Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Hamburg University

Optimization

Notes

Department of Mathematics, Hamburg University, Bundesstrasse 55 , 20146, Hamburg, Germany

Abstract:

Keywords: Optimization• Convexity

Introduction

0.1. Basic Definitions

Definition 0.1 (Notation for special sets.).

Let X be a real vector space, let C and D be subsets of X, and let $z \in X$. Then we use the following notation:

- $C + D = \{x + y \mid x \in C, y \in D\}$.
- $C - D = \{x - y \mid x \in C, y \in D\}$.
- $z + C = \{z\} + C$.
- $C - z = C - \{z\}$.
- For every $\lambda \in \mathbb{R}$, $\lambda C = \{\lambda x \mid x \in C\}$.
- If Λ is a non-empty subset of $\mathbb{R}$, then $\Lambda C = \bigcup_{\lambda \in \Lambda} \lambda C$, and $\Lambda z = \Lambda\{z\} = \{\lambda z \mid \lambda \in \Lambda$.

Definition 0.2 (Cone).

Let X be a vector space, and $C \subset X$. C is called a cone with vertex O if $C = \mathbb{R}^+ C$. The cone is pointed or unpointed according to whether $O \in C$ or $\emptyset \notin C$ respectively. A pointed cone with vertex O is salient if $C \cap \{-C\} = \{O\}$

Remark 0.1 (Ordering relation and Cones).

We can associate a partial ordering relation denoted by $\preceq$(or $\succeq$) with a pointed cone C by setting the following relation,

$$p \preceq q \iff q - p \in C.$$

We can see the following properties, $p \preceq p$, $\forall p \in X$. If $p \preceq q$ and $q \preceq r$, then $p \preceq r$; the partial ordering relation is compatible with the structure of a vector space in the sense that

$$\begin{aligned} p \succeq 0 &\implies \lambda p \succeq 0, \quad \forall \lambda > 0 \\ p \succeq q &\implies p + r \succeq q + r, \quad \forall r \in X \end{aligned}$$

We see that the cone C is the set of positive elements for this ordering relation,

$$C = \{p \in X \mid p \succeq 0\}$$

The set $\{-C\}$ is the set of negative elements,

$$\{-C\} = \{p \in X \mid p \preceq 0\}$$

If the cone C is salient, the relation $\preceq$ is an ordering relation:

$$p \preceq q,\, q \preceq p \implies p = q$$

Definition 0.3.
Let C be a subset of U Banach space. The conical hull of C is the intersection of all the cones in U containing C, i.e., the smallest cone in U containing C. It is denoted by $\operatorname{cone} C$. The closed conical hull of C is the smallest closed cone in U containing C. It is denoted by $\overline{\operatorname{cone}}\, C$.

Definition 0.4.
We say a functional J is proper if $\operatorname{dom} J \neq \emptyset$ and $J > -\infty$.

Definition 0.5.
Let U a vector topological space. We define as the indicator function $I_C : U \to \overline{\mathbb{R}}$ of a set $C \subset U$, as follows:

$$I_C(x) = \begin{cases} 0 & x \in C \\ \infty & \text{otherwise.} \end{cases}$$

0.2. Useful lemmas and Theorems.

Lemma 0.1.
Let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence in a Hilbert Space H. Then $(x_n)_{n\in\mathbb{N}}$ possesses a weakly convergent subsequence.

Lemma 0.2.
Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in a Hilbert Space H. Then $(x_n)_{n\in\mathbb{N}}$ converges if and only if it is bounded and possesses at most one weak sequential cluster point.

Fact 0.1.
A Banach space is reflexive if its unit ball is compact in the weak topology. This implies that every bounded sequence admits a weakly converging subsequence. Hilbert spaces and L^p spaces $(1 < p < \infty)$ are reflexive.

Theorem 0.1.
Let $f : H \to (\infty, \infty]$ be a convex functional on a Hilbert space. Then the following are equivalent:

(i) f is weakly sequentially lower semicontinuous.

(ii) f is sequentially lower semicontinuous.

(iii) f is lower semicontinuous.

(iv) f is weakly lower semicontinuous.

Lemma 0.3.
A convex set is closed if and only if it is weakly closed.

Lemma 0.4.
Every bounded linear operator over a Banach Space is weakly continuos.

Lemma 0.5 (Parallelogram law).

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

Lemma 0.6.
Let $\mathcal{X}$ be a Hausdorff space and let $(f_i)_{i\in I}$ be a family of lower semicontinuous functions from $\mathcal{X}$ to $[-\infty,\infty]$. Then $\sup_{i\in I} f_i$ is lower semi-continuous. If I is finite, then $\min_{i\in I} f_i$ is lower-semicontinuous.

Definition 0.6.
Let $\mathcal{X}$ be a Hausdorff space. The lower semicontinuous envelope of $f : \mathcal{X} \to [-\infty,\infty]$ is

$$\overline{f} = \sup\left\{g : \mathcal{X} \to [-\infty,\infty] \mid \quad g \leq f \text{ and } g \text{ is lower semicontinuous}\right\}.$$

Proposition 0.1.
If C is a compact set in a normed space U, and G is a closed subset of C. Then G is compact.

Proof. Let $\{g_n\}$ a sequence contained in G. Since $G \subset C$ and C compact. $\exists\{g_n\}_k$ subsequence of $\{g_n\}$, contained in G such that $\{g_n\}_k \to g$, as $k \to \infty$, and then since G is closed $g \in G$. Therefore G is compact. □

Definition 0.7.
Let U a vector real space. We denote the set of all

Proposition 0.2.
Let $\phi : \mathbb{R} \to (-\infty,\infty]$ be a proper function that is differentiable on a nonempty open interval C in $\operatorname{dom}\phi$. Then the following hold:

- *ϕ' is increasing on C. Then $\phi + I_C$ is convex.*
- *Suppose that ϕ' is strictly increasing on C. Then ϕ is strictly convex on C.*

1. Lecture 1

1.1. Infinite-Dimensional Optimization

Let (U,d) be a metric space and $J : U \to \overline{\mathbb{R}}$. We call a minimization problem.

$$\min_{u\in C} J(u)$$

Definition 1.1.
A point $u \in U$ is called:

- **Local Minimizer.** If there is a neighborhood $V \in U$ such that $J(u) \leq J(v)$, $\forall v \in V$.
- **Global Minimizer.** If $J(u) \leq J(v)$, $\forall v \in U$.

Definition 1.2.
Let be $\{u_k\} \in U$, a convergent sequence in U, such that converges to $u \in U$. The functional J is called lower semicontinuous at $u \in U$ if

$$J(u) \leq \liminf_{k\to\infty} J(u_k).$$

In general if J is lower semicontinuous at u, for all the $u \in U$. J is lower semicontinuous (l.s.c).

Theorem 1.1.
Let $J : U \to \mathbb{R}$ lower semicontinuous functional and $\exists \xi \in \mathbb{R}$, such that the level set $\mu_\xi = \{u \in U \mid J(u) \leq \xi\}$ be non-empty and compact set of U. Then there exists a global minimum.

Proof. Let $\alpha := \inf_{u\in U} J(u)$. Then $\exists\{u_n\} \in U$ such that $J(u_n) \to \alpha$. Then $\exists N \in \mathbb{N}$, such that $\forall k \geq N$, $J(u_k) \leq \xi$ (otherwise $\xi = \alpha$), since μ_ξ is not empty, we have $u_k \in \mu_\xi$. Since μ_ξ is compact, $\exists\{u_k\}_l$ a subsequence of $\{u_k\}$ that converges in μ_ξ, i.e. $\{u_k\}_l \to \overline{u} \in \mu_\xi$, as $l \to \infty$. Since α is the infimimum and J is lower semicontinuous and,

$$\alpha \leq J(\overline{u}) \leq \liminf_{l\to\infty} J(u_{k_l})$$

On the other hand, since $J(u_k) \to \alpha$,

$$\liminf_{l\to\infty} J(u_k) \leq \alpha$$

Therefore $J(\overline{u}) = \alpha$, and hence $\overline{u}$ exists and it is a global minimizer. □

Corollary 1.1.
Let U be a Banach space. If the following conditions hold:

- $\exists\mu_\xi \in U$ *(level set) non-empty and compact.*
- $J : U \to \overline{\mathbb{R}}$ *is lower semicontinuous.*

Then set of global minimizers G is compact.

Proof. The theorem 1.1 implies that all minimizers are in the set μ_ξ. Therefore by proposition 0.1, G is precompact. Since J is lower semicontinuous, for any convergent sequence $(u_k) \in G$, we have

$$\alpha \leq J(u) \leq \liminf_{k\to\infty} J(u_k) = \alpha$$

Implying that the limit is also a global minimizer. Hence G is closed. □

1.2. Derivatives

Let U and V Banach spaces and $F : U \to V$ a mapping from U to V (that could be non linear).

Definition 1.3.
Let C be a subset of U, let $F : C \to V$, and let $x \in C$ be such that, for all $y \in U$, $\exists\alpha > 0$ and the set $[x, x + \alpha y] \subset C$. Then F is Gâteaux differentiable at x if there exists an operator $\mathsf{D}F(x) \in \mathcal{B}(U, V)$, called the Gâteaux derivative of F at x, such that,

$$\forall(y \in U) \quad \mathsf{D}F(x) = \lim_{\alpha\downarrow 0} \frac{F(x + \alpha y) - F(x)}{\alpha}$$

Thus, the second Gâteaux derivative of F at x is the operator $\mathsf{D}^2F(x) \in \mathcal{B}(U, \mathcal{B}(U, K))$ that satisfies

$$(\forall y \in U) \quad \mathsf{D}^2F(x)y = \lim_{\alpha\downarrow 0} \frac{\mathsf{D}F(x + \alpha y) - \mathsf{D}F(x)}{\alpha}$$

Remark 1.1.

The Gâteaux derivative $\mathsf{D}F(x)$ is unique whenever it exists.

Definition 1.4.

Let $x \in U$, let C a set contained in a neighborhood $\mathcal{V}(x)$ de x, and let $F : C \to V$. Then F is Fréchet differentiable at x if there exists an operator $\mathsf{D}F(x) \in \mathcal{B}(U, V)$, called the Fréchet derivative of F at x, such that

$$\lim_{0 \neq \|y\| \to 0} \frac{\|F(x+y) - F(x) - \mathsf{D}F(x)y\|}{\|y\|} = 0.$$

Higher-order Fréchet derivatives are defined inductively. Thus, the second Fréchet derivative of F at x is the operator $\mathsf{D}^2F(x) \in \mathcal{B}(U, \mathcal{B}(U, V))$ that satisfies,

$$\lim_{0 \neq \|y\| \to 0} \frac{\left\|\mathsf{D}F(x+y) - \mathsf{D}F(x) - \mathsf{D}^2F(x)y\right\|}{\|y\|} = 0.$$

Lemma 1.1.

Let $x \in U$, let be C a set $\mathcal{V}(x)$ contained in a neighborhood of x, and let $F : C \to V$. Suppose that F is Fréchet differentiable at x. Then the following hold:

- *F is Gâteaux differentiable at x and the two derivatives coincide.*
- *F is continuous at x.*

Proof. Denote the Fréchet derivative of F at x by L_x.

- Let $\alpha > 0$ and $y \in U \backslash \{0\}$. Then

$$\left\|\frac{F(x+\alpha y) - Fx}{\alpha} - L_x y\right\| = \|y\| \frac{\|F(x+\alpha y) - Fx - L_x(\alpha y)\|}{\|\alpha y\|}$$

converges to 0 as $\alpha \downarrow 0$, since F is Fréchet differentiable.

- Fix $\epsilon > 0$. By definition 1.4, we can find $\delta \in (0, \frac{\epsilon}{\epsilon + \|L_x\|}]$, such that for all y in the open ball of radius δ and center in zero, (i.e. $\forall y \in B_\delta(0)$),

$$\|F(x+y) - Fx - L_x y\| \leq \epsilon \|y\|$$

Thus, $\forall y \in B_\delta(0)$, by triangle inequality,

$$\begin{aligned} \|F(x+y) - Fx\| &\leq \|F(x+y) - Fx - L_x y\| + \|L_x y\| \\ &\leq \epsilon \|y\| + \|L_x\| \|y\| \\ &\leq \delta(\epsilon + \|L_x\|) \\ &\leq \epsilon. \end{aligned}$$

It follows that F is continuous at x.

□

Fact 1.1.

Let $x \in U$, let $\mathcal{U}$ be a neighborhood of x, and let G be a real Banach space, let $T : \mathcal{U} \to G$ a mapping from $\mathcal{U}$ to G, let $\mathcal{V}$ be a neighborhood of Tx, and let $R : V \to K$. Suppose that T is Fréchet differentiable at x and that R is Gâteaux differentiable at Tx. Then $R \circ T$ is Gateâux differentiable at x and $\mathsf{D}(R \circ T)(x) = (\mathsf{D}R(Tx)) \circ \mathsf{D}T(x)$. If R is Fréchet differentiable at x, then so is $R \circ T$.

Fact 1.2.
Let $x \in U$, let $\mathcal{U}$ be a neighborhood of x, let G be a real Banach space, and let $F : U \to G$. Suppose that F is twice Fréchet differentiable at x. Then $\forall (y, z) \in U \times U$, $(\mathsf{D}^2 F(x)y)z = (\mathsf{D}^2 F(x)z)y$.

Remark 1.2 (Notation).
The result of applying the operator $\mathsf{D}F(x)$ (derivative at x) applied to some vector y will be denote by apostrophe to the function $F'(x; y) = \mathsf{D}F(x)y$. If we are in a Hilbert Space $(H, \langle\cdot\rangle)$ by Riesz representation formula and the fact $\mathsf{D}F(x) \in \mathcal{B}(H, H)$, we have

$$F'(x; y) = \mathsf{D}F(x)y = \langle y, \nabla F(x)\rangle,$$

where $\nabla F(x) \in H$ and it represents the operator $\mathsf{D}F(x)$. Sometimes we also denote $F'(x) = \nabla F(x)$.

2. Lecture 2

2.1. Convexity

Definition 2.1.
Let U be linear space. A functional $J : U \to \overline{\mathbb{R}}$ is called convex, if for $t \in [0, 1]$ and $u_1, u_2 \in U$.

$$J(tu_1 + (1-t)u_2) \le tJ(u_1) + (1-t)J(u_2) \tag{1}$$

holds such that the right hand side is well defined.

- J is strictly convex if (1) holds strictly for $\forall u_1, u_2 \in U$, $u_1 \neq u_2$ and $t \in (0, 1)$ with $J(u_1) < \infty$ and $J(u_2) < \infty$.
- An optimization problem,

$$\min_{u \in C} J(u)$$

  is called convex if both C and J are convex.

Lemma 2.1.
If C and V are convex subsets in U, then

1. *αV is convex, for $\alpha \in \mathbb{R}$.*
2. *$C + V$ is convex.*

Proof. If V or C are empty sets then the statement is immediately true. Consider $V, C \neq \emptyset$.

1. Let αv_1 and αv_2 be elements of αV. By definition 0.1, v_1 and v_2 are elements of V. For $t \in [0, 1]$, $v_t = tv_1 + (1-t)v_2$. Therefore,

$$\alpha v_t = \alpha\,(tv_1 + (1-t)v_2) = t(\alpha v_1) + (1-t)(\alpha v_2)$$

   Since V is convex $v_t \in V$, and hence $\alpha v_t \in \alpha V$.

2. Let $w_1, w_2 \in C + V$, therefore $\exists c_1, c_2 \in C$ and $\exists v_1, v_2 \in V$ such that $c_1 + v_1 = w_1$ and $c_2 + v_2 = w_2$. Since C and V are convex. For $t \in [0, 1]$. Let $w_t = tw_1 + (1-t)w_2$.

   Let $c_1, c_2 \in C$ and $v_1, v_2 \in V$. For $t \in [0, 1]$, $C \ni c_t = tc_1 + (1-t)c_2$ and $V \ni v_t = tv_1 + (1-t)v_2$. Therefore,

$$w_t = tw_1 + (1-t)w_2 = t(c_1 + v_1) + (1-t)(c_2 + v_2) = c_t + v_t$$

   Therefore by definition 0.1, $c_t + v_t = w_t \in C + V$

□

Lemma 2.2.
Let V be a collection of convex sets in U, then $C = \bigcap_{K\in V} K$ is convex.

Proof. If $C = \emptyset$, then C the statement is vacuously true. Consider $C \neq \emptyset$ and $u_1, u_2 \in C$ then $u_1, u_2 \in K$ for all $K \in V$

$$\implies tu_1 + (1-t)u_2 \in K, \quad \forall K \in V \implies tu_1 + (1-t)u_2 \in \bigcap_{K\in V} K$$

□

Lemma 2.3.
Let U a Banach Space, let $C \subset U$ convex subset of U and $J : C \to \mathbb{R}$ a functional defined over C. Define $\alpha = \inf_{u\in C} J(u)$. Then the set $\Psi = \{u \in C | J(u) = \alpha\}$ is convex, i.e. the solution of

$$\min_{u\in C} J(u)$$

is a convex set.

Proof. Let $u_1, u_2 \in \Psi$ and $u_t = tu_1 + (1-t)u_2$. Since J is convex, it holds that $J(u_t) \leq tJ(u_t) + (1-t)J(u_t) = \alpha$. Thus $J(u_t) = \alpha$, $\forall t \in [0, 1]$. Implying $u_t \in \Psi$. Hence Ψ is convex. □

Lemma 2.4.
Let U be linear normed space, and $C \subset U$ a convex set and $J : U \to \overline{\mathbb{R}}$ convex functional. Let $\overline{u} \in C$ such that

$$J(\overline{u}) \leq J(u) \quad \forall u \in B_\epsilon(\overline{u}) \cap C,$$

for some ball $B_\epsilon(\overline{u})$ in U with center in $\overline{u}$. Then $J(\overline{u}) = \inf_{u\in C} J(u)$. In other words, the local minimizer of a convex optimization problem is also a global minimizer.

Proof. Let $B_\epsilon(\overline{u})$ be an open neighborhood of $\overline{u}$ with $J(\overline{u}) \leq J(u)$ for all $u \in B_\epsilon(\overline{u}) \cap C$. Take an arbitrary $u^* \in C$ and consider $u_t = t\overline{u} + (1-t)u^*$. Since C is convex $u_t \in C$.
For some $t \in (0, 1)$, $u_t \in B_\epsilon(\overline{u})$.
Thus,

$$J(\overline{u}) \leq J(u_t) \leq tJ(\overline{u}) + (1-t)J(u^*).$$

We have $\forall t \in [0, 1]$ that $(1-t) \geq 0$, then

$$(1-t)J(\overline{u}) \leq (1-t)J(u^*) \qquad \forall u^* \in C$$

Therefore, $\overline{u}$ is a local minimizer for C. □

Theorem 2.1.
Let U is Banach Space, $C \subset U$ convex and $J : C \to \mathbb{R}$ Gateâux differentiable. Consider the minimization problem.

$$\min_{u\in C} J(u)$$

1. *Let $\overline{u}$ be a local solution. Then $J'(\overline{u}; u - \overline{u}) \geq 0$, $\forall u \in C$.*
2. *If J is convex on C, then $J'(\overline{u}; u - \overline{u}) \geq 0$, $\forall u \in C$ is necessary and sufficient for global optimality of $\overline{u}$*
3. *If J is strictly convex on C, then the minimization problem admits at most one solution.*

4. *If C is closed, and J is convex and continuous with*

$$\lim_{\substack{u \in C \\ \|u\| \to \infty}} J(u) = \infty.$$

Then a global solution $\overline{u} \in C$ exists.

Proof.

1. Let $\overline{u}$ be a local solution $J(\overline{u}) \leq J(u)$, $\forall u \in B_\epsilon(\overline{u}) \cap C$, let $t \in [0,1]$, $u_t = \overline{u} + t(u - \overline{u})$, then $u_t \in C$, since C is convex.

   For small $t > 0$,

$$0 \leq \frac{1}{t}\left[J(u_t) - J(u)\right] \leq \frac{1}{t}\left[J(\overline{u} + t(u - \overline{u})) - J(u)\right] \xrightarrow{t \downarrow 0} J'(\overline{u}; u - \overline{u})$$

2. Since J is convex we have for $u \in C$, $J(\overline{u} + t(u - \overline{u})) \leq J(\overline{u}) + t\left[J(u) - J(\overline{u})\right]$, for $t > 0$

$$\implies J(u) - J(\overline{u}) \geq \frac{1}{t}\left[J(\overline{u} + t(u - \overline{u})) - J(\overline{u})\right] \xrightarrow{t \downarrow 0} J'(\overline{u}; u - \overline{u}) \geq 0.$$

   Therefore $\overline{u}$ is a global minimizer.

3. Assume, that there are two solution for the minimization problem, $\overline{u}, u^* \in C$, such that $\overline{u} \neq u^*$ and $J(\overline{u}) = J(u^*) = \inf_{u \in C} J(u)$. Since J is strictly convex $J(u_t) = J(t\overline{u} + (1-t)u^*) < tJ(\overline{u}) + (1-t)J(u^*) = \alpha$ for all $t \in [0,1]$. Contradicting our assumption that u^* and $\overline{u}$ are solutions.

4. $\alpha = \inf_{u \in C} J(u) \in \mathbb{R} \cup \{-\infty\}$, choose a minimizing sequence $(u_k)_k \subset C$ with $J(u_k) \xrightarrow{k \to \infty} \alpha$

   $\implies (u_k)_k$ is bounded, because $J \to \infty$ as $\|u\| \to \infty$.
   $\implies (u_k)_k$ contains a weakly convergent subsequence $u_{k_e} \xrightarrow[e \to \infty]{} \overline{u} \in C$. Since C is closed and convex.
   $\implies J$ is weakly-lower semicontinuos because it is convex and continuos.

$\square$

3. Lecture 3

Now consider Hilbert space $(H, \langle \cdot, \cdot \rangle)$ with the norm defined as $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

Let be $J : H \to \mathbb{R}$ a functional over a Hilbert space H, we define the set,

$$\operatorname*{arg\,min}_{v \in C \subseteq H} J(x) := \{x \mid x \in H \wedge \forall v \in C : \; J(x) \leq J(v)\}.$$

By Riesz-Fréchet representation formula, exists a unique vector $\nabla J(x) \in H$ such that,

$$(\forall y \in H) \quad J'(x; y) = \langle y, \nabla J(x) \rangle$$

namely Gateâux gradient of J at x.

Lemma 3.1.
Let H Hilbert space and $C \subset H$ closed and convex. Define $P_C : H \to C$,

$$P_C(x) = \arg\min_{v \in C} [\|v - x\|].$$

Then,

1. *P_C is well defined, i.e. $\forall x \in H$, $\exists! u \in C$ such that $P_C(x) = \{u\}$.*
2. *$\forall u, x \in H$, we have $u = P_C(x) \iff u \in C$ and $\langle x - u, v - u\rangle \leq 0$ $\forall v \in C$.*
3. *$\forall x, y \in H$, $\langle y - x, P_C(y) - P_C(x)\rangle \geq 0$.*
4. *The projection P_C is non expansive. That is $\|P_C(x) - P_C(y)\| \leq \|x - y\|$, $\forall x, y \in H$,*
5. *Let be $t > 0$ a real number, then $\forall u \in C$, and $\forall v \in H$, $\phi(t) = \frac{1}{t}\left\|P_C\left(u + tv\right) - u\right\|$ is not increasing.*

Proof.

1. First we prove existence, let be $(v_k)_k$ a minimizing sequence in C, such that

$$\|x - v_k\| \to \alpha = \inf_{v \in C} \|x - v\|,$$

By the parallelogram law,

$$\begin{aligned}
2\left\|v_j - x\right\|^2 + 2\left\|v_i - x\right\|^2 &= \left\|v_j - v_i\right\|^2 + \left\|v_j + v_i - 2x\right\|^2 \\
2\left\|v_j - x\right\|^2 + 2\left\|v_i - x\right\|^2 &= \left\|v_j - v_i\right\|^2 + 4\left\|\frac{v_j + v_i}{2} - x\right\|^2 \\
\implies 2\left\|v_j - x\right\|^2 + 2\left\|v_i - x\right\|^2 - 4\left\|\frac{v_j + v_i}{2} - x\right\|^2 &= \left\|v_j - v_i\right\|^2
\end{aligned}$$

Since C is convex $\frac{v_i + v_j}{2} \in C$, then by definition of α,

$$0 \leq \alpha \leq \left\|\frac{v_j + v_i}{2} - x\right\|$$

Therefore the above equations become in the following inequality,

$$2\left\|v_j - x\right\|^2 + 2\left\|v_i - x\right\|^2 - 4\alpha^2 \geq \left\|v_j - v_i\right\|^2$$

Since $\|v_i - x\| \to \alpha$ and $\|v_j - x\| \to \alpha$, we have that $\|v_j - v_i\| \to 0$, therefore the series is Cauchy and then converges. Since C is closed the series converges to a point $v \in C$.

Second we prove uniqueness, we proceed by contradiction, take $v, v' \in C$ such $v \neq v'$, and both of them minimizing the distant with respect the point x, i.e.

$$\|x - v\| = \left\|x - v'\right\| = \alpha = \min_{u \in C} \|u - x\|$$

By the parallelogram law,

$$2\left\|x - v\right\|^2 + 2\left\|x - v'\right\|^2 = \left\|2x - v - v'\right\|^2 + \left\|v - v'\right\|^2$$

Since C is convex, $\left\| \frac{v+v'}{2} - x \right\| \geq \alpha$

$$\begin{aligned}
\left\| v - v' \right\|^2 &= 2\left\| x - v \right\|^2 + 2\left\| x - v' \right\|^2 - \left\| 2x - v - v' \right\|^2 \\
\left\| v - v' \right\|^2 &= 2\left\| x - v \right\|^2 + 2\left\| x - v' \right\|^2 - 4\left\| x - \frac{v - v'}{2} \right\|^2 \\
\left\| v - v' \right\|^2 &= 2\alpha^2 + 2\alpha^2 - 4\left\| x - \frac{v - v'}{2} \right\|^2 \leq 0
\end{aligned}$$

Therefore $\|v - v'\| = 0$, and $v = v'$.

By the uniqueness and existence $\arg\min_{u \in C} [\|u - x\|]$ is not empty set and has only one element for each $x \in H$. Thus, P_C is well defined.

2. First consider that $u \in C$ and $\langle x - u, v - u \rangle \leq 0$, for all $v \in C$. Then,

$$\begin{aligned}
\langle x - u, v - u \rangle &= \langle x - u, x - x + v - u \rangle \\
&= \langle x - u, x - u \rangle + \langle x - u, v - x \rangle \\
&= \|x - u\|^2 + \langle x - u, v - x \rangle \\
&\leq 0
\end{aligned}$$

Therefore, applying Cauchy-Schwartz inequality we have

$$\begin{aligned}
\|x - u\|^2 &\leq \langle x - u, x - v \rangle \\
&\leq \|x - u\| \, \|x - v\| \\
\implies \|x - u\| &\leq \|x - v\|
\end{aligned}$$

And the above holds for every $v \in C$, therefore $u = P_C(x)$. Now consider $u = P_C(x)$, since $P_C(x)$ is well defined by definition $u \in C$. Since C is convex for all $v \in C$, $u_t = u + t(v - u) \in C$, for all $t \in [0, 1]$. Since $u = P_C(x)$ and $u_t \in C$, we have $\|x - u\|^2 \leq \|x - u_t\|^2$

$$\begin{aligned}
\|x - u\|^2 &\leq \langle x - u_t, x - u_t \rangle \\
&= \langle x - u - t(v - u), x - u - t(v - u) \rangle \\
&= \langle x - u, x - u \rangle - 2\langle x - u, t(v - u) \rangle + t^2 \langle v - u, v - u \rangle \\
0 &\leq t^2 \langle v - u, v - u \rangle - 2t\langle x - u, v - u \rangle \\
0 &\leq t^2 \|v - u\|^2 - 2t\langle x - u, v - u \rangle
\end{aligned}$$

For $t = 0$ the above inequality holds (since $u_t = u$), and for all $t > 0$, we have

$$\langle x - u, v - u \rangle \leq \frac{t}{2} \|v - u\|^2 .$$

Taking $t \downarrow 0$,

$$\begin{aligned}
&\langle x - u, v - u \rangle \leq \inf_{t \in (0,1)} \frac{t}{2} \|v - u\|^2 \\
&\langle x - u, v - u \rangle \leq 0
\end{aligned}$$

3. By the above result, we have that $P_C(x) \in C$ and $P_C(y) \in C$, and

$$\langle x - P_C(x), P_C(y) - P_C(x)\rangle \leq 0 \text{ and } \langle y - P_C(y), P_C(x) - P_C(y)\rangle \leq 0$$

Rearranging and adding both inequalities we have,

$$\begin{aligned}
\langle x - P_C(x), P_C(y) - P_C(x)\rangle &\leq \langle y - P_C(y), P_C(y) - P_C(x)\rangle \\
\langle x, P_C(y) - P_C(x)\rangle - \langle P_C(x), P_C(y) - P_C(x)\rangle &\leq \langle y, P_C(y) - P_C(x)\rangle - \langle P_C(y), P_C(y) - P_C(x)\rangle \\
\langle P_C(y), P_C(y) - P_C(x)\rangle - \langle P_C(x), P_C(y) - P_C(x)\rangle &\leq \langle y, P_C(y) - P_C(x)\rangle - \langle x, P_C(y) - P_C(x)\rangle \\
\|P_C(y) - P_C(x)\|^2 &\leq \langle y - x, P_C(y) - P_C(x)\rangle.
\end{aligned}$$

The above inequality immediately implies $0 \leq \langle y - x, P_C(y) - P_C(x)\rangle$.

4. Using, Cauchy-Schwartz inequality we obtain,

$$\|P_C(y) - P_C(x)\|^2 \leq \langle y - x, P_C(y) - P_C(x)\rangle \leq \|y - x\| \, \|P_C(y) - P_C(x)\| .$$

Hence $\|P_C(y) - P_C(x)\| \leq \|y - x\|$, for all $x, y \in H$.

5. If $u + tv \in C$, then $P_C(u + tv) = u + tv$, therefore $\phi(t) = \|v\|$. And the function is not increasing. Now if $u + tv \notin C$, we can bound $\phi(t)$ as follows,

$$\phi(t) = \frac{1}{t}\|P_C(u + tv) - u\| = \frac{1}{t}\|P_C(u + tv) - P_C(u)\| \leq \|v\|$$

Since $\lim_{t\to 0} \phi(t) = \|v\|$, and for $t > 0$ we have the above inequality, then ϕ cannot be increasing.

□

Theorem 3.1.
Let H be Hilbert space, $C \subset H$ closed and convex, $J : C \to \mathbb{R}$, Gateâux differentiable at the local solution $\overline{u}$ of $\min_{u\in C} J(u)$. Thus, $J'(\overline{u}; u - \overline{u}) \geq 0$, $\forall u \in C$ and it is equivalent to $\overline{u} = P_C(\overline{u} - \delta\nabla J(\overline{u}))$, $\forall \delta > 0$.

Proof. Since every Hilbert Space is a Banach space, and C is closed and Convex subset of H, and $\overline{u}$ is a solution of minimization problem; we can apply 2.1.
Thus $J'(\overline{u}; u - \overline{u}) \geq 0 \iff \langle u - \overline{u}, \nabla J(\overline{u})\rangle \geq 0 \ \forall u \in C$.
For all $\delta > 0$, we multiply the Gateâux gradient $(-\delta)$ and we have,

$$\langle u - \overline{u}, -\delta\nabla J(\overline{u})\rangle \leq 0 \ \forall u \in C,$$

adding zero to the gradient, $\langle u - \overline{u}, \overline{u} - \delta\nabla J(\overline{u}) - \overline{u}\rangle \leq 0$. Then we set $w \in H$ as $w := \overline{u} - \delta\nabla J(\overline{u})$, and applying lemma 3.1 we have,

$$\overline{u} = P_C(w) \iff \langle u - \overline{u}, w - \overline{u}\rangle$$

Thus,

$$\overline{u} = P_C(\overline{u} - \delta\nabla J(\overline{u}))$$

□

3.1. Application

Consider U, Y, Z Hilbert spaces. Let be $J : Y \times U \to \mathbb{R}$ a functional. Consider the minimization problem,

$$\begin{cases} \overline{u} = \min\limits_{y,u} J(y,u) \\ Ay = Bu \quad u \in U_{ad} \subset U \end{cases}$$

For some set U_{ad} closed, convex and bounded. And $A \in \mathcal{B}(Y,Z)$ bounded and invertible with $A^{-1} \in \mathcal{B}(Z,Y)$ and $B \in \mathcal{B}(U,Z)$.

Then we can write $y \in Y$ as a function of $u \in U$,

$$y = y(u) = A^{-1}Bu.$$

Where $A^{-1}B \in \mathcal{B}(U,Y)$, and adjoint with respect to the inner product $(A^{-1}B)^{\star}$. Consider the reduced cost functional $F : U \to \mathbb{R}$, such that $u \mapsto J(y(u), u)$, then our problem is equivalent to

$$\overline{u}_{ad} = \min_{u \in U_{ad}} F(u)$$

Let $(u_k)_k \in U_{ad}$ denote a minimizing sequence, i.e. $F(u_k) \to \inf_{u \in U_{ad}} F(u)$, since $u_k \in U_{ad}$ the sequence is bounded. Therefore we can find a convergent subsequence $u_{k_l} \xrightharpoonup[l\to\infty]{} \overline{u}$, moreover since U_{ad} is closed and convex U_{ad} is weakly closed, implying $\overline{u} \in U_{ad}$

Proposition 3.1.
If J is continuous and weakly lower semicontinuous, then $\overline{u} = \arg\min\limits_{u \in U_{ad}} [F(u)]$.

Proof. If J is weakly lower semicontinuos

$$J(y(\overline{u}), \overline{u}) \leq \liminf_{l \to \infty} J(y(u_k), u_k)$$

That is,

$$F(\overline{u}) \leq \liminf_{l \to \infty} F(u_k) = \alpha$$

Since $u_{k_l} \xrightharpoonup[l\to\infty]{} \overline{u}, \implies y(u_k) \rightharpoonup y(\overline{u}) = \overline{y}$, and $A^{-1}Bu_k \rightharpoonup A^{-1}B\overline{u}$ □

If J is Gateâux differentiable, applying the chain rule (fact 1.1) to F and valuating in u we have

$$F_u(u) = \left. J_y(y,u) \right|_{y=y(u)} \circ y_u(u) + \left. J_u(y,u) \right|_{y=y(u)},$$

where $F_u(u) \in U^*$, $J_y(y,u) \in Y^*$, $y_u(u) \in \mathcal{B}(U,Y)$ and $J_u \in U^*$. Since U and Y are Hilbert spaces, there is a member of each space respectively representing this operator in the inner product. We take $\nabla_u F(u) \in U$,

$\nabla_y J(u) \in Y$, and $\nabla_u J(y,u) \in U$. Therefore we can write $F_u(u; u-\overline{u})$ as follows, and we evaluate our optimality condition,

$$\begin{aligned} 0 &\leq \langle u-\overline{u}, \nabla_u F(\overline{u})\rangle_{UU^*} \quad \forall u \in U_{ad} \\ &= \langle A^{-1}B(u-\overline{u}), \nabla_y J(\overline{y},\overline{u})\rangle_{YY^*} + \langle u-\overline{u}, \nabla_u J(\overline{y},\overline{u})\rangle_{UU^*} \\ &= \langle A^{-1}B(u-\overline{u}), \nabla_y J(\overline{y},\overline{u})\rangle_{YY^*} + \langle u-\overline{u}, \nabla_u J(\overline{y},\overline{u})\rangle_{UU^*} \\ &= \langle u-\overline{u}, (A^{-1}B)^\star \nabla_y J(y,\overline{u})\rangle_{UU^*} + \langle u-\overline{u}, \nabla_u J(\overline{y},\overline{u})\rangle_{UU^*} \\ &= \langle u-\overline{u}, (A^{-1}B)^\star \nabla_y J(\overline{y},\overline{u}) + \nabla_u J(\overline{y},\overline{u})\rangle_{UU^*} \end{aligned}$$

Setting $p^* = (A^{-1}B)^\star \nabla_y J(\overline{y},\overline{u})$. We have that $\overline{u} = P_{U_{ad}}(\overline{u} - \delta(p^* + \nabla_u J(\overline{y},\overline{u})))$

4. Lecture 4

Lemma 4.1.
Let U be linear space and $J : U \to \overline{\mathbb{R}}$. Then

1. *If J is convex, then the effective domain* $\mathrm{dom}(J) = \{u \in U |\ J(u) < \infty\}$ *is convex.*
2. *J is convex $\iff$* $\mathrm{epi}(J) = \{(u,\alpha) \in U \times \mathbb{R} | J(u) \leq \alpha\}$ *is convex.*

Proof. Since U and $\mathbb{R}$ are linear spaces, is easy to see that scalar multiplications and sums are well defined over $U \times \mathbb{R}$ and so over $\mathrm{epi}\,(J)$.

1. Assume J convex. If $u_1 \in \mathrm{dom}\,(J)$ and u_2 are elements of $\mathrm{dom}\,(J)$. Therefore, $J(u_1) < \infty$, and $J(u_2) < \infty$, therefore for $t \in [0,1]$, we have $tJ(u_1) < \infty$ and $(1-t)J(u_2) < \infty$. Since J is convex,

   $$J(tu_1 + (1-t)u_2) \leq tJ(u_1) + (1-t)J(u_2) < \infty$$

   ,
   Therefore, $tu_1 + (1-t)u_2 \in \mathrm{dom}\,(J)$. Hence $\mathrm{dom}\, J$ is convex.

2. First consider J a convex functional, then we have for all $u_1, u_2 \in U$,

   $$J(tu_1 + (1-t)u_2) \leq tJ(u_1) + (1-t)J(u_2) \quad \forall t \in [0,1]$$

   Let (u_1,α_1), (u_2,α_2) elements of $\mathrm{epi}\,(J)$, then $J(u_1) < \alpha_1$ and $J(u_2) < \alpha_2$. Since J is convex.

   $$J(tu_1 + (1-t)u_2) \leq t\alpha_1 + (1-t)\alpha_2$$

   Then $(tu_1 + (1-t)u_2, t\alpha_1 + (1-t)\alpha_2) \in \mathrm{epi}\,(J)$. Therefore, if J is convex, and (u_1,α_1), (u_2,α_2) are elements of $\mathrm{epi}\,(J)$ then,

   $$t(u_1,\alpha_1) + (1-t)(u_2,\alpha_2) = (tu_1 + (1-t)u_2, t\alpha_1 + (1-t)\alpha_2) \in \mathrm{epi}\,(J)$$

   Hence $\mathrm{epi}\,(J)$ is convex.
   Now assume $\mathrm{epi}\,(J)$ convex. Let (u_1,α_1), (u_2,α_2) elements of $\mathrm{epi}\,(J)$ then $(tu_1 + (1-t)t\alpha_1 + (1-t)\alpha_2)$, then

   $$J(tu_1 + (1-t)u_2) \leq t\alpha_1 + (1-t)\alpha_2 \quad \forall t \in [0,1]$$

   By definition of $\mathrm{epi}\,(J)$, if $u_1, u_2 \in \mathrm{dom}\, J$, then $(u_1, J(u_1))$ and $(u_2, J(u_2))$, are elements of $\mathrm{epi}\,(J)$, therefore

   $$J(tu_1 + (1-t)u_2) \leq tJ(u_1) + (1-t)J(u_2) \quad \forall t \in [0,1]$$

   Implying that J is convex.

□

Definition 4.1.

Let U a Banach space. Then the function $J : U \to \overline{\mathbb{R}}$ is called lower semi-continuous at $u_0 \in U$ if the following conditions holds:

- If $\forall \epsilon > 0$ there is a neighborhood $B_\delta(u_0)$ of u_0, such that $J(u_0) - \epsilon \leq J(u)$ $\forall u \in B_\delta(u_0)$.
- If $J(u_0) \leq \liminf_{n\to\infty} J(u_n)$ holds for each sequence $u_n \in U$.

Remark 4.1.

If the second condition holds, J is called sometimes sequentially semi-continuous. If J is continuous it is also lower semi-continuous.

Theorem 4.1.

Let U be a Banach space and $J : U \to \overline{\mathbb{R}}$. Then the following conditions are equivalent.

1. *J is lower semi-continuous, i.e., J is lower semi-continuous at every point in U.*
2. *The* epi (J) *is closed.*
3. *The level sets $\mu_\xi = \{u \in U | J(u) \leq \xi\}$ is a closed set. Note that the sets μ_ξ are closed if and only if the sets $\gamma_\xi = \{u \in U | J(u) > \xi\}$ are open. (Since $\mu_\xi^c = \gamma_\xi$).*

Proof.

- (1) $\implies$ (2) Let (u_n, ξ_n), be a sequence in epi (J), such that converges to (u, ξ) in $U \times \mathbb{R}$. Then

$$J(u) \leq \liminf_{n\to\infty} J(u_n) \leq \liminf_{n\to\infty} \xi_n = \xi.$$

  Hence $(u, \xi) \in$ epi (J).

- (2) $\implies$ (3)Let $\xi \in \mathbb{R}$ and assume that $(u_n)_{n\in\mathbb{N}}$ is a sequence in μ_ξ that converges to u. Then the set $(u_n, \xi)_{n\in\mathbb{N}}$ is in epi (J). Since epi (J) is closed, we conclude that $(u, \xi) \in$ epi (J), and hence $u \in \mu_k$.

- (3) $\implies$ (1) Let bet $u \in U$ an arbitrary member of the Banach space U, and let $(u_n)_{n\in\mathbb{N}}$ be a sequence that converges to u. And we set the number $\eta = \liminf_{n\to\infty} J(u_n)$. Then we have to prove that $J(u) \leq \eta$. When $\eta = \infty$, the inequality is clear. Therefore we assume that $\eta < +\infty$. Since every sequence in $\mathbb{R}$ has a subsequence that converges to the lim inf, the sequence $(u_n)_n$ has a subsequence $(u_k)_k$, such that $J(u_k) \xrightarrow{k\to\infty} \eta$. Now, we can fix $\xi \in (\eta, \infty)$. By convergence we can find c such that $k \geq c$ implies that $(J(u_k))$ belongs to $(-\infty, \xi)$, therefore the set

$$\{u_k | \; k \geq c \in \mathbb{N}\} \subset \mu_\xi.$$

  Since the sequence $u_n \to u$, the subsequence $u_k \to u$. And μ_ξ closed implies $u \in \mu_\xi$. Since this holds for all $\eta < \infty$, we take $\xi \downarrow \eta$. Implying $J(u) \leq \eta$.

□

Example 4.1.

The indicator function of a set $C \subset U$, i.e. the function $I_C : U \to [-\infty, \infty]$

$$I_C(u) = \begin{cases} 0, & \text{if } u \in C \\ +\infty, & \text{otherwise,} \end{cases}$$

is lower semi-continuous if and only if C is closed.

Proof. Take $\xi \in \mathbb{R}$. If $\xi < 0$, the set $\mu_\xi = \emptyset$. If $\xi > 0$, the set $\mu_\xi = C$. Therefore the sets m_ξ, for all $\xi \in \mathbb{R}$ is closed if and only if C is closed. By the theorem 4.1 I_C is lower semi-continuous if and only if C is closed. □

The Dual Systems of Linear Spaces

Two linear spaces X and Y over the same scalar field Γ define a dual system if a fixed bilinear functional on their product is given:

$$(\cdot \mid \cdot) : X \times Y \to \Gamma$$

.

The dual system is called separated if the following two properties hold:

1. $\forall x \in X \backslash \{0\}$ there is $y \in Y$ such that $(x \mid y) \neq 0$.

2. $\forall y \in Y \backslash \{0\}$ there is $x \in X$ such that $(x \mid y) \neq 0$.

In other words, X separates points in Y, and Y separates points in X. We consider only separated dual systems. For each $y \in Y$, we define the application $f_y : X \to \Gamma$ by

$$f_y(x) = (x \mid y) \quad \forall x \in X$$

We observe that f_y is a linear functional on X and the mapping $y \to f_y$, $\forall y \in Y$, is linear and injective, as can be seen from condition (1). Hence, the correspondence is an embedding. Thus, the elements of Y can be identified with the linear functionals on X. In a similar way, the elements of X can be considered as linear functionals of Y, identifying an element $x \in X$ with $g_x : Y \to \Gamma$, defined by

$$g_x(y) = (x \mid y), \quad \forall y \in Y.$$

Therefore, each dual system of linear spaces defines a mapping from either of the two linear spaces into the space of linear functionals on the other.

We set,

$$p_x(y) = |(x \mid y)| = |g_x(y)|, \quad \forall y \in Y$$
$$q_y(x) = |(x \mid y)| = |f_y(x)|, \quad \forall x \in X$$

and we observe that $\mathcal{P} = \{p_x |\ x \in X\}$ is a family of seminorms on Y and $\mathcal{Q} = \{q_y |\ y \in Y\}$ is a family of seminorms on X.

Definition 4.2.

If U is a normed space, the the dual space $U^* = \mathcal{B}(U, \mathbb{R})$. Consists of all linear and bonded functionals mapping from U to $\mathbb{R}$.

Theorem 4.2.

Let be U a Banach space, then the dual U^* is also a Banach space relative to the norm of the functionals defined by

$$\|u^*\| = \sup_{\|u\|_U \leq 1} |u^*(u)|$$

Example 4.2.

Let $\Omega \subset \mathbb{R}$ be a measurable set. Let $f \in L^p(\Omega)$. Consider the functional $\phi_g : L^p(\Omega) :\to \mathbb{R}$ defined by,

$$\phi_g(f) = \int_\Omega fg d\mu$$

characterized for some g mapping Ω to the real line. This is a linear functional with respect $L^p(\Omega)$. We want an estimate of the norm of this functional. For this purpose we apply Hölder inequality, with $\frac{1}{p} + \frac{1}{q} = 1$, and $p, q > 1$,

$$\begin{aligned}
\|\phi_g\| &= \sup_{1 \geq \|f\|_{L^p(\Omega)}} \left| \int_\Omega gf d\mu \right| \\
&\leq \sup_{1 \geq \|f\|_{L^p(\Omega)}} \int_\Omega |gf| \, d\mu \\
&\text{By Hölder inequality} \\
&\leq \sup_{1 \geq \|f\|_{L^p(\Omega)}} \left(\int_\Omega |f|^p \, d\mu \right)^{\frac{1}{p}} \left(\int_\Omega |g|^q \, d\mu \right)^{\frac{1}{q}} \\
&= \left(\int_\Omega |g|^q \, d\mu \right)^{\frac{1}{q}} \sup_{1 \geq \|f\|_{L^p(\Omega)}} \left(\int_\Omega |f|^p \, d\mu \right)^{\frac{1}{p}} \\
&= \left(\int_\Omega |g|^q \, d\mu \right)^{\frac{1}{q}} = \|g\|_{L^q(\Omega)}
\end{aligned}$$

This result implies that if $g \in L^q(\Omega)$, then ϕ_g is bounded, hence for all $g \in L^q(\Omega)$ we have that the functionals characterized by g, $\phi_g \in (L^p(\Omega))^*$. It's possible to demonstrate that all $\phi \in (L^p(\Omega))^*$ can be characterized by some g in $L^q(\Omega)$. Thus,

$$L^q(\Omega) = (L^p(\Omega))^*$$

Remark 4.2.

There is a natural duality between U and U^* determined by the bilinear functional $(\cdot \mid \cdot) : U \times U^* \to \mathbb{R}$, defined as

$$(u \mid u^*) = u^*(u), \quad \forall u \in U, \forall u^* \in U^*$$

Definition 4.3.

A subset $A \in X \times X^*$ is called monotone if $(x_1 - x_2 \mid y_1 - y_2) \geq 0$, for any $(x_1, y_1), (x_2, y_2) \in A$. A monotone subset of $X \times X^*$ is said to be maximal monotone if it is not properly contained in any other monotone subset of $X \times X^*$.

If Λ is a single-valued operator from X to X^*, then the monotonicity condition becomes,

$$(x_1 - x_2 \mid Ax_1 - Ax_2) \geq 0, \quad \forall x_1, x_2 \in \operatorname{dom} A$$

Definition 4.4 (locally bounded.).

A subset $A \subset X \times X^*$ is said to be locally bounded at $x_0 \in X$ if there exists a neighborhood V of x_0 such that $A(V) = \{Ax;\ x \in (\operatorname{dom} A) \cap V\}$ is a bounded subset of X^*. The operator $A : X \to X^*$ is said to be bounded if it maps every bounded subset of X into a bounded set of X^*.

Remark 4.3.

If X and X^* are two vector spaces in duality, we can associate with a cone C of X its polar cone C^*

$$C^* = \{p^* \in X^* \mid (p \mid p^*) \geq 0,\ \forall p \in C\}$$

Since C^* is a pointed cone with vertex O in X^*, it defines a partial ordering relation denoted by $\preceq$ or $\succeq$:

$$p^* \preceq q^* \iff q^* - p^* \in C^*$$

Hence C is evidently the cone of positive elements in X^*.
If, X and X^* are two dual topological vector spaces and if C is a pointed closed convex cone with vertex O, then from the above we have the following property:

$$p \in C \iff p \succeq 0 \iff (p \mid p^*) \leq 0, \qquad \forall p^* \in C^*$$

Indeed $C^{**} = C$, C^{**} being the polar cone of C^* and so $p \in C \iff p \in C^{**} \iff (p \mid p^*) \geq 0$, $\forall p^* \in C^*$.

Definition 4.5.

A sequence $(u_n)_n$ in a Banach space is called weakly convergent to some $u \in U$ if for all linears continuous functionals $u^* \in U^*$ we have

$$\lim_{n\to\infty} u^*(u_n) = u^*(u)$$

u is also called the weak-limit and we write $u_n \xrightharpoonup[n\to\infty]{} u$.

Theorem 4.3.

A sequence $(u_n)_n$ in U converges to $u \in U$ if and only if $\sup_{n\in\mathbb{N}} \|u_n\| < \infty$ and $u_n \xrightharpoonup[n\to\infty]{} u$

Theorem 4.4 (Bourbaki-Alaoglu-Katulami).

The closed unit ball in a Banach space U is weakly compact if and only if U is reflexive. If U is in an addition separable, then it's weakly sequentially compact.

Definition 4.6.

Let U be a Banach space and $J : U \to \mathbb{R}$, J is called weakly (sequentially) lower semi-continuous at point u_0 if for every weakly convergent sequence $(u_n)_n$ converges to u_0, i.e. $u_n \rightharpoonup u_0$, it holds

$$J(u) \leq \liminf_{n\to\infty} J(u_n)$$

Definition 4.7.

A non empty set $C \subset U$ is called weakly closed if for every weakly convergent sequence $(u_n)_n$ in C follows that the weak limit belongs to C. i.e. $u_n \rightharpoonup u$, with $u_n \in C$, implies $u \in C$.

Definition 4.8.

A non empty set $C \subset U$ is called weakly sequentially compact if for every sequence in C contains a weakly convergent subsequence whose limit belongs to C.

Theorem 4.5.

Let U be a Banach space and $J : U \to \overline{\mathbb{R}}$ the the following conditions are equivalent:

- *J is weakly lower semi-continuous on U for all $u \in U$.*
- *The level sets $\mu_\xi = \{u \in U | J(u) \leq \xi\}$ is weakly closed for each $\xi \in \mathbb{R}$.*

Lemma 4.2.

Let be $J : U \to \overline{\mathbb{R}}$ a convex and lower semicontinuous functional. Assume there is $u_0 \in U$ such that $J(u_0) = -\infty$, then J is nowhere finite.

Proof. Assume that there is $v \in U$ such that $-\infty < J(v) < \infty$. Then by convexity $J(\lambda u_0 + (1-\lambda)v) = -\infty$, $\forall \lambda \in [0,1]$. Because J is lower semicontinuos it follows that in the limit $\lambda \to 0$,

$$(\lambda u_0 + (1-\lambda)v) \to v \implies J(v) \leq J(\lambda u_0 + (1-\lambda)v) = -\infty$$

□

Lemma 4.3.
Every lower semi-continuous and convex function on a linear space U is weakly lower semi-continuous.

Corollary 4.1.
Assume that U is a reflexive Banach space, then every bounded sequence $(u_n)_n \in U$ that is $\sup_{n\in\mathbb{N}} \|u_n\| < \infty$ has a subsequence $(u_k)_k$ which is weakly convergent to some $u \in U$.

Remark 4.4.
Since every Hilbert space is reflexive the corollary applies to this case.

Lemma 4.4.
A closed set C is weakly closed if and only if the set is convex.

Definition 4.9.
Let U be a real linear space and $J : U \to \overline{\mathbb{R}}$. We said that J is sublinear if:

$$\begin{aligned} J(\lambda u) &= \lambda J(u) \qquad \forall u \in U, \text{ and } \mathbb{R} \ni \lambda > 0 \\ J(u+v) &\leq J(u) + J(v) \qquad \forall u, v \in U \end{aligned}$$

Remark 4.5.
Every sublinear function is convex.

Theorem 4.6.
Let U be a real linear space $J : U \to \overline{\mathbb{R}}$ a sublinear functional. Then there is a linear functional f on U such that,

$$f(u) \leq J(u) \quad \forall u \in U$$

Remark 4.6.
Let $J : U \to \overline{\mathbb{R}}$, from definition 4.4, we have J is locally bounded around u_0 if $\exists V \subset U$ neighborhood of u_0 such that for some $M \in \mathbb{R}$,

$$|J(u)| < M \qquad \forall u \in V.$$

Lemma 4.5.
Let U be a Banach space, and let $J : U \to \mathbb{R}$ convex. If J is locally bounded around u, then J is lower semi-continuous in u.

Proof. Let $u_k \to u$ as $k \to \infty$. For each $\epsilon > 0$ we can find a sequence α_k such that $\left\| \frac{u-u_k}{\alpha_k} \right\| < \epsilon$, and $\alpha_k \to 0$ as $k \to \infty$. Moreover, for k sufficiently large we have $\|u - u_k\| < \epsilon$. Choose ϵ such that J is bounded in $\overline{B_{2\epsilon}(u)}$ by M and define $v_k = u_k + \frac{u-u_k}{\alpha_k} \in \overline{B_{2\epsilon}(u)}$, since $\|v_k - u\| \leq \|u_k - u\| + \left\| \frac{u-u_k}{\alpha_k} \right\| \leq 2\epsilon$.
Since J is convex

$$J(u) \leq \alpha_k J(v_k) + (1-\alpha_k) J(u_k) \leq \alpha_k M + J(u_k)$$

.
Since $\alpha_k \to 0$, then

$$J(u) \leq \liminf_{k\to\infty} (\alpha_k M + J(u_k)) = \liminf_{k\to\infty} J(u_k)$$

Thus is if J is convex and locally bounded around u, then is lower semi-continuous around u. □

Remark 4.7.
The result that convexity and local boundedness imply lower semi-continuity is similar to classical result for linear operators where local boundedness implies continuity. In general convexity plays in optimization the same role as linearity in solving equations.

5. Lecture 5

5.1. Subgradients

Proposition 5.1.
Let $J : U \to (-\infty, \infty]$ be proper and convex, let $x \in \operatorname{dom} J$, and let $y \in U$. Then the following hold:

1. *Let $\phi : \mathbb{R}_+ \to (-\infty, \infty]$, such that $\phi(\alpha) := \frac{J(x+\alpha y)-J(x)}{\alpha}$. The function ϕ is increasing.*
2. *$J'(x; y)$ exists in $[-\infty, \infty]$ and*
$$J'(x;y) = \inf_{\alpha\in\mathbb{R}_+} \frac{J(x+\alpha y) - J(x)}{\alpha}.$$
3. *$J'(x; y - x) + J(x) \leq J(y)$.*

Proof.

1. Fix α and β in $\mathbb{R}_+$ such that $\alpha < \beta$, and set $\lambda = \alpha/\beta$ and $z = x + \beta y$. If $J(z) = \infty$, then certainly $\phi(\alpha) \leq \phi(\beta) = \infty$. Consider $J(x) < \infty$, since J is convex,
$$J(x + \alpha y) = J(\lambda z + (1 - \lambda)x) \leq \lambda J(z) + (1 - \lambda)J(x) = J(x) + \lambda(J(z) - J(x)).$$
Then, substituting λ and z
$$\begin{aligned} J(x + \alpha y) - J(x) &\leq \frac{\alpha}{\beta}(J(x + \beta y) - J(x)) \\ \frac{J(x + \alpha y) - J(x)}{\alpha} &\leq \frac{J(x + \beta y) - J(x)}{\beta} \\ \phi(\alpha) &\leq \phi(\beta) \end{aligned}$$
Hence, if $\alpha < \beta \implies \phi(\alpha) \leq \phi(\beta)$, i.e. ϕ is increasing.
2. Since ϕ is increasing,
$$\inf_{\alpha\in\mathbb{R}_+} \phi(\alpha) = \inf_{\alpha\in\mathbb{R}_+} \frac{J(x + \alpha y) - J(x)}{\alpha} = \lim_{\alpha\downarrow 0} \frac{J(x + \alpha y) - J(x)}{\alpha} = J'(x; y)$$
3. If $y \notin \operatorname{dom} J$ the inequality holds. Consider $y \in \operatorname{dom} J$, and we invoke convexity of J for $\alpha \in (0, 1)$,
$$\begin{aligned} J((1 - \alpha)x + \alpha y) &\leq (1 - \alpha)J(x) + \alpha J(y) \\ J((x + \alpha(y - x)) &\leq J(x) + \alpha(J(y) - J(x)) \\ J((x + \alpha(y - x)) &\leq J(x) + \alpha(J(y) - J(x)) \\ \frac{J((x + \alpha(y - x)) - J(x)}{\alpha} &\leq J(y) - J(x) \end{aligned}$$
Taking $\alpha \downarrow 0$, we have
$$J'(x; y - x) \leq J(y) - J(x)$$

$\square$

Remark 5.1.

Let $J : U \to (-\infty, \infty]$ be convex, let $x \in U$, and suppose that J is Gâteaux differentiable at x. The above result is equivalent to say $\forall y \in U \ (y - x \mid \nabla J(x))_{UU^*} + J(x) \leq J(y)$.

Proposition 5.2.

Let $J : U \to (-\infty, \infty]$ be proper. Suppose that $\operatorname{dom} J$ *is open and convex, and that J is Gâteaux differentiable on* $\operatorname{dom} J$ *. Then the following are equivalent:*

1. *J is convex.*
2. $\forall x, y \in \operatorname{dom} J, \quad (x - y \mid \nabla J(x))_{UU^*} \leq J(x) - J(y)$.
3. $\forall x, y \in \operatorname{dom} J, \quad 0 \leq (x - y \mid \nabla J(x) - \nabla J(y))_{UU^*}$, *i.e.* $\nabla J(x)$ *is monotone.*

Proof. Let $x, y \in \operatorname{dom} J$, and $z \in U$. Since $\operatorname{dom} J$ is open, $\exists \epsilon > 0$, such that $x + \epsilon(x - y) \in \operatorname{dom} J$ and $y + \epsilon(y - x) \in \operatorname{dom} J$. Set $C = (-\epsilon, 1 + \epsilon)$ and the function $\phi : \mathbb{R} \to (-\infty, \infty)$

$$\phi(\alpha) = J(y + \alpha(x - y)) + I_C(\alpha).$$

Where I_C is the indicator function. Then ϕ is Gateâux differentiable on C and $\forall \alpha \in C$,

$$\phi'(\alpha) = (x - y \mid \nabla J(y + \alpha(x - y)))_{UU^*}$$

- (1) $\implies$ (2). Since J is proper and convex, assumption number (1), by preposition (5.1) part (3), we have that
$$\forall y \in \operatorname{dom} J \quad (y - x \mid \nabla J(x))_{UU^*} + J(x) \leq J(y)$$
- (2) $\implies$ (3). Since assumption (2) holds $y, x \in \operatorname{dom} J$,
$$(y - x \mid \nabla J(x))_{UU^*} + J(x) \leq J(y)$$
and
$$(x - y \mid \nabla J(y))_{UU^*} + J(y) \leq J(x).$$
Adding both inequalities, we have
$$0 \leq (x - y \mid \nabla J(x) - \nabla J(y))_{UU^*}$$
- (3) $\implies$ (1). Take α and β in C, such that $\alpha < \beta$, and set $y_\alpha = y + \alpha(x - y)$ and $y_\beta = y + \beta(x - y)$. Then the assumption that for all $x, y \in \operatorname{dom} J$, $(x - y \mid \nabla J(x) - \nabla J(y))_{UU^*} \geq 0$ implies that
$$\begin{aligned} 0 &\leq (y_\beta - y_\alpha \mid \nabla J(y_\beta) - \nabla J(y_\alpha))_{UU^*} \\ &= (\beta(x - y) - \alpha(x - y) \mid \nabla J(y_\beta) - \nabla J(y_\alpha))_{UU^*} \\ &= (\beta - \alpha)((x - y) \mid \nabla J(y_\beta) - \nabla J(y_\alpha))_{UU^*} \\ &= (\beta - \alpha)\left(\phi'(\beta) - \phi'(\alpha)\right) \end{aligned}$$
Therefore we have that, if $\beta > \alpha$ and assumption (3) holds, ϕ' is increasing on C, by proposition (0.2), we have that ϕ is convex. Therefore,
$$J(\alpha y + (1 - \alpha)y) = \phi(\alpha) \leq \alpha\phi(1) + (1 - \alpha)\phi(0) = \alpha J(x) + (1 - \alpha)J(y)$$
Hence J is convex.

$\square$

Theorem 5.1.

Let $J : U \to (-\infty, \infty]$ be proper. Suppose that $\operatorname{dom} J$ *is open and convex, and that J is twice Gâteaux differentiable on* $\operatorname{dom} J$ *. Then,*

$$(\forall x \in \operatorname{dom} J)(\forall y \in \operatorname{dom} J) \quad (z \mid \nabla^2 f(x) z)_{UU^*} \geq 0 \iff J \text{ is convex}$$

Proof. First assume J convex. Let $z \in U$. Since $\operatorname{dom} J$ is open, $\exists \alpha > 0$ small enough, such that $x + \alpha z \in \operatorname{dom} J$, and by proposition (5.1) part (3),

$$0 \leq (z \mid \nabla J(x + \alpha z) - \nabla J(x))_{UU^*} = \frac{1}{\alpha}((x + \alpha z) - x \mid \nabla J(x + \alpha z) - \nabla J(x))_{UU^*},$$

letting $\alpha \downarrow 0$ we obtain

$$0 \leq (z \mid \nabla^2 J(x) z)_{UU^*} = (\mathsf{D}^2 J(x) z) z.$$

Now assume that ϕ is twice Gâteaux differentiable on $\operatorname{dom} J$ with $\forall \alpha \in C$, given by

$$\phi''(\alpha) = (x - y \mid \nabla^2 J(y + \alpha(x - y))(x - y))_{UU^*}.$$

Hence ϕ' is increasing and by proposition (0.2), ϕ is convex. Therefore J is convex. $\square$

Definition 5.1.

Let U be a Banach space and let $J : U \to (-\infty, \infty]$ be a convex and proper function. The subdifferential at a point $u \in \operatorname{dom} J$ is a mapping,

$$\partial J : U \to 2^{U^*}, \qquad \partial J(u) := \{p^* \in U^* \mid J(v) \geq J(u) + p^*(v - u), \ \forall v \in U\}$$

The elements of $p^* \in \partial J(u)$ are called subgradients of J at u.

Example 5.1.

Consider $J : \mathbb{R} \to \mathbb{R}$, $u \to |u|$ which is not differentiable at $u = 0$. If $u > 0$, then $J(u) = u$ and we can find $0 < v < u < w$. Then $p^* \in \partial J(u)$ implies by definition of subdifferential

$$\begin{aligned} v - u \geq p^*(v - u) &\equiv (1 - p^*)(u - v) \leq 0 \\ w - u \geq p^*(w - u) &\equiv (1 - p^*)(w - u) \geq 0. \end{aligned}$$

which implies for $u > 0$, $p^* \leq 1 \leq p^*$, then $p^* = 1$.
In the same way we obtain for $u < 0$, $p^* \geq -1 \geq p^*$. In the case $u = 0$, we need to satisfy $|v| \geq p^* v$, which is fulfilled if and only if $|p^*| \leq 1$. Hence for $J(u) = |u|$,

$$\partial |u| = \begin{cases} \{1\}, & u > 0 \\ [-1, 1], & u = 0 \\ \{-1\}, & u < 0 \end{cases}.$$

Example 5.2.

A convex function which is not subdifferentiable everywhere $J : \mathbb{R} \to \mathbb{R}$,

$$J(u) = \begin{cases} -\sqrt{1 - |u|^2} & |u| \leq 1 \\ \infty & \text{otherwise} \end{cases}$$

For $|u| \geq 1$, we have $\partial J(u) = \emptyset$.

Example 5.3.
Let C be a convex and closed subset of U and I_C function defined by

$$I_C(u) = \begin{cases} 0 & u \in C \\ \infty & \text{otherwise} \end{cases}$$

The subdifferentiable is the definition of normal cone at u

$$\partial I_C(u) = \{u^* \in U^* \mid u^*(u - v) \geq \forall v \in C\} = \mathcal{N}_C(u)$$

.

Theorem 5.2.
Let U be a Banach space. And $J : U \to \overline{\mathbb{R}}$ a subdifferentiable function. Then $\partial J(u)$ is convex and weakly closed.

Remark 5.2.
Most of the rules for derivates also hold for subdifferentials with some additional assumptions,

- $J : U \to \overline{\mathbb{R}}$, $\lambda > 0$, $\partial J(\lambda u) = \lambda J(u)$.
- $\partial(J + F)(u) \supseteq \partial J(u) + \partial F(u)$.

Theorem 5.3 (Moreau-Rockafellar).
Let U be a Banach space and $J_i : U \to \mathbb{R}$ proper and convex functions for $i = 1, \ldots, n$. The sum-rule

$$\partial(J_1 + \cdots + J_n)(u) = \partial J_1(u) \ldots \partial J_n(u), \qquad n \geq 2$$

holds if there exists $u_0 \in U$ such that all $J_i(u_0)$ are finite and all J_i except at most one J_k, $k \in \{1, 2, \ldots n\}$ are continuous at u_0

6. Lecture 6

Theorem 6.1.
Let V, U, Banach Spaces. Let $J : V \to \overline{\mathbb{R}}$ a convex functional. And consider the mapping $A : U \to V$ linear and continuous with the adjoint $A^* : U^* \to V^*$. Moreover, J is lower semi-continuous and let $A\overline{u}$ be a point where J is continuous and finite. Then te compose function $J \circ A : U \to \overline{\mathbb{R}}$ is subdifferentiable for all $u \in V$ and,

$$\partial(J \circ A)(u) = A^* (\partial J(Au))$$

Proof. Let $p^* \in \partial J(Au)$,

$$J(p) \geq J(Au) + p^*(p - Au) \quad \forall p \in V$$

where $p = Av$ with $v \in U$,

$$\begin{aligned} (J \circ A)(v) &\geq (J \circ A)(u) + p^*(A(v - u)) \quad \forall v \in U \\ &= (J \circ A)(u) + A^* p^*(v - u) \quad \forall v \in U \end{aligned}$$

i.e. $A^* p^* \in \partial(J \circ A)(u) \implies A^* \partial J(Au) \subseteq \partial(J \circ A)(u)$. Proof based again on the weak separation theorem of convex sets. (We have to check Bauschke) $\square$

Theorem 6.2.
If $J : U \to \overline{\mathbb{R}}$ is convex and Frechét-differentiable at $u \in U$, then $\partial J(u) = \{J'(u)\}$

Proof. Let $p^* \in \partial J(u)$. Then for each $t > 0$, $J(u+tv) - J(u) \geq p^*(tv) = tp^*(v)$, diving by t and takin the limit $t \to 0$ we obtain,

$$\begin{aligned} J'(u;v) &\geq p^*(v) \quad \forall v \in U \\ \Longrightarrow (\nabla J(u) - p^*)(v) &\geq 0 \quad \forall v \in U. \end{aligned}$$

Since $J(u)$ is Frechét differentiable the operator $J'(u;v)$ is linear with respect to v and $p^* \in U^*$ implies $(\nabla J(u) - p^*)$ is linear, taking $-v \in U$, we obtain that $(\nabla J(u) - p^*)(v) \leq 0$. Therefore $p^* = \nabla J(u)$.
On the other hand, if J is differentiable, it follows that $\nabla J(u) \in \partial J(u)$. For $v \in U$, we set $w = v - u$, $u \in U$ we have by (5.2),

$$\begin{aligned} J(u+w) - J(u) &\geq J'(u;w) = (w \mid \nabla J(u))_{UU^*} \\ \Longrightarrow \quad J(v) - J(u) &\geq J'(u; v-u) = (v - u \mid \nabla J(u))_{UU^*} \end{aligned}$$

Since the above inequality holds for all $v \in U$ implies $\nabla J(u) \in \partial J(u)$. $\square$

Remark 6.1.

The subgradient can be used to obtain local optimality conditions that are necessary and sufficient for convex problem.

Theorem 6.3.

Let U be a Banach Space and $J : U \to \mathbb{R}$ convex and proper. Then each local minimum is global minimum. Moreover $\overline{u} \in U$ is a minimizer if and only if $0 \in \partial J(\overline{u})$.

Proof. If $0 \in \partial J(\overline{u})$: $J(v) \geq J(\overline{u}) + (0)(v - \overline{u}) = J(\overline{u})$, $\forall v \in U$, and hence $\overline{u}$ is a global minimizer.
Assume that $0 \notin J(\overline{u})$, then $\exists v \in U$, such that

$$J(v) < J(\overline{u}) + (0)(v - \overline{u}) = J(\overline{u}).$$

Therefore $\overline{u}$ cannot be a minimizer. $\square$

Definition 6.1 (Duality).

Let $J : U \to \overline{\mathbb{R}}$, and U a Banach space. Then the convex conjugate function $J^* : U^* \to \mathbb{R}$ is defined by

$$J^*(p^*) = \sup_{u \in U} \{p^*(u) - J(u)\}$$

implies that $-\sup_{u \in U} \{p^*(u) - J(u)\} = -J^*(p^*) = \inf_{u \in U} \{J(u) - p^*(u)\}$.

Example 6.1.

Consider the indicator function of a convex set C, $I_C : U \to \overline{\mathbb{R}}$

$$I_C(u) = \begin{cases} 0 & u \in C \\ \infty & \text{otherwise.} \end{cases}$$

Then we have that the convex conjugate is given by

$$J^*(p^*) = \sup_{u \in U} \{p^*(u) - I_C(u)\} = \sup_{u \in C} \{p^*(u)\}.$$

Example 6.2.
$J : \mathbb{R}_+ \to \mathbb{R}$, such that

$$J(u) = \begin{cases} u(\ln u - 1) & u > 0 \\ 0 & u = 0 \end{cases}$$

Then $J^*(p^*) = \sup_{u>0}\{p^* u - u(\ln u - 1)\}$, then for $u > 0$,

$$f'(u) = p^* - (\ln u - 1) - 1 = p^* - \ln(u) \implies \overline{u} = \exp p^*$$
$$f'(u) = p^* - (\ln u - 1) - 1 = p^* - \ln(u) \implies \overline{u} = \exp p^*$$

Example 6.3.
Let $J : \mathbb{R} \to \mathbb{R}$, such that $J(u) = \exp u$, then $J^*(p^*) = \sup_{u\in\mathbb{R}}\{p^* u - \exp u\}$. Let $f(u) = p^* u - \exp(u)$, therefore $f'(u) = p^* - \exp u$, $\forall u \in \mathbb{R}$. Which is zero for $\overline{u} = \ln p^*$, if $p^* > 0$. Since $f''(u) < 0$, then $\overline{u}$ is indeed maximum. And we see that $\lim_{u\to\pm\infty} f(u) = -\infty$. If $p^* = 0$, $f(u) = -\exp u < 0$ and therefore the $\sup_{u\in\mathbb{R}} f(u) = 0$ (Consider the limit when $u \to -\infty$). Then we have,

$$J^*(p^*) = \begin{cases} p^*(\ln p^* - 1) & p^* > 0 \\ 0 & p^* = 0 \end{cases}$$

Example 6.4.
Let H be a Hilbert space and $J(u) = \frac{1}{2}\|u\|^2$. Since H is Hilbert, by Riesz, for each linear and bounded functional $\phi_{p^*} \in H$, $\exists p^* \in H$ such that, $\phi_{p^*}(u) = \langle u, p^*\rangle$. Using the definition of conjugate function,

$$\begin{aligned} J^*(p^*) &= \sup_{u\in U}\left\{\langle u, p^*\rangle - \frac{1}{2}\|u\|^2\right\} \\ &= -\inf_{u\in U}\left\{\frac{1}{2}\|u\|^2 - \langle u, p^*\rangle\right\} \end{aligned}$$

Note that,

$$\frac{1}{2}\|u - p^*\|^2 = \frac{1}{2}\|u\|^2 - \langle u, p^*\rangle + \frac{1}{2}\|p^*\|^2$$

Therefore we can substitute in the above equation to find an equivalent form to the conjugate function,

$$\begin{aligned} J^*(p^*) &= -\inf_{u\in U}\left\{\frac{1}{2}\left(\|u - p^*\|^2 - \|p^*\|^2\right)\right\} \\ &= -\frac{1}{2}\inf_{u\in U}\left\{\|u - p^*\|^2\right\} + \frac{1}{2}\|p^*\|^2 \end{aligned}$$

We have $\|u - p^*\| \geq 0$, $\forall u \in H$, then,

$$\inf_{u\in U}\left\{\|u - p^*\|^2\right\} = 0,$$

since we can take $u = p^*$. Hence,

$$J^*(p^*) = \frac{1}{2}\|p^*\|^2 \tag{2}$$

Theorem 6.4.
Let U be a Banach space and $J : U \to \overline{\mathbb{R}}$. Then J^* is convex.

Proof. Let $p^*, q^* \in U^*$, and $\lambda \in [0,1]$,

$$\begin{aligned}
J^*(\lambda p^* + (1-\lambda)q^*) &= \sup_{u\in U}\{(\lambda p^* + (1-\lambda)q^*)(u) - J(u)\} \\
&= \sup_{u\in U}\{\lambda p^*(u) - \lambda J(u) + (1-\lambda)q^*(u) - (1-\lambda)J(u)\} \\
&\leq \sup_{v,u\in U}\{\lambda p^*(u) - \lambda J(u) + (1-\lambda)q^*(v) - (1-\lambda)J(v)\} \\
&= \sup_{u\in U}\{\lambda p^*(u) - \lambda J(u)\} + \sup_{v\in U}\{(1-\lambda)q^*(v) - (1-\lambda)J(v)\} \\
&= \lambda J^*(p^*) + (1-\lambda)J^*(q^*).
\end{aligned}$$

Hence J^* is convex. □

7. Lecture 7

Remark 7.1.
Some elementary properties of conjugate functions

- **Young inequality** $J(u) + J^*(p^*) \geq p^*(u) \quad \forall u \in U, \forall p^* \in U^*$
- $J^*(0) = \sup_{u\in U}((0,u) - J(u)) = \sup_{u\in U}(-J(u)) = -\inf_{u\in U} J(u)$

In many applications in optimization, is used the equivalent formulation,

$$\inf_{u\in U} J(u) = -J^*(0).$$

$J \leq F \implies J^* \geq F^*$

Theorem 7.1.
Let U a Banach space and $J^* : U^* \to \overline{\mathbb{R}}$ be the conjugate of the $J : U \to \overline{\mathbb{R}}$. Then for all $u \in U$.

$$p^* \in \partial J(u) \iff J(u) + J^*(p^*) = p^*(u)$$

.

Proof. Let $p^* \in \partial J(u)$, then for all $v \in U$ we have,

$$\begin{aligned}
J(v) &\geq J(u) + p^*(v) - p^*(u) \\
p^*(u) - J(u) &\geq p^*(v) - J(v) \\
p^*(u) - J(u) &\geq \sup_{v\in U}\{p^*(v) - J(v)\} \\
p^*(u) - J(u) &\geq J^*(p^*) \\
p^*(u) &\geq J^*(p^*) + J(u)
\end{aligned}$$

In the other hand (Young inequality), by definition of J^*,

$$\begin{aligned}
J^*(p^*) &\geq p^*(u) - J(u) \\
J^*(p^*) + J(u) &\geq p^*(u)
\end{aligned}$$

Therefore we have that $p^*(u) = J(u) + J^*(p^*)$.

Now assume $p^* \notin \partial J(u)$, then $\exists v \in U$ such that,

$$J(v) < J(u) + p^*(v - u)$$

Rearranging the inequality we get,

$$p^*(u) - J(u) < p^*(v) - J(v) \leq J^*(p^*)$$

Which implies that we have the inequality strictly and the equality cannot hold.

$$p^*(u) < J^*(p^*) + J(u)$$

□

Corollary 7.1.
It follows from previous theorem that $\partial J(u) = \{p^* \in U^* | J(u) + J^*(p^*) = (p^*, u)\}$.

Theorem 7.2.
Let U *be a Banach space and* $J : U \to \mathbb{R}$ *be proper function. If* $p^* \in \partial J(u)$ *then* $u \in \partial J^*(p^*)$

Proof. Let $p^* \in \partial J(u)$. For any $g^* \in U^*$, it follows

$$J^*(g^*) = \sup_{v \in U} \{g^*(v) - J(v)\} \geq g^*(u) - J(u) \geq g^*(u) - J(u)$$

From theorem 7.1

$$J^*(g^*) \leq g^*(u) - p^*(u) + J^*(p^*) = (g^* - p^*)(u) + J^*(p^*) \implies u \in \partial J^*(p^*).$$

□

By iteration the definition, we obtain the bipolar function $(J^*)^* = J^{**} : U^{**} \to \overline{\mathbb{R}}$,

$$J^{**}(u) = \sup_{p^* \in U^*} \{p^*(u) - J^*(p^*)\}$$

Theorem 7.3 (Convex envelope theorem.).
Let U *be a reflexive Banach space. The* J^{**} *is the maximum convex functional below* J *(also called convex envelope). That is, if* F *is convex and* $F(u) \leq J(u), \forall u \in U$*. Then,* $J^{**}(u) \leq J(u)$*, and* $F(u) \leq J^{**}(u), \forall u \in U$*. In particular* $J^{**} = J$ *if and only if* J *is convex.*

Proof. Let $\phi_u \in U^{**}$. Since U is reflexive every $\phi_u \in U^* *$ can be related with a member $u \in U$, defining $\phi_u(p^*) = p^*(u)$, we proceed

$$\begin{aligned}
J^{**}(u) &= \sup_{p^* \in U^*} \{\phi_u(p^*) - J^*(p^*)\} \\
&= \sup_{p^* \in U^*} \{p^*(u) - J^*(p^*)\} \\
&= \sup_{p^* \in U^*} \left\{p^*(u) - \sup_{v \in U} \{p^*(v) - J(v)\}\right\} \\
&= \sup_{p^* \in U^*} \left\{p^*(u) + \inf_{v \in U} \{J(v) - p^*(v)\}\right\} \\
&= \sup_{p^* \in U^*} \left\{\inf_{v \in U} \{p^*(u) + J(v) - p^*(v)\}\right\} \\
&= \sup_{p^* \in U^*} \left\{\inf_{v \in U} \{p^*(u - v) + J(v)\}\right\}
\end{aligned}$$

Taking $v = u$ in the expression and comparing it with its infimum the inequality holds,

$$\inf_{v \in U} \{p^* (u - v) + J(v)\} \leq p^*(u - u) + J(u)$$
$$\inf_{v \in U} \{p^* (u - v) + J(v)\} \leq J(u)$$

We have that $J^{**}(u) \leq J(u)$.

$$\sup_{p^* \in U^*} \inf_{v \in U} \{p^* (u - v) + J(v)\} \leq J(u)$$
$$J^{**}(u) \leq J(u)$$

Now we assume that F is a convex functional and $g^* \in \partial F(u)$ for $u \in U$.

$$\begin{aligned} \implies F(v) &\geq F(v) + q^*(v - u) \\ F^{**}(u) &= \sup_{p^* \in U^*} \inf_{v \in U} \{p^*(u - v) + F(u) + q^*(v - u)\} \\ &\geq \sup_{p^* \in U^*} \inf_{v \in U} \{(p^* - q^*)(u - v) + F(u)\} \\ &\geq \inf_{v \in U} \{(q^* - q^*)(u - v) + F(u)\} \\ &= F(u) \end{aligned}$$

If F is convex,

$$\implies F(u) \leq F^{**}(u) \leq F(u) \implies F(u) = F^{**}(u),$$

Hence, we have

$$F(u) = F^{**}(u) = \sup_{p^* \in U^*} \inf_{v \in U} \{p^*(u - v) + F(v)\} \leq J^{**}(u).$$

□

8. Lecture 8

Definition 8.1.

Let U and Y Banach spaces and $J : U \to \overline{\mathbb{R}}$ is a proper function. We consider an optimization problem (P) called primal problem

$$\inf_{u \in U} J(u) \tag{P}$$

Then the problem is said to be nontrivial if there is $\overline{u} \in U$ such that $J(\overline{u}) < \infty$. A function $\Phi : U \times Y \to \overline{\mathbb{R}}$ is said to be a perturbation function of J,

$$\inf_{u \in U} \Phi(u, p) \tag{Pp}$$

if $\Phi(u, 0) = J(u)$ for all $u \in U$. For each $p \in Y$, the minimization problem (Pp) is called a perturbation problem. The variable p is called perturbation parameter. If we denote by Φ^* the convex conjugate function of Φ, the *dual problem*, with respect to Φ is defined by

$$\sup_{p^* \in Y^*} -\Phi^*(0, p^*) \tag{P*}$$

where $\Phi^* : (U \times Y)^* \cong U^* \times Y^* \to \overline{\mathbb{R}}$, a function defined as follows.

$$\Phi^* (u^*, p^*) = \sup_{\substack{u \in U \\ p \in Y}} \{u^*(u) + p^*(p) - \Phi(u, p)\}$$

Remark 8.1.
For $p = 0$, (P*) $\equiv$ (Pp). We denote the infimum for problem (P) by inf (P) and the supremum for problem (P*) by sup (P*)

Lemma 8.1 (Weak duality).
For the problem (P) and (P) it holds that*

$$-\infty \leq \sup\,(\text{P*}) \leq \inf\,(\text{P}) \leq \infty.$$

Proof. Let $p^* \in Y^*$. It follows

$$\begin{aligned}
-\Phi^*(0, p^*) &= -\sup_{\substack{u\in U\\ p\in Y}} \{0(u) + p^*(p) - \Phi(u,p)\} \\
&= \inf_{\substack{u\in U\\ p\in Y}} \{\Phi(u,p) - p^*(p)\} \\
&\leq \Phi(u,0) - p^*(0) \quad \forall u \in U, p^* \in Y^* \\
\implies \sup_{p^*\in Y^*} \{-\Phi\,(0,p^*)\} &\leq \inf_{u\in U} \Phi(u,0) = \inf(P)
\end{aligned}$$

$\square$

By iteration we can define, a bidual problem

$$-\sup_{u\in U}\left\{-\Phi^{**}(u,0)\right\} = \inf_{u\in U} \Phi^{**}(u,0) \tag{P**}$$

In case the space U is reflexive then $U^{**} = U$.

If the perturbation function $\Phi(u,p)$ is proper, convex and weakly lower semicontinuous. Then $\Phi^{**} = \Phi$. In this case $\Phi(u,0) = \Phi^{**}(u,0)$ i.e (P)$\equiv$ (P**)

Definition 8.2.
Consider the infimal value function

$$h(p) = \inf\,(\text{Pp}) = \inf_{u\in U} \Phi(u,p)$$

The problem (P) is called stable if $h(0)$ is finite and its sub-differentiable in zero is not empty.

Theorem 8.1.
The primal problem (P) is stable if and only if the following conditions are simultaneously satisfied:

- *The dual problem (P*) has a solution.*
- *There is no duality gap, i.e.*

$$\inf\,(\text{P}) = \sup\,(\text{P*}) \leq \infty$$

Theorem 8.2 (Extremal relation).
Let $\Phi : U \times Y \to \overline{\mathbb{R}}$, be convex the the following statements are equivalent:

1. (P) *and* (Pp) *have solutions* $\overline{u}$ *and* $\overline{p^*}$ *and* $\inf(P) = \sup(P^*)$
2. $\Phi(\overline{u}, 0) + \Phi^*\left(0, \overline{p^*}\right) = 0$

3. $\left(0,\overline{p^*}\right) \in \partial\Phi(u,0)$ *and* $(\overline{u},0) \in \partial\Phi^*(0,p^*)$

Proof. We proceed by parts:

1. (1) $\implies$ (2): $\overline{u}$ solution of $\inf$ (P) and $\overline{p^*}$ solution of $\sup$ (P*) and $\inf$ (P) $=$ $\sup$ (P*). This properties implies, $\Phi(\overline{u},0) = \inf (\mathrm{P}) = \sup (\mathrm{P}^*) = -\Phi(0,\overline{p^*}) \implies \Phi(\overline{u},0) + \Phi^*(0,\overline{p^*}) = 0$.

2. (2) $\implies$ (1): $-\Phi^*(0,\overline{p^*}) = \sup (\mathrm{P}^*) \leq \inf (\mathrm{P}) = \Phi(\overline{u},0) = -\Phi^*(0,\overline{p^*}) \implies \sup (\mathrm{P}^*) = \inf (\mathrm{P})$

3. (2) $\iff$ (3): $\Phi(\overline{u},0) + \Phi^*(0,\overline{p^*}) = 0 = (0,\overline{u}) + (\overline{p^*},0) = \left((0,\overline{p^*}),(\overline{u},0)\right) \iff (0,\overline{p^*}) \in \partial\Phi(\overline{u},0)\ \forall u \in U, \forall p^* \in \partial J(u) \iff J(u) + J^*(p^*) = (p^*,u)$

□

Fencel duality.

Consider the functional $J : U \to \overline{\mathbb{R}}$,

$$J(u) = F(u) + G(Au)$$

with $F : U \to \overline{\mathbb{R}}$, G convex function $G : V \to \overline{\mathbb{R}}$ and $A : U \to V$ bounded and linear.

We introduce the perturbation $\Phi(u,p) = F(u) + G(Au - p)$. The dual problem is obtained with,

$$\Phi^*(0,p^*) = \sup_{\substack{u\in U\\ p\in V}} \{p^*(p) - F(u) - G(Au-p)\}$$

For fixed u we set $q : Au - p$.

$$\begin{aligned}
\Phi^*(0,p^*) =& \sup_{u\in U}\sup_{q\in V} \{p^*\left(Au-q\right) - F(u) - G(q)\} \\
=& \sup_{u\in U}\sup_{q\in V} \{p^*\left(Au\right) - p^*(q) - F(u) - G(q)\} \\
=& \sup_{u\in U} \{p^*\left(Au\right) - F(u)\} + \sup_{q\in V} \{(-p^*)(q) - G(q)\} \\
=& \sup_{u\in U} \{\left(A^\star \circ p^*\right)(u) - F(u)\} + \sup_{q\in V} \{(-p^*)(q) - G(q)\} \\
=& F^*(A^\star \circ p^*) + G^*(-p^*)
\end{aligned}$$

Where $(A^\star \circ p^*) \in U^*$, defined as $(A^\star \circ p^*) : U \to \overline{\mathbb{R}}$

$$(A^\star \circ p^*)(u) = p^*(Au)$$

In case U is a Hilbert space $A^\star$ is the adjoint operator of A.

9. Lecture 9

We check the optimality conditions.

$$\begin{aligned}
0 &= \Phi(\overline{u}, 0) + \Phi^*(0, \overline{p^*}) \\
&= F(\overline{u}) + G(A\overline{u}) + F^*(A^\star \overline{p^*}) + G^*(-\overline{p^*}) \\
&= [F(\overline{u}) + F^*(A^\star \overline{p^*}) - A^\star \circ p^*(\overline{u})] + [G(A\overline{u}) + G^*(-\overline{p^*}) - (-p^*)(A\overline{u})]
\end{aligned}$$

Using Young inequality $J(u) + J^*(u^*) - u^*(u) \geq 0$, $\forall u \in U$, and $\forall u^* \in U^*$, we see that both square brackets are nonnegative; and the sum is zero. Then

$$\begin{aligned}
F(\overline{u}) + F^*(A^\star \overline{p^*}) = A^\star \circ p^*(u) &\implies A^\star p^* \in \partial F(\overline{u}) \\
G(A\overline{u}) + G^*(-\overline{p^*}) = (-p^*)(A\overline{u}) &\implies -p^* \in \partial G(A\overline{u})
\end{aligned}$$

F, G are convex and locally bounded, one can show that $\sup$ (P*) $= \inf$ (P).

Example 9.1 (Denoising with bounded variation.).

Let be $u, v \in L^2(\Omega)$. And let be $g : \Omega \to \mathbb{R}^n$, such that, $g \in C_0^\infty(\Omega, \mathbb{R}^n)$. Consider the following functional $J : L^2(\Omega) \to \mathbb{R}$, defined as follows,

$$J(u) = \frac{1}{2}\int_\Omega |u(x) - v(x)|^2 + \alpha \sup_{\|g\|\leq 1} \int_\Omega u \mathrm{div}\,(g) dx$$

Also consider the minimization problem

$$\min_{u \in BV(\Omega)} J(u),$$

restricted to the set of functions with bounded total variations,

$$BV(\Omega) = \left\{ u \in L^1(\Omega) \mid V(u, \Omega) < \infty \right\},$$

where a total bounded variation is defined as,

$$V(u, \Omega) = \sup \left\{ \int_\Omega u \ \mathrm{div}\,(g) dx; \text{ such that } g \in C_0^\infty(\Omega, \mathbb{R}^n), \ \|g\|_\infty \leq 1 \right\}$$

Remark 9.1.

For u smooth enough, it is possible to apply integration by parts, considering the contributions due g has compact support and $\Omega \subset \mathbb{R}^n$, $\int_\Omega u \ \mathrm{div}\, g dx = -\int_\Omega g \cdot \nabla u dx$.

Consider the norm defined on $BV(\Omega)$ as follows,

$$\|u\|_{BV} := \|u\|_{L^1(\Omega)} + V(u, \Omega).$$

If we consider $J(u) = F(u) + G(Au)$, we can set

$$F(u) = \frac{1}{2}\int_\Omega |u(x) - v(x)|^2\, dx = \frac{1}{2}\,\|u - v\|^2_{L^2(\Omega)}$$
$$G(Au) = \alpha\int_\Omega |\nabla u|\, dx$$

Where $A := \alpha\nabla$, and $G(u) = \int_\Omega |u|\, dx$. We introduce the convex functional of each function,

$$F^*(q^*) = \frac{1}{2}\int_\Omega |q^*(x) - v(x)|^2 - \frac{1}{2}v^2(x)dx \qquad \forall q^* \in L^2(\Omega)$$
$$G^*(p^*) = \begin{cases} 0, & \|p^*\| \le 1 \\ -\infty, & \text{otherwise} \end{cases} \qquad \forall p^* \in C_0^\infty(\Omega, \mathbb{R}^n)$$

In order to apply the Fencel duality we see that the , adjoint of A is given by $A^\star = -\alpha(\nabla\cdot)$, thus

$$-J(p^*) = \frac{1}{2}\int_\Omega \left|-\alpha\nabla\cdot p^* + v^2\right|^2 + \frac{1}{2}v^2 dx$$

9.1. Lagrangians

Definition 9.1.
The function $L : U \times Y^* \to \overline{\mathbb{R}}$, $-L(u, p^*) = \sup_{p\in Y}\{p^*(p) - \Phi(u, p)\}$, is called Lagrangian of (P) relative to the perturbation Φ. If we denote by Φ_u for fixed $u \in U$ the function $p \to \Phi(u, p)$, then $-L(u, p^*) = \Phi_u^*(p^*)$

Lemma 9.1.
For all $u \in U$, the function $L_u : Y^* \to \overline{\mathbb{R}}$, $p^* \to L(u, p)$ is a concave function (i.e. $-L_u$ is convex) and weak upper semi-continuous. If Φ is convex then for all $p^* \in Y^*$ the function $L_{p^*} : U \to \overline{\mathbb{R}}$, $u \to L(u, p^*)$ is convex.

Proof. □

Without assuming anything about Φ, we obtain

$$\begin{aligned}\Phi^*(u^*, p^*) &= \sup_{u\in U, p\in Y}\{u^*(u) + p^*(p) - \Phi(u, p)\} \\ &= \sup_{u\in U}\left\{u^*(u) + \sup_{p\in Y}[p^*(p) - \Phi(u, p)]\right\} \\ &= \sup_{u\in U}\{u^*(u) - L(u, p^*)\}\end{aligned}$$

This implies that,

$$(\text{P*})\ \sup_{p^*\in Y^*}\{-\Phi^*(0, p^*)\} = \sup_{p^*\in Y^*}\inf_{u\in U} L(u, p^*)$$

Now we assume that Φ is convex and weak lower semi-continuous, then for $u \in U$, the function $\Phi_u : Y \to \overline{\mathbb{R}}$ is convex and weak lower semi-continuous and thus $\Phi_u^* * = \Phi_u$. Moreover

$$\begin{aligned}\Phi(u, p) &= \Phi_u^{**}(p) \\ &= \sup_{p^*\in Y^*}\{p^*(p) - \Phi_u^*(p)\} \\ &= \sup_{p^*\in Y^*}\{p^*(p) + L(u, p^*)\} \\ &= \sup_{p^*\in Y^*}\{L(u, p^*)\}\end{aligned}$$

Thus,

$$(\text{P}) \quad \inf_{u,p} \Phi(u,p) = \inf_{u \in U} \sup_{p^* \in Y^*} L(u,p^*) \tag{3}$$

Remark 9.2.
The problems (P) and (P*) are related to min-max problem we have that the weak duality means

$$\sup \inf L \leq \inf \sup L$$

Definition 9.2.
An element $(\overline{u}, \overline{p^*}) \in U \times Y^*$ is called saddle point of L if

$$L(\overline{u}, p^*) \leq L(\overline{u}, \overline{p^*}) \leq L(u, p^*), \quad \forall u \in U, \forall p^* \in Y^*.$$

Theorem 9.1.
Assume that Φ convex and weak lower semicontinuous. Then $(u^*, \overline{p^*})$ is a saddle point of L if and only if $\overline{u}$ is solution of (P), $\overline{p^*}$ is solution of (P) and* $\inf(\text{P}) = \sup(\text{P*})$.

Proof. Let $(\overline{u}, \overline{p^*})$ be a saddle point of L. We have that,

$$\left.\begin{array}{r} L(\overline{u}, \overline{p^*}) = \inf_{u \in U} L(u, \overline{p^*}) = -\Phi^*(0, \overline{p^*}) \\ L(\overline{u}, \overline{p^*}) = \sup_{p^* \in Y^*} L(\overline{u}, \overline{p^*}) = -\Phi^*(\overline{u}, 0) \end{array}\right\} \implies \Phi(\overline{u}, 0) + \Phi^*(0, \overline{p}^*) = 0$$

Theorem about extremal conditions $\implies$ $\overline{u}$ is a solution of (P), $\overline{p^*}$ solution of (P*) and

$$\inf(\text{P}) = \sup(\text{P*})$$

"other direction" follows the same argumentation. $\square$

Theorem 9.2 (Saddle point theorem.).
Let $\Phi : U \times Y \to \overline{\mathbb{R}}$ be convex, weak lower semicontinuous and (P) is stable. Then $\overline{u} \in U$ is a solution of (P) if and only if then exist $\overline{p^*} \in Y^*$ such that $(\overline{u}, \overline{p^*})$, is a saddle point of L.

10. Lecture 10

10.1. Minimization

Two approaches for the solution of infinite-dimensional optimization problems.

- **Discretize then Optimize:** this approach consists of a direct discretization of the problem which leads to a non-linear programming problem in a finite subspace $U_n \subset U$. If the discretization is accurate enough, the approximate solution $\overline{u_n} \in U_n$ is close to the real solution $\overline{u} \in U$

- **Optimize then Discretize:** The idea is to formulate the problem in infinite dimensions and to apply discretization only for the solution of sub-problems and for the evluation of the objective function. The main advantage is that quantitative estimates for the convergence of the optimization method can be combined with error estimates of the discretization.

The general framework for minimization algorithms for

$$\min_{u \in U} J(u)$$

with $J : U \to \overline{\mathbb{R}}$, U Banach space delivers a sequence $(u_k)_k \in U$.

Desired Concepts

1. **Global convergence:** We need to measure if a limit is a candidate for a solution of (P).

   ### Definition 10.1 (Stationary measure).
   Let $\Sigma : U \to \mathbb{R}_+$ a functional, given by

   $$\Sigma(u) = \begin{cases} 0 & u \text{ is stationary point.} \\ \alpha(u) \in \mathbb{R}_+ & \text{everywhere else} \end{cases}$$

   We call this function an stationary measure.

   ### Example 10.1.

   - For unconstrained problems:
     $$\Sigma(u) = \|\nabla J(u)\|_{U^*}$$
   - In the case U is a Hilbert space and the problem restricted over a closed and convex set $C \subset U$.
     $$\Sigma(u) = \|u - P_c(u - \nabla J(u))\|$$

   We have global convergence if:

   - Every accumulation point of $(u_k)_k$ is a stationary point.
   - $\lim_{k\to\infty} \Sigma(u_k) = 0$ for some continuous stationary measure.
   - $(u_k)_k$ has an accumulation point which is stationary.
   - $\liminf_{k\to\infty} \Sigma(u_k) = 0$ and Σ continuous.

2. **Fast local convergence:** Let $\overline{u}$ be a stationary point of (P). We say that $u_k \to \overline{u}$ locally with *q-superlinear rate* if $\|u_{k+1} - \overline{u}\|_U = C_k \|u_k - \overline{u}\|_U$, as $k \to \infty$ in a neighborhood of $\overline{u}$ and $C_k \to 0$ as $k \to 0$. We have convergence of order $\alpha + 1$, $\alpha > 0$ if $\|u_{k+1} - \overline{u}\|_U = C_k \|u_k - \overline{u}\|_U^{\alpha+1}$. With $\alpha = 1$ we obtain *q-quadratic* convergence.

10.2. Gradient methods

In the following, we assume H to be Hilbert space.

Motivation

Dynamical systems in physics are often based on the idea of gradient flow with respect to the energy which the system follows. Consider for example the heat equation. Hence, the thermal energy is given by

$$E(u) = \frac{1}{2} \int |\nabla u|^2 \, dx.$$

The gradient flow is defined by,

$$\frac{\partial u}{\partial t} = -E(u),$$

which yields the heat equation.

$$\dot{u}(t) = \Delta u$$

In order to obtain a minimization method in U, we introduce the gradient flow $\frac{\partial u}{\partial t} = -\nabla J(u)$, with $\nabla J(u) \in H$ which can be associated with the gradient of J at u, i.e.

$$\left\langle v, \frac{\partial u}{\partial t} \right\rangle = \langle v, \nabla J(u) \rangle = -J'(u; v), \quad \forall v \in H$$

The evolution of J corresponding to the gradient flow is given by

$$\frac{\partial}{\partial t} (J(u)) = \left\langle \frac{\partial u}{\partial t}, \nabla J(u) \right\rangle = -\left\| \frac{\partial u}{\partial t} \right\|^2 \leq 0$$

i.e. the objective function is decreasing and

$$\frac{\partial}{\partial t} J(u) = 0 \iff \frac{\partial u}{\partial t} = 0$$

.

Consequently, the gradient flow will decrease the objective function J until the evolution arrives a stationary point. In order to obtain an iterative optimization method, we use an explicit time discretization of the flow, i.e.

$$\begin{aligned} u_{k+1} &= u_k - \tau_k \nabla J(u_k) \quad \text{for } k = 0, 1, 2, .. \\ \nabla J(u_{k+1}) &= 0 \quad \text{Stop the iteration.} \end{aligned}$$

with appropriate (small) choice of the time step $\tau_k > 0$.

Remark 10.1.
We are only interested in the minimization of the objective function and not in the accurate approximation of the solution of the gradient flow.

Therefore we select the step size purely based on the suitable descent of objective function. A classical way to do this is the so called *Armijo-Goldstein* rules, which are based on the "effective descent",

$$\mathcal{D}_{\text{eff}}(\tau) = J(u_k + \tau s) - J(u_k)$$

and the "expected descent"

$$\mathcal{D}_{\text{exp}} = \tau J'(u_k; s) = \tau \langle s, \nabla J(u_k) \rangle$$

where $s = -\nabla J(u_k)$. They are related to each other by the Taylor formula,

$$\mathcal{D}_{\text{eff}}(\tau) = \mathcal{D}_{\text{exp}}(\tau) + O(\tau)$$

Therefore, we can test if

$$\alpha \mathcal{D}_{\text{exp}}(\tau) \leq \mathcal{D}_{\text{eff}}(\tau) \leq \beta \mathcal{D}_{\text{exp}}(\tau) \tag{*}$$

with constants $0 < \beta < \alpha < 1$.

If $\mathcal{D}_{\text{eff}}(\tau) > \beta \mathcal{D}_{\text{exp}}(\tau)$, then τ is too large and we decrease it. If $\mathcal{D}_{\text{eff}}(\tau) < \alpha \mathcal{D}_{\text{exp}}(\tau)$, we could still increase τ. We accept τ, if (*) is fulfilled. Typical choices of the constants are $\alpha \approx 0.9$ and $\beta \approx 0.1$.

Moreover, the strategy for increasing and decreasing τ should be different, i.e. multiplying with 1.5, for increasing and dividing by 2 for decreasing.

Theorem 10.1.
Let $J : H \to \overline{\mathbb{R}}$ be twice Fréchet-differentiable and weakly lower semicontinuous on a Hilbert space H. Moreover, let the level sets,

$$\mu_\xi = \{u \in U \mid J(u) \leq \xi\}$$

be bounded in H for each $\xi \in \mathbb{R}$ and empty for sufficiently small ξ. Then the sequence $(u_k)_k$ generated by the gradient method with the Amijo-Goldstein line search has a weakly convergent subsequence, whose limit is a stationary point.

Proof. Since the gradient method is a descent method we have $J(u_k) \leq J(u_0)$, $\forall k \geq 0$, i.e. $(u_k)_k$ is bounded, the lemma (0.1), implies that $\exists (u_{k_l})_l \xrightarrow[l \to \infty]{} \overline{u}$.
Therefore,

$$\sum_{k=0}^{N} \|u_{k+1} - u_k\|^2 = \sum_{k=0}^{N} \langle u_{k+1} - u_k, -\tau_k \nabla J(u_k) \rangle$$

Considering that,

$$\frac{1}{\beta} \mathcal{D}_{\text{eff}}(\tau) = \frac{1}{\beta} \left(J(u_k + \tau s) - J(u_k) \right) \leq \mathcal{D}_{\text{exp}}(\tau) = \tau J'(u_k; s)$$

Then we have substituting,

$$\begin{aligned}\sum_{k=0}^{N}\|u_{k+1}-u_k\|^2 &\leq \frac{1}{\beta}\sum_{k=0}^{N}(J(u_k)-J(u_{k+1}))\\ &= \frac{1}{\beta}(J(u_0)-J(u_{N+1}))\\ &\leq \frac{1}{\beta}(J(u_0)-\inf_{u\in U}J(u)) = q\end{aligned}$$

q is independent of N; for $N\to\infty$ we obtain

$$\sum_{l=0}^{\infty}\|u_{k_l+1}-u_{k_l}\|^2 = \sum_{k=0}^{\infty}\|u_{k+1}-u_k\|^2 \leq q$$

Thus, there is a subsequence $(u_k)_l$, such that

$$\|\tau_{k_l}\nabla J(u_{k_l})\| = \|u_{k_l+1}-u_{k_l}\| \xrightarrow{l\to\infty} 0.$$

$\exists c>0,\ J''(u_{k_l};(v,v)) \leq c\|v\|^2,\ \forall v\in U$. Amijo-Goldstein implies,

$$\begin{aligned}\alpha\mathcal{D}_{\exp}(\tau) &= \alpha J'(u_{k_l};u_{k_l+1}-u_{k_l})\\ &\leq \alpha\mathcal{D}_{\exp}(\tau_{k_l})\\ &= J(u_{k_l+1})-J(u_{k_l})\\ &= J'(u_{k_l};u_{k_l+1}-u_{k_l}) + \int_0^1 J''\left(u_{k_l}+t(u_{k_l+1}-u_{k_l});((u_{k_l+1}-u_{k_l},u_{k_l+1}-u_{k_l}))\right)dt\\ &\leq J'(u_{k_l};u_{k_l+1}-u_{k_l}) + C\|u_{k_l+1}-u_{k_l}\|^2\end{aligned}$$

implies,

$$(\alpha-1)J'(u_{k_l};u_{k_l+1}-u_{k_l}) \leq C\|u_{k_l+1}-u_{k_l}\|^2$$

Since $u_{k_l+1}-u_{k_l} = -\tau_k\nabla J(u_{k_l})$.

$$(1-\alpha)\tau_{k_l}\|\nabla J(u_{k_l})\|^2 \leq c\tau_{k_l}^2\|\nabla J(u_{k_l})\|^2$$

Therefore, $\nabla J(u_{k_l})=0$, or $1-\alpha\leq c\tau_{k_l}$. But since $1-\alpha>0$ and $c<0$. We have $\nabla J(u_{k_l})=0$, implying the algorithm reached stationary point;

$$\forall j\geq k_l \quad u_j = u_{k_l} \implies \|\nabla J(u_{k_l})\|\to 0 \implies \nabla J(\overline{u}) = 0$$

. □

Remark 10.2.

The Armijo-Goldstein rule can be applied to any method which yields a descent direction, i.e. an element $s\in U$ of a Banach space satisfying

$$(s \mid \nabla J(u_k))_{UU^*} < 0$$

Descent Method:

For $k = 0, 1, \dots$

- Choose descent direction $s^k \in U : (s^k \mid \nabla J(u_k))_{UU^*} < 0$.
- Choose step size $\tau_k > 0 : J(u_k + \tau_k s^k) < J(u_k)$.
- $u_{k+1} = u_k + \tau_k s^k$. Stop iteration if $\nabla J(u_{k+1}) = 0$.

11. Lecture 11

11.1. The proximal point algorithm

In the following, we assume that $J : U \to \overline{\mathbb{R}}$ is proper, lower semicontinuous and convex. We denote the optimal value by

$$\alpha = \inf_{u \in U} J(u)$$

and the set of solutions by S. Under these assumptions, the Moreau-Yosida regularization $J_{(\lambda,z)}$ of J defined as

$$J_{(\lambda,z)} = \frac{1}{2\lambda} \|u - z\|^2 + J(u)$$

is proper, lower-semicontinuous and convex.

Definition 11.1.

The proximal mapping $\operatorname{prox} : U \to U$ is defined as

$$\operatorname*{prox}_{\lambda,J}(z) = \operatorname*{arg\,min}_{u \in U} \left[J_{(\lambda,z)}(u) \right] = \operatorname*{arg\,min}_{u \in U} \left[\frac{1}{2\lambda} \|u - z\|^2 + J(u) \right]$$

with $\lambda > 0, z \in U$.

The proximal point algorithm

The proximal point algorithm is given by iterating

- Choose u_0.
- $u_{n+1} = \operatorname*{prox}_{\lambda_n,J}(u_n)$, for $n = 1, 2, \dots$.

and the generated sequence is called proximal sequence.

By theorem 6.3 and 5.3, we know that

$$0 \in \partial J_{(\lambda_n,u_n)}(u_{n+1}) \iff -\frac{u_{n+1} - u_n}{\lambda_n} \in \partial J(u_{n+1})$$

This can be written with the resolvent operator as

$$u_{n+1} = (I + \lambda_n \partial J)^{-1}(u_n)$$

Clearly, stationary points of the proximal sequence are minimizers of the objective function J:

$$u_{n+1} = u_n \iff 0 \in \partial J(u_{n+1}),$$

i.e. if we find a fix point of the proximal operator.

Lemma 11.1.
Let $(u_k)_k$ be a proximal sequence. Then:

1. *Let $(J(u_k))_k$ is non-increasing. Moreover, it decreases strictly, as long as $u_{n+1} \neq u_n$*
2. *For each $u \in U$ and $n \in \mathbb{N}$, we have*

$$2\sigma_n[J(u_{n+1}) - J(u_n)] \leq \|u_0 - u\|^2$$

with $\sigma_n = \sum_{k=0}^{n} \lambda_k$

Proof. Let $(u_k)_k$ be a proximal sequence. Then by the definition of proximal point algorithm, we see

$$\frac{1}{2\lambda_n}\|u_{n+1} - u_n\|^2 + J(u_{n+1}) \leq \frac{1}{2\lambda_n}\|u_n - u_n\|^2 + J(u_n) = J(u_n) \quad \forall n \geq 0$$

Implies $J(u_{n+1}) \leq J(u_n)$, $\forall n$, i.e. the sequence $(J(u_k))_k$ is non increasing. Indeed, it is decreasing strictly as long as $u_{n+1} \neq u_n$. By definition of the subgradient, we obtain,

$$J(u) \geq J(u_{n+1}) + \left\langle -\frac{u_{n+1} - u_n}{\lambda_n}, u - u_{n+1} \right\rangle \quad \forall u \in U.$$

With $-\frac{u_{n+1}-u_n}{\lambda_n} \in \partial J(u_{n+1})$.

$$\begin{aligned} 2\lambda_n[J(u_{n+1}) - J(u)] &\leq 2\langle u_{n+1} - u_n, u - u_{n+1}\rangle \\ &= \|u_n - u\|^2 - \|u_{n+1} - u\|^2 - \|u_{n+1} - u\|^2 \\ &\leq \|u_n - u\|^2 - \|u_{n+1} - u\|^2 \end{aligned}$$

Taking the sum we find

$$2\sum_{n=0}^{N} \lambda_n[J(u_{n+1}) - J(u)] \leq \|u_n - u\|^2 - \|u_{N+1} - u\|^2 \leq \|u_0 - u\|^2$$

$J(u_{n+1}) \leq J(u_{N+1})$.
Implies 2 $\sigma_n(J(u_{N+1}) - J(u)) \leq \|u_0 - u\|^2$ □

Theorem 11.1 (Weak convergence of the proximal sequence).
Let $(u_k)_k$ be a proximal sequence.

1. *If $\sigma_n \xrightarrow[n\to\infty]{} \infty$, then $J(u_n) \downarrow \alpha$ as $n \to \infty$*
2. *Every weak limit point of the $(u_k)_k$ lies in S.*

3. *If $S \neq \{\}$, then u_n converges weakly to a minimizer of J and*

$$J(u_{n+1}) - \alpha \leq \frac{\mathrm{d}(u_0, S)}{2\sigma_n}$$

Proof.

1. By the previous lemma (11.1), we obtain $J(u_{n+1}) - J(u) \leq \frac{\|u_0 - u\|^2}{2\sigma_n}$. Consequently, if $\sigma_n \to \infty$, $J(u_n) \to \alpha = \inf_{u \in U} J(u)$.
2. Follows by the weak lower semi-continuity of J.
3. Replacing $J(u) \leq \alpha$

□

Theorem 11.2 (Strong convergence of proximal sequence).

Let $(u_k)_k$ be a proximal sequence, with $(\lambda_n)_n \notin \ell^1$. If J is strongly convex with parameter $\alpha > 0$, i.e.

$$J(\lambda u + (1-\lambda)u) \leq \lambda J(u) + (1-\lambda)J(v) - \frac{\alpha}{2}\|u - v\|^2 \lambda(1-\lambda),$$

$\lambda \in [0,1]$, with $u, v \in U$. Then $S = \{\overline{u}\}$ and

$$\|u_{n+1} - \overline{u}\| \leq \|u_0 - u\| \prod_{k=0}^{n} (1 + \alpha\lambda_k)^{-1} \qquad (\clubsuit)$$

In particular, $(u_n)_n$ converges strongly to the unique minimizer $\overline{u} \in S$ as $n \to \infty$.

Lemma 11.2.

Let $J : U \to \overline{\mathbb{R}}$, be strongly convex with parameter α. Then for $u^* \in \partial J(u)$ and $v \in U$, the inequality,

$$J(v) \geq J(u) + \langle u^*, v - u\rangle + \frac{\alpha}{2}\|v - u\|^2$$

holds. Moreover, $v^* \in \partial J(v)$, it holds that

$$\langle u^* - v^*, u - v\rangle \geq \alpha \|u - v\|^2$$

Example 11.1.

If the function $J : U \to \mathbb{R}$ is simple, the proximal map can be obtained by an explicit formula.

1. The indicator function $J(u) = I_C(u)$. The proximal iterations: $u_{n+1} = P_C(u_n)$, with projection $P_C : U \to U$ to the set C
2. $J : \mathbb{R}^n \to \mathbb{R}^n$, $u \mapsto \|u\|_1$. Then the proximal iterations are completed via the soft shrinkage operator.

$$u_{n+1} = \mathcal{S}_\lambda(u_n) = \left(I + \lambda\partial \|\cdot\|_1\right)^{-1}(u_n)$$

with $\mathcal{S}_\lambda(u) = (S_\lambda(u_i))_{i=1}^n$.

$$S_\lambda(x) = \begin{cases} x - \lambda, & x > \lambda \\ 0, & -\lambda \leq x \leq \lambda \\ x + \lambda, & x \leq -\lambda \end{cases}$$

Symbols

$\mathbb{N}$	Natural numbers
$\mathbb{R}$	The real line $(-\infty, \infty)$
$\mathbb{R}_+$	Positive real numbers $(0, \infty)$
$\mathbb{R}_-$	Negative real numbers $(-\infty, 0)$
$\overline{\mathbb{R}}$	Extended real line $[-\infty, \infty]$
$\mathsf{D}J(u; v)$	First Derivative at u with direction v.
$\operatorname{dom} J$	Effective Domain of J.
$(x_k)_k$	Sequence with $k \in \mathbb{N}$ indexing the elements x_k.
$(x_{k_l})_l$	Subsequence of a sequence $(x_k)_k$, indexed by $l \in \mathbb{N}$
$\mathcal{L}(X, Y)$	Set of linear operators from X to Y.
$\mathcal{B}(X, Y)$	Set of linear and bounded operators from X to Y.
$\mathcal{L}(X)$	Set of linear functionals from X to $\mathbb{R}$.
$\mathcal{B}(X)$	Set of linear and bounded functionals from X to $\mathbb{R}$.
$C^k(X, Y)$	Set of continuously k-differentiable operators from X to Y. If Y is not specified assume the real line.
$C_0^k(X, Y)$	Set of continuously k-differentiable operators from X to Y with compact support.
$\operatorname{epi} J$	Epigraph of J.
$\operatorname{gra} J$	Graph of J.
I_C	Indicator function of the set C.
$\overline{C}$	Closure of the set C
$A^\star$	Adjoint of operator A.
J^*	Convex conjugate of the functional J.

References

[1] Lecture notes of Prof. Dr. Anna Christina Brandt.

[2] V. Barbu and Th. Precupanu, Convexity and optimization in Banach spaces.

[3] I. Ekeland and R. Teman, Convex analysis and variational problems.

[4] H. Bauschke and P. Combettes, Convex analysis and monotone operator theory in Hilbert spaces.