilon} L(\nabla x(q), x(q), q) \, dq = 0. \]

5. This implies $\lim_{\varepsilon \to 0} F(x_{k_\ell}) = F(x)$. This implies
   \[ \liminf_{k \to \infty} F(x_k) = \liminf_{\varepsilon \to 0} F(x_{k_\ell}) \geq \int_{G_\varepsilon} L(\nabla x(q), x(q), q) \, dq. \]

6. Monotone convergence as $\varepsilon \to 0$:
   \[ \liminf_{k \to \infty} F(x_k) \geq \int_{\Omega} L(\nabla x(q), x(q), q) \, dq = F(x). \]

   Interestingly, this result also holds in the other direction...!
Generalised convexity notions

Functionals on matrices

If \( x : \Omega \to \mathbb{R}^m \), \( \Omega \subset \mathbb{R}^n \) then \( \forall x \in \mathbb{R}^{mxn} \).

\[
F(x) = \int \left( \sum_{i,j} (x_i(x_j), x_j, q) \right) dq
\]

**Def:** \( f : \mathbb{R}^{mxn} \to \mathbb{R}(\omega) \) is **polyconvex** iff

\[
f(A) = \varphi(A, \text{minors}_A) \quad \text{for some convex } \varphi.
\]

(recall a minor is the determinant of the matrix where row and columns have been removed)

\( f : \mathbb{R}^{mxn} \to \mathbb{R} \) (locally bounded & measurable) is **quasiconvex** iff

\[
f(A) \leq \frac{1}{|\omega|} \int f(A + \theta q) dq \quad \text{for every } \omega \in \mathbb{R}^n \text{ held fixed } \quad \text{and } \varphi \in C^0(\omega; \mathbb{R}^n).
\]

**Formal Theorem**

\[
F(x) = \int L(\nabla x, x, q) dq
\]

\( L \) is quasiconvex in its first argument \( \Rightarrow \)

\( F \) is sequentially weakly lsc in \( W^{1,p}(\Omega) \)

(Dacorogna - Direct Methods in the Calculus of Variations, 2nd ed., chapter 5 & 6)

\[\Rightarrow \] In practice, quasiconvexity is hard to check. Therefore one often replaces it by different notions (Dacorogna Th. 1.7)

**Th**: \( f : \mathbb{R}^{mxn} \to \mathbb{R} \)

\( f \) convex \( \Rightarrow \) \( f \) polyconvex \( \Rightarrow \) \( f \) quasiconvex \( \Rightarrow \) \( f \) "rank one convex"

\[\Rightarrow \] The exact terminology "quasi-convexity" changes in the literature...!
$\lambda$-convexity (aka. strong convexity or semi-convexity)

Recall: Let $A = \text{dom} F \subset X$ be open and convex and $F: A \to \mathbb{R}$ be twice Gateaux differentiable.

1) $F$ is $\lambda$-convex if
   \[ F(y) \geq F(x) + \langle DF(y), y-x \rangle \quad \forall x, y \in A \quad \text{(gradient inc.)} \]
2) $\langle DF(x) - DF(y), x-y \rangle \geq \lambda \|x-y\|^2 \quad \forall x, y \in A \quad \text{(monotonicity)}$
3) $\nabla^2 F(x) \succeq 0$ (positive semidefinite)

What if we replace $\lambda$ by a different constant?

Definition: $F: X \to \mathbb{R}$ is $\lambda$-convex if
\[ F((1-\lambda)x + \lambda y) \leq (1-\lambda)F(x) + \lambda F(y) - \frac{\lambda}{2} (1-\lambda)\lambda \|x-y\|^2 \]

Remarks:
- $\lambda = 0$ \iff convex
- $\lambda > 0$ \iff convex (stronger)
- $\lambda < 0$ \iff convex (weaker)
- $F$ is $\lambda$-convex if \iff $F - \frac{\lambda}{2} \|x\|^2$ is convex.

Let $A = \text{dom} F \subset X$ be open and convex and $F: A \to \mathbb{R}$ be twice differentiable, $\lambda \in \mathbb{R}$.

1) $F$ is $\lambda$-convex if
   \[ F(y) \geq F(x) + \langle DF(x), y-x \rangle + \frac{1}{2} \lambda \|y-x\|^2 \quad \text{(gradient inc.)} \]
2) $\langle DF(x) - DF(y), x-y \rangle \geq \lambda \|y-x\|^2 \quad \text{(monotonicity)}$
3) $\nabla^2 F(x) \succeq \lambda I$ (positive definiteness)

Convex (and $\lambda$-convex) functions have supporting hyperplanes. Similarly, $\lambda$-convex functions have supporting "hyperplanes."

- $\lambda < 0$: can still do a lot of convex analysis even though $F$ not convex!
- $\lambda > 0$: more control & better estimates.
Uniform integrability (Introduction to Orlicz spaces)

In what follows, $(\Omega, \mathcal{A}, \mu)$ will always be a prob. mea. space.
Recall that $L^p$ for $p \geq 1$ has a predual;
- A uniform $L^p$-bound yields weak-$\star$ compactness; (Banach-Alaoglu)
- $L^1$ doesn't have a predual, so a uniform $L^1$-bound (e.g. from Jensen) doesn't work.

However:
(Brezis Th. 4.29/4.30)

Uniform integrability

A sequence $(x_n)_n \in L^1(\mu)$ is relatively weakly compact
$\iff$

it is "uniformly integrable"

Def A sequence $(x_n)_n \in L^1(\mu)$ is uniformly integrable iff
$\forall \epsilon > 0 \exists S > 0 \forall A \in \mathcal{A} \forall x \in A$
$|\mu(A) - S| \Rightarrow \sup_n \|x_n\|_\mu < \epsilon$

This is nice, but how to prove uniform integrability, e.g. of
level sets $F \subseteq C$ of some functional $F(x) = \int f(x(\omega)) \mu(\omega)$?
Can we exploit the information that we throw away by Jensen?

$f(\int x(\omega) \mu(\omega)) \leq \int f(x(\omega)) \mu(\omega) \leq C$. (f convex)

(Bogachev, Th I.4.5.9)

Th (De la Vallée-Poussin)

A sequence $(x_n)_n \in L^1(\mu)$ is uniformly integrable iff
there exists a non-negative increasing convex $\Phi: [0,\infty] \to [0, \infty)$
for which
$\lim_{t \to \infty} \frac{\Phi(t)}{t} = \infty$ (superlinear growth)
$\sup_n \int \Phi(\|x_n\|_\mu) \mu(\omega) < \infty$

Hence $\{\|x\| \leq C\}$ is (seg.) weakly compact in $L^1$

Another application of uniform integrability is Vitali's
convergence theorem:
$x_n(\omega) \to x(\omega) \mu$ a.e.
$x_n \text{ unif. int } \Rightarrow x_n \overset{\text{up}}{\Rightarrow} x$

Instead of working with $L^1$ weakly, maybe we can work
directly with the "Orlicz class":
$L^{\Phi}(\Omega) = \{x: \Omega \to \mathbb{R} \text{ measurable} : \int \Phi(\|x(\omega)\|_\mu) \mu(\omega) < \infty\}$.
Relation with $L^1$

\[ L^1_\mu(\Omega) = U \{ L^{\psi}_\mu(\Omega) : \psi \text{ non-negative, increasing, convex,} \]
\[ \lim_{t \to \infty} \frac{\psi(t)}{t} = \infty \} \]

**Proof:**

\[ L^{\psi}_\mu \subseteq L^1_\mu \quad \forall \psi \text{ by Jensen.} \]

The one-element set \( \{ x \} \subset L^1_\mu \) is clearly "uniformly integrable," so by de la Vallée-Poussin \( x \in L^\psi_\mu \) for some \( \psi \) satisfying the conditions.

→ Compare this to:

\[ L^{1, \mu}_\mu(\Omega) \supseteq U_{\mu \geq 1} L^\mu_\mu(\Omega). \]

→ Can we put more structure on \( L^{\psi}_\mu \)? Challenges:
- \( \int \psi(|x|) \, d\mu \) can not be scaled to be a norm.
- \( L^\psi_\mu \) is not a vector space! (In general)

(D) Setting: (Young and) N-functions.

(we shall assume \( \psi \) is defined on the whole \( \mathbb{R} \), but even, i.e. \( \psi(-z) = \psi(z) \))

**Def** \( \psi : \mathbb{R} \to \mathbb{R}^+ \) is an N-function iff it is:

- continuous
- convex
- \( \psi(0) = 0 \iff \psi(z) \geq 0 \)
- \( \psi(-z) = \psi(z) \) (even)
- \( \lim_{z \to \infty} \frac{\psi(z)}{z} = \infty \) (superlinear growth)
- \( \lim_{z \to 0} \frac{\psi(z)}{z} = 0 \) (differentiable in \( 0 \))

**Prop** If \( \psi \) is an N-function then so is \( \psi^* \)

**proof (exercise).**

→ Many results require less assumptions, e.g. \( \psi : \mathbb{R} \to [0, \infty] \), or \( \mu \) may even be an infinite measure: \( \mu(\Omega) = \infty \).
For consistency we focus on N-functions \( \psi \) and prob. meas. \( \mu \).
Examples:

\[ \varphi(z) = |z| \]

(P, P > 1)

\[ \varphi(z) = \lambda z (z+1) \]

almost linear growth

\[ \varphi(z) = \lambda \log(z+1) - z \]

exponential growth

\[ \int \varphi(|x|) d\nu = \|x\|_{L_p}^{p} \]

\[ \|x\|_{L_p}^{p} \]

... = ?

... = ?
Lecture 12

A. Linearizing the Orlicz class

Theorem. Let $\varphi : \mathbb{R} \to \mathbb{R}^+$ be an N-function and $(\Omega, \mathcal{A}, \mu)$ a non-atomic probability space. Then:

(i) $L^\varphi(\Omega)$ is a vector space iff:

\[ \exists K, \varrho > 0 \quad \forall z \geq 0 \quad \varphi(2z) \leq K \varphi(z) \quad \text{"$\Delta_2$-property"} \]

(or $\varphi \in \Delta_2$)

(ii) In general,

\begin{itemize}
  \item $x, y \in L^\varphi(\Omega)$, $|x| + |y| \leq 1 \Rightarrow \alpha x + \beta y \in L^\varphi(\Omega)$,
  \item $y \in L^{\varphi'}(\Omega)$ and $x$ measurable with $|x| \leq |y| \Rightarrow x \in L^\varphi(\Omega)$.
\end{itemize}

"Circled solid subset"

(implicitly identifying a.e. equal functions)

Definition:

\[ L^\varphi(\Omega) := \{ x : \Omega \to \mathbb{R} \text{ measurable : } \exists \alpha > 0 \text{ s.t. } \int \varphi(\alpha x) \, d\mu < \infty \} \]

"Orlicz space"

Proposition. $L^\varphi(\Omega)$ is a vector space. (Rao & Ren Prop. 3.1.6)

Proof. Take $x, y \in L^\varphi(\Omega)$, hence $\exists \alpha_x, \alpha_y > 0$ s.t. $\alpha_x x, \alpha_y y \in L^{\varphi'}(\Omega)$.

Take arbitrary $a, b \in \mathbb{R}$, and set $c = \frac{\alpha_x}{\alpha_x + \beta_y} a + \frac{\alpha_y}{\alpha_x + \alpha_y} b$.

Then $c(ax + by) = \frac{\alpha_x}{\alpha_x + \beta_y} ax + \frac{\alpha_y}{\alpha_x + \alpha_y} by \in L^\varphi(\Omega)$.

$\square$
Norming the Orlicz space \( I \)

**Prop.** \( \forall x \in L^\varphi(\mathbb{R}) \exists \beta \text{ s.t. } \int \varphi(\beta x) \, dy \leq 1 \)

**Proof:** Take an (arbitrary) sequence \( x_n \to 0 \) and set \( x_n := x \cdot a_n \), where \( a \) is such that \( ax \in L^\varphi(\mathbb{R}) \).

Then \( 0 \leq \varphi(ax_n) \leq \varphi(ax) \in L^\varphi(\mathbb{R}) \) and \( \varphi(ax_n) \to 0 \) for \( (a, x) = 0 \).

Dominated convergence: \( \int \varphi(ax_n) \, dy \to 0 \). Hence there exists an \( n_0 \) (set \( \beta := x_{n_0} \)) such that \( \int \varphi(\beta x) \leq 1 \).

\[ N_\varphi(x) := \inf \{ k > 0 : \int \varphi(\frac{x}{k}) \, dy \leq 1 \} \]

\( \hat{\varphi} \) Luxemburg norm

**Prop.** \( N_\varphi \) is a norm

**Sketch of Proof:**
1. \( N_\varphi(ax) = \inf \{ k > 0 : \int \varphi(\frac{ax}{k}) \, dy \leq 1 \} \)
2. \( = \inf \{ axk > 0 : \int \varphi(\frac{x}{k}) \, dy \leq 1 \} \)
3. \( = a \cdot N_\varphi(x) \).

(Triangle inequality follows from convexity of \( \varphi \))

**Remark:** More generally, norms of the type \( x \mapsto \inf \{ k > 0 : \frac{x}{k} \in B \} \) for some closed solid set \( B \) are called gauge or Minkowski norms.

\( \textbf{Th.} (L^\varphi(\mathbb{R}), N_\varphi) \text{ is a Banach space (i.e. complete)} \)

(when p-a.e. equivalent functions are identified)

**Th.** \( N_\varphi(x) <, =, > 1 \iff \int \varphi(\beta x) <, =, > 1 \)

Very useful unit ball property:

(\( \text{Rao & Ren Th. 3.2.3} \))
Let $x \in L^p_\varphi$ and $y \in L^{p^*}_\varphi$. By Young's inequality

$$\frac{|x| \cdot |y|}{N_\varphi(x) N_{p^*}(y)} \leq \varphi\left(\frac{|x|}{N_\varphi(x)}\right) + \varphi^\ast\left(\frac{|y|}{N_{p^*}(y)}\right).$$

Integrating:

$$\frac{1}{N_\varphi(x) N_{p^*}(y)} \int |xy| \, d\mu \leq \int \varphi\left(\frac{|x|}{N_\varphi(x)}\right) \, d\mu + \int \varphi^\ast\left(\frac{|y|}{N_{p^*}(y)}\right) \, d\mu \leq 1 + 1 = 2,$$

and so we obtain a Hölder-type estimate:

$$\int |xy| \, d\mu \leq 2 N_\varphi(x) N_{p^*}(y)$$

(Prop. 3.3.1, Rao & Ren)

The factor 2 is not very nice here; we will see that one obtains a better estimate when choosing a different norm. However, this estimate does show that $L^p_\varphi$ and $L^{p^*}_\varphi$ may act as dual spaces (in some sense; we will be more precise later).

Cl Norming the Orlicz space $\mathbb{R}^n$

Recall that, for a general Banach space,

$$\|x^*\|_{p^*} := \sup\{\langle x^*, x \rangle : \|x^*\| \leq 1\}$$

and

$$\|x\|_{p^*} := \sup\{\langle x^*, x \rangle : \|x\| \leq 1\}.$$

Motivated by this we define

$$\|x\|_\varphi := \sup\{\int |xy| \, d\mu : y \in L^{p^*}_\varphi, \int \varphi^\ast(|y|) \, d\mu \leq 1\}$$

Not very difficult to see that this is a norm!

(Rao & Ren, Th. 3.3.13)

\[\|x\|_\varphi = \min_{k \in \mathbb{R}_+} \left\{ \frac{1}{k} \left( 1 + \int \varphi^\ast(|kx|) \, d\mu \right) \right\} \text{ (Amemiya norm)}\]

This is a very practical expression, since:

$$\|x\|_\varphi \leq 1 + \int \varphi^\ast(|x|) \, d\mu$$

Hence a uniform bound on $\int \varphi^\ast(|x|) \, d\mu$ yields a uniform bound on the Orlicz norm!
Similarly to the unit ball property of the Luxemburg norm, we now have:

$$\text{Prop} \quad \mathcal{S}_\varphi \left( \frac{x}{\|x\|_{\varphi}} \right) \leq 1 \quad \forall x \neq 0 \in L_\varphi^{\infty}(\Omega)$$

(Rao & Ren Prop. 3.3.3)

From this for any \(x \in L_\varphi^{\infty}, y \in L_\varphi^{\infty}^*:\)

$$\|x \cdot y\|_{\varphi} := \sup \{ \int \! x y \, d\mu : y \in L_\varphi^{\infty}^* \text{ with } \mathcal{S}_\varphi(y) \leq 1 \}$$

$$\geq \int \! x y \, d\mu \quad \text{since } \mathcal{S}_\varphi \left( \frac{y}{\|y\|_{\varphi}^*} \right) \leq 1.$$

Hence we find again a Hölder-type estimate:

$$\int \! x y \, d\mu \leq \|x\|_{\varphi} \|y\|_{\varphi}^*$$

Another Hölder-type estimate follows from the Luxemburg ball property:

$$\int \! x y \, d\mu \leq \|x\|_{\varphi} \mathcal{N}_\varphi(y)$$

**D**

Relation between the Orlicz and Luxemburg norms

The two norms are certainly not equal. In fact

$$\|x\|_{\varphi} = \mathcal{N}_\varphi(x) \iff x = 0 \text{ a.e.}$$

However, (Rao & Ren Prop. 3.3.4)

$$\text{Prop} \quad \mathcal{N}_\varphi(x) \leq \|x\|_{\varphi} \leq 2 \mathcal{N}_\varphi(x) \quad \forall x \in L_\varphi^{\infty}(\Omega)$$

(equivalence of norms!)

**proof.** Note that \(k = \|x\|_{\varphi}\) satisfies \(\mathcal{S}_\varphi \left( \frac{x}{k} \right) d\mu \leq 1\), hence

$$\mathcal{N}_\varphi(x) := \inf \{ k > 0 : \mathcal{S}_\varphi \left( \frac{x}{k} \right) \leq 1 \} \leq \|x\|_{\varphi}.$$

On the other hand,

$$\|x\|_{\varphi} = \sup \{ \int \! x y \, d\mu : y \in L_\varphi^{\infty}^* \text{ with } \mathcal{N}_\varphi(y) \leq 1 \}$$

Hölder

$$\leq \sup \{ 2 \mathcal{N}_\varphi(x) \mathcal{N}_\varphi(y) : y \in L_\varphi^{\infty}^* \text{ with } \mathcal{N}_\varphi(y) \leq 1 \}$$

$$= 2 \mathcal{N}_\varphi(x).$$

In other words the topologies generated by \(\mathcal{N}_\varphi\) and \(\|x\|_{\varphi}\) are the same! This also implies that

$$L_\varphi^{\infty}(\Omega), \|x\|_{\varphi} \text{ is a (complete) Banach space} \quad \text{(Rao & Ren Prop. 3.3.11)}$$
Last week we introduced the Orlicz class and space:
\[ L^\varphi (\mathbb{R}) := \{ x : \mathbb{R} \to \mathbb{R} \text{ meas. s.t. } \int \varphi (1 + |x|) \, dx < \infty \} \]
\[ L^\varphi (\mathbb{R}) := \{ x : \mathbb{R} \to \mathbb{R} \text{ meas. } \exists C > 0, \int \varphi (1 + |x|) \, dx < \infty \} . \]

Indeed, \( L^\varphi \) is the linear hull of \( L^\varphi \), i.e. the smallest vector space containing \( L^\varphi \). This can be seen as follows:

- Closedness under addition is no problem, if the space is closed under rescaling since by convexity
  \[ \varphi (ax + y) \, dx \leq \frac{1}{a} \int \varphi (1 + |x|) \, dx + \frac{1}{a} \int \varphi (2y) \, dx. \]

- In order to make \( L^\varphi \) closed under any rescaling \( \alpha > 0 \), we need to take \( L^\varphi \)!

Recall that \( L^\varphi = L^\varphi \) iff \( \varphi \) has the \( \Delta_2 \)-property:
\[ \Leftrightarrow \exists K \forall z \text{ (suff. large)} \quad \varphi (Kz) \leq K \varphi (z) . \]

Note that \( \varphi(z) = \frac{1}{p} |z|^p \) is special in that \( \varphi(2z) = \frac{2^p}{K} \varphi(z) \).

- The factor \( 2 \) is arbitrary, since
  \[ \varphi (2z) \leq K \varphi (z) . \]

Two equivalent norms:
\[ N_\varphi (x) := \inf \{ k > 0 : \int \varphi (1 + k|x|) \, dx \leq 1 \} \quad \text{Luxemburg (Cajane/Minlos)} \]
\[ \| x \|_\varphi := \sup \{ \int |xy| \, dx : y \in L^\varphi_*, \int \varphi (1 + |y|) \, dx \leq 1 \} \quad \text{Orlicz} \]

\[ A \]

The \( M^\varphi \)-space and its dual
\[ M^\varphi (\mathbb{R}) := \{ x : \mathbb{R} \to \mathbb{R} \text{ meas. s.t. } \forall \alpha > 0, \int \varphi (1 + \alpha|y|) \, dx < \infty \} \]

Clearly, \( M^\varphi \) is a vector space, and
\[ M^\varphi \subset L^\varphi \subset L^\varphi . \]

In fact, \( M^\varphi \) is also closed under either norm, and hence also a (complete) Banach space! (Rao & Ren prop. 3.4.3)

\[ \text{Th} \quad \text{If } \varphi \text{ has the } \Delta_2 \text{-property then } M^\varphi = L^\varphi = L^\varphi . \]

(Rao & Ren cor. 3.4.5)
proof. Take \( x \in L^p \) with \( \alpha > 0 \) s.t. \( \int x(\alpha x) dy < \infty \).

Need to prove that \( \int x(\alpha x) dy < \infty \) for any \( \beta > 0 \), so that \( x \in M^p_\alpha \).

Let \( \beta \) be such that \( \beta \leq 2^n \alpha \). Then
\[
\int x(\beta x) dy \leq \int x(2^n \alpha x) dy \leq K^n \int x(\alpha x) dy < \infty.
\]

---

\[
\text{Th.} \quad (M^p_\alpha, N_\alpha)^* = (L^p_\alpha, \| \cdot \|_{N_\alpha}) \quad \text{(Rao-Ren Th. 4.1.7)}
\]

and
\[
(M^p_\alpha, \| \cdot \|_{N_\alpha})^* = (L^p_\alpha, N_\alpha^*),
\]

i.e.
\[
(M^p_\alpha)^* = L^p_\alpha^* \quad \text{and}
\]
\[
\| x \|_{(M^p_\alpha, N_\alpha)^*} = \sup \{ \int x(x) dy : x \in L^p_\alpha, \quad N_\alpha(x) \leq 1 \} = \| x \|_{L^p_\alpha}
\]
\[
\| x \|_{(M^p_\alpha, \| \cdot \|_{N_\alpha})^*} = \sup \{ \int x(x) dy : x \in L^p_\alpha, \quad \| x \|_{N_\alpha} \leq 1 \} = N_\alpha^*(x).
\]

(Either norm corresponds to the other norm in the dual space!)

Cor. 0. if \( \varphi \) has the \( A_2 \)-property then \( (M^p_\alpha)^* = (L^p_\alpha)^* = L^p_\alpha \).

2. if both \( \varphi \) and \( \varphi^* \) have the \( A_2 \)-property then \( L^p_\alpha \) is reflexive.

Important: \( L^p_\alpha \) always has a predual (whether \( A_2 \) holds or not)! Hence weak-* compact level sets by Banach-Alaoglu! If in addition, \( L^p_\alpha \) is separable then weak-* sequentially complete level sets.

Prop. Let \( A \subset R^n \) with Borel \( \sigma \)-algebra \( A \) and non-atomic measure \( \mu \in P(A) \). \( L^p_\alpha \) is separable iff \( \varphi \) has the \( A_2 \)-property.

(Rao-Ren Th. 3.5.1)
An example (last week's exercise)

\[ \Psi(Z) = Z \log Z - Z \]
\[ \Psi_2(Z) = -Z \log(-3Z) + Z \]
\[ \Psi(Z) = (\Psi_1 + \Psi_2)(Z) \]

\[ F(x) = \int \psi(x) \, dy \]

Minimiser? grows slightly faster than linearly, but slower than polynomial, for any \( p > 1 \).

\[ \phi \]

hence \( \Delta_2 \)-property satisfied:
\[ \phi(Z) \leq 2^p \phi(Z) \]
for \( Z \) sufficiently large

Horrible expressions if you'd want to calculate \( \Psi \) and \( \phi \) explicitly! Instead, work with their duals:
\[ \Psi^*(Z^*) = \frac{1}{2} e^{-Z} \]
\[ \Psi_2^*(Z^*) = \frac{3}{2} e^{-Z} \]
\[ \Psi^*(Z^*) = \Psi_1^*(Z^*) + \Psi_2^*(Z^*) = \frac{1}{2} e^{-Z} + \frac{3}{2} e^{-Z} \]
\[ \Psi^*(Z^*) = \cosh(Z^*) - 1 \]

Clearly \( \Psi^*(Z^*) \leq \phi^*(Z^*) + 1 \), and so

\[ \Psi(Z) \geq \phi(Z) - 1 \]

Hence on level sets \( \{ F \leq C \} \):

\[ \| x \|_\phi \leq 1 + \int \phi(1|x|) \, dy \leq 2 + \int (\phi(1|x|) - 1) \, dy \]
\[ \leq 2 + \int \psi(x) \, dy \leq 2 + C \quad \text{uniformly bounded!} \]

compactness:

Banach-Alaoglu: \( \{ F \leq C \} \) is weakly-* compact in \( L_1^\phi \).

This is meaningful since \( M_{\phi}^* \) is the predual of \( L_1^\phi \).

Be aware however that \( \phi^* \) grows exponentially and can not satisfy the \( \Delta_2 \)-property, hence \( M_{\phi}^{**} \not\subseteq L_1^{\phi^*} \).

If we take \( \Omega \in \mathbb{R}^n \) and \( \mu \) a (rescaled) Lebesgue measure, then \( L_1^\phi \) is separable and \( \{ F \leq C \} \) is weakly-* seq. cpt. lsc:

Write:

\[ F(x) = \int [\sup_{z^*} z^* x - \Psi_1^*(z^*) - \Psi_2^*(z^*)] \, dy \]
\[ = \sup_{x^* \text{ meas.}} \int [x^* x - \Psi_1^*(x^*) - \Psi_2^*(x^*)] \, dy \]
(to prove)
\[ = \sup_{x^* \in M_{\phi}^*} \int [x^* x - \Psi_1^*(x^*) - \Psi_2^*(x^*)] \, dy \rightarrow F \text{ weakly-* lsc.} \]

Direct method: \( \exists \) minimiser \( x \in L_1^\phi \) of \( F(x) \) (unique by strict convexity)

Remark: we could also have worked with \( L_1^\phi \)-weak and unif. integrability.