

Optimization

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

Abstract:

Keywords: Optimization • Convexity

Introduction

0.1. Definitions

Definition 0.1.

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

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.

Theorem 0.1.

Let $f : H \rightarrow (\infty, \infty]$ 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 continuous.

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 0.2.

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

$$\bar{f} = \sup \{g : \mathcal{X} \rightarrow [-\infty, \infty] \mid g \leq f \text{ and } g \text{ is lower semicontinuous}\}.$$

1. Lecture 1

Fact 1.1.

Let $x \in H$, let U be a neighborhood of x , let G be a real Banach space, let $T : U \rightarrow G$, let V be a neighborhood of Tx , and let $R : V \rightarrow K$. Suppose that T is Frchet differentiable at x and that R is Gateaux differentiable at Tx . Then $R \circ T$ is Gateaux differentiable at x and $D(R \circ T)(x) = (DR(Tx)) \circ DT(x)$. If R is Fréchet differentiable at x , then so is $R \circ T$.

Fact 1.2.

Let $x \in H$, let U be a neighborhood of x , let K be a real Banach space, and let $T : U \rightarrow K$. Suppose that T is twice Fréchet differentiable at x . Then $\forall (y, z) \in H \times H$, $(D^2T(x)y)z = (D^2T(x)z)y$.

Definition 1.1.

Let $x \in H$, let $C \in \mathcal{V}(x)$, and let $T : C \rightarrow K$. Then T is Fréchet differentiable at x if there exists an operator $DT(x) \in B(H, K)$, called the Frchet derivative of T at x , such that

$$\lim_{0 \neq \|y\| \rightarrow 0} \frac{\|T(x+y) - Tx - DT(x)y\|}{\|y\|} = 0$$

2. Lecture 2

2.1. Convexity

Definition 2.1.

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

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

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 in U , then

- $\alpha V = \{w = \alpha v, v \in C\}$ is convex.
- $C + V$ is convex.

Proof. □

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 $C \in U$ convex and $J : C \rightarrow \mathbb{R}$. Define $\alpha = \inf_{u \in C} J(u)$. Then the set $\Psi = \{u \mid J(u) = \alpha\}$ 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 $J(u_t) \leq tJ(u_1) + (1-t)J(u_2) = \alpha$. 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 $J : U \rightarrow \overline{\mathbb{R}}$ convex functional. Let $\bar{u} \in C$ such that

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

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

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

For some $t \in (0, 1)$, $u_t \in B_\epsilon(\bar{u})$.

Thus,

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

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

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

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

Theorem 2.1.

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

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

1. Let \bar{u} be a local solution. Then $J'(\bar{u}; u - \bar{u}) \geq 0, \forall u \in C$.

2. If J is convex on C , then $J'(\bar{u}; u - \bar{u}) \geq 0, \forall u \in C$ is necessary and sufficient for global optimality of \bar{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

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

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

Proof.

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

For small $t > 0$,

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

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

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

Therefore \bar{u} is a global minimizer.

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

$\implies (u_k)_k$ is bounded, because $J \rightarrow \infty$ as $\|u\| \rightarrow \infty$.

$\implies (u_k)_k$ contains a weakly convergent subsequence $u_{k_e} \xrightarrow{e \rightarrow \infty} \bar{u} \in C$. Since C is closed and convex.

$\implies J$ is weakly-lower semicontinuous because it is convex and continuous.

□

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

$$\arg \min_{v \in C \subseteq H} J(x) := \{x \mid x \in C \wedge \forall v \in C : J(x) \leq J(v)\}.$$

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 Gateaux gradient of J at x .

Lemma 3.1.

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

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

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. $\|P_C(u) - P_C(\bar{u})\| \leq \|u - \bar{u}\| \ \forall u, \bar{u} \in H$, i.e. The projection P_C is non expansive.
4. $\langle P_C(u) - P_C(\bar{u}), u - \bar{u} \rangle \leq 0, \ \forall u, \bar{u} \in H$
5. Let be $t > 0$ a real number, then $\forall u \in C$, and $\forall v \in H$, $\phi(t) = \frac{1}{t} \|P_C(u + tv) - u\|$ is non-increasing.

Proof.

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

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

By the parallelogram law,

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

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,

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

Since $\|v_i - x\| \rightarrow \alpha$ and $\|v_j - x\| \rightarrow \alpha$, we have that $\|v_j - v_i\| \rightarrow 0$, 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.

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

By the parallelogram law,

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

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

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

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. □

Theorem 3.1.

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

Proof. Since every Hilbert Space is a Banach space, and C is closed and Convex subset of H , and \bar{u} is a solution of minimization problem; we can apply 2.1.

Thus $J'(\bar{u}; u - \bar{u}) \geq 0 \iff \langle u - \bar{u}, \nabla J(\bar{u}) \rangle \geq 0 \forall u \in C$.

For all $\delta > 0$, we multiply the Gateaux gradient $(-\delta)$ and we have,

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

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

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

Thus,

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

□

3.1. Application

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

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

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 $u \in U$,

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

Consider the reduced cost functional $F(u) := J(y(u), u)$, then our problem is equivalent to

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

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

Proposition 3.1.

If J is continuous and weakly lower semicontinuous, then $\bar{u} = \arg \min_{u \in U_{ad}} [F(u)]$.

Proof. If J is weakly lower semicontinuous

$$J(y(\bar{u}), \bar{u}) \leq \liminf_{l \rightarrow \infty} J(y(u_k), u_k)$$

That is,

$$F(\bar{u}) \leq \liminf_{l \rightarrow \infty} F(u_k) = \alpha$$

Since $u_{k_l} \xrightarrow{l \rightarrow \infty} \bar{u}$, $\implies y(u_k) \rightharpoonup y(\bar{u})$ and $A^{-1}Bu_k \rightharpoonup A^{-1}B\bar{u}$ □

J is Gateaux differentiable, applying the chain rule to F and valuating in u we have

$$F_u(u; h) = J_y(y; \alpha(u; h), u) + J_u(y, u; h)$$

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

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

4. Lecture 4

Lemma 4.1.

Let U be linear space and $J : U \rightarrow \bar{\mathbb{R}}$. Then

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

1. Assume J convex. If $u_1 \in \text{dom}(J)$ and u_2 are elements of $\text{dom}(J)$. 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, $tu_1 + (1-t)u_2 \in \text{dom}(J)$. Hence $\text{dom } J$ 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 $(u_1, \alpha_1), (u_2, \alpha_2)$ elements of $\text{epi}(J)$, 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 $(tu_1 + (1-t)u_2, t\alpha_1 + (1-t)\alpha_2) \in \text{epi}(J)$. Therefore, if J is convex, and $(u_1, \alpha_1), (u_2, \alpha_2)$ are elements of $\text{epi}(J)$ then,

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

Hence $\text{epi}(J)$ is convex.

Now assume $\text{epi}(J)$ convex. Let $(u_1, \alpha_1), (u_2, \alpha_2)$ elements of $\text{epi}(J)$ then $(tu_1 + (1-t)u_2, t\alpha_1 + (1-t)\alpha_2)$, 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 $\text{epi}(J)$, if $u_1, u_2 \in \text{dom } J$, then $(u_1, J(u_1))$ and $(u_2, J(u_2))$, are elements of $\text{epi}(J)$, 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 $J : U \rightarrow \overline{\mathbb{R}}$ 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 $J(u_0) - \epsilon \leq J(u) \quad \forall u \in B_\delta(u_0)$.
- If $J(u_0) \leq \liminf_{n \rightarrow \infty} J(u_n)$ 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 $J : U \rightarrow \overline{\mathbb{R}}$. 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 $\text{epi}(J)$ is closed.

3. The level sets $\mu_\xi = \{u \in U \mid J(u) \leq \xi\}$ is a closed set. Note that the sets μ_ξ are closed if and only if the sets $\gamma_\xi = \{u \in U \mid J(u) > \xi\}$ are open. (Since $\mu_\xi^c = \gamma_\xi$).

Proof.

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

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

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

- (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 $\text{epi}(J)$. Since $\text{epi}(J)$ is closed, we conclude that $(u, \xi) \in \text{epi}(J)$, and hence $u \in \mu_\xi$.
- (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 $\eta = \liminf_{n \rightarrow \infty} J(u_n)$. 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 liminf, the sequence $(u_n)_n$ has a subsequence $(u_k)_k$, such that $J(u_k) \xrightarrow{k \rightarrow \infty} \eta$. 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

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

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

□

Example 4.1.

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

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

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 μ_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:

$$(\cdot, \cdot) : X \times Y \rightarrow \Gamma$$

.

The bilinear functional is sometimes omitted.

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

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

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

$$f_x(y) = (x, y) \quad \forall y \in Y$$

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

$$g_y(x) = (x, y), \quad \forall x \in X.$$

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,

$$p_y(x) = |(x, y)| = |g_y(x)|, \quad \forall x \in X$$

$$q_x(y) = |(x, y)| = |f_x(y)|, \quad \forall y \in Y$$

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

Definition 4.2.

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

Theorem 4.2.

Let U be 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 $\phi_g : L^p(\Omega) \rightarrow \mathbb{R}$ 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 $p, q > 1$,

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

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 $(\cdot, \cdot) : U \times U^* \rightarrow \mathbb{R}$, defined as

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

Definition 4.3.

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

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

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

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 $u_n \xrightarrow[n \rightarrow \infty]{} u$

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 addition separable, then it's weakly sequentially compact.

Definition 4.4.

Let U be a Banach space and $J : U \rightarrow \mathbb{R}$, 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

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

Definition 4.5.

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 4.6.

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 $J : U \rightarrow \overline{\mathbb{R}}$ 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 $J : U \rightarrow \overline{\mathbb{R}}$ 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 semicontinuous it follows that in the limit $\lambda \rightarrow 0$,

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

□

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 4.3.

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 4.7.

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

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

Remark 4.4.

Every sublinear function is convex.

Theorem 4.6.

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

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

Definition 4.8.

Let $J : U \rightarrow \overline{\mathbb{R}}$, we said that J is locally bounded around u_0 if $\exists V \subset U$ neighborhood of u_0 such that for some $M \in \mathbb{R}$

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

Lemma 4.5.

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

Proof. Let $u_k \rightarrow u$ as $k \rightarrow \infty$. For each $\epsilon > 0$ we can find a sequence α_k such that $\left\| \frac{u - u_k}{\alpha_k} \right\| < \epsilon$, and $\alpha_k \rightarrow 0$ as $k \rightarrow \infty$. (Please read Maximal Monotone Operators and Evolution Systems in Banach Spaces of Barbu. Details Still to be recovered).

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 $\alpha_k \rightarrow 0$, then

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

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

Remark 4.5.

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

Characterization for J'' still to be analyzed the professor's notes are flawed.

Definition 5.1.

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

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

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

Example 5.1.

Consider $J : \mathbb{R} \rightarrow \mathbb{R}$, $u \rightarrow |u|$ 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 $J : \mathbb{R} \rightarrow \mathbb{R}$,

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

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

Example 5.3.

Let C be a convex and closed subset of U and \mathcal{X}_C function defined by

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

The subdifferentiable is the definition of normal cone at u

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

.

Theorem 5.1.

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

Remark 5.1.

Most of the rules for derivatives also hold for subdifferentials with some additional assumptions,

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

Theorem 5.2 (Rockefeller).

Let U be a Banach space and $J : U \rightarrow \mathbb{R}$ proper and convex functions for $i = 1, \dots, n$. The sum-rule

$$\partial(J_1 + \dots + J_n)(u) = \partial J_1(u) \dots \partial J_n(u), \quad n \geq 2$$

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 $J_k, k \in \{1, 2, \dots, n\}$ are continuous at u_0

6. Lecture 6

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 $J^* : U^* \rightarrow \overline{\mathbb{R}}$ be the conjugate of the $J : U \rightarrow \overline{\mathbb{R}}$. Then for all $u \in U$.

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

.

Proof. content...

□

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 $J : U \rightarrow \mathbb{R}$ 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 $(J^*)^* = J^{**} : U^{**} \rightarrow \overline{\mathbb{R}}$,

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

Theorem 7.3.

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

Proof.

$$J^{**}(u) = \sup_{p^* \in U^*} \{p^*(u) - J^*(p^*)\} \tag{2}$$

$$= \sup_{p^* \in U^*} \left\{ p^*(u) - \sup_{v \in U} \{p^*(v) - J(v)\} \right\} \tag{3}$$

$$= \sup_{p^* \in U^*} \left\{ p^*(u) + \inf_{v \in U} \{p^*(v) - J(v)\} \right\} \tag{4}$$

$$\tag{5}$$

Since for any $p^* \in U^*$,

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

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

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

$$\implies F(v) \geq F(v) + q^*(v - u) \quad (6)$$

$$F^{**}(u) = \sup_{p^* \in U^*} \inf_{v \in U} \{p^*(u - v) + F(v)\} \quad (7)$$

$$\geq \sup_{p^* \in U^*} \inf_{v \in U} \{(p^* - q^*)(u - v) + F(v)\} \quad (8)$$

$$\geq \inf_{v \in U} \{(q^* - q^*)(u - v) + F(v)\} \quad (9)$$

$$= F(u) \quad (10)$$

If F is convex,

$$\implies F(u) \leq F^{**}(u) \leq F(u) \implies F(u) = F^{**}(u), \quad (11)$$

$$F(u) = F^{**}(u) = \sup_{p^* \in U^*} \inf_{v \in U} \{p^*(u - v) + F(v)\} \leq J^{**}(u) \quad (12)$$

□

8. Lecture 8

Definition 8.1.

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

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

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

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

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

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

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

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

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

$$-\infty \leq \sup(P^*) \leq \inf(P) \leq \infty.$$

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

$$-\Phi^*(0, p^*) = -\sup_{\substack{u \in U \\ p \in Y}} ((0, u) + (p^*, p) - \Phi(u, p)) \quad (13)$$

$$= \inf_{\substack{u \in U \\ p \in Y}} (\Phi(u, p) - (p^*, p)) \quad (14)$$

$$\leq (\Phi(u, 0) - (p^*, 0)) \quad \forall u \in U, p^* \in Y^* \quad (15)$$

$$\implies \sup_{p^* \in Y^*} (-\Phi(0, p^*)) \leq \inf_{u \in U} \Phi(u, 0) = \inf(P) \quad (16)$$

□

By iteration we can define, a bidual problem

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

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

$$h(p^*) = \inf(P_p) = \inf_{u \in U} \Phi(u, p)$$

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.

$$\inf(P) = \sup(P^*) \leq \infty$$

Theorem 8.2 (Extremal relation).

Let $\Phi : U \times Y \rightarrow \overline{\mathbb{R}}$, be convex the the following statements are equivalent:

1. (P) and (P_p) have solutions \bar{u} and \bar{p}^* and $\inf(P) = \sup(P^*)$
2. $\Phi(\bar{u}, 0) + \Phi^*(0, \bar{p}^*) = 0$
3. $(0, \bar{p}^*) \in \partial\Phi(u, 0)$ and $(\bar{u}, 0) \in \partial\Phi^*(0, \bar{p}^*)$

Proof. We proceed by parts:

1. (1) \implies (2): \bar{u} solution of $\inf(P)$ and \bar{p}^* solution of $\sup(P^*)$ and $\inf(P) = \sup(P^*)$. This properties implies, $\Phi(\bar{u}, 0) = \inf(P) = \sup(P^*) = -\Phi(0, \bar{p}^*) \implies \Phi(\bar{u}, 0) + \Phi^*(0, \bar{p}^*) = 0$.
2. (2) \implies (1): $-\Phi^*(0, \bar{p}^*) = \sup(P^*) \leq \inf(P) = \Phi(\bar{u}, 0) = -\Phi^*(0, \bar{p}^*) \implies \sup(P^*) = \inf(P)$
3. (2) \iff (3): $\Phi(\bar{u}, 0) + \Phi^*(0, \bar{p}^*) = 0 = (0, \bar{u}) + (\bar{p}^*, 0) = ((0, \bar{p}^*), (\bar{u}, 0)) \iff (0, \bar{p}^*) \in \partial\Phi(\bar{u}, 0) \quad \forall u \in U, \forall p^* \in \partial J(u) \iff J(u) + J^*(p^*) = (p^*, u)$

□

Fenchel duality.

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

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

with $F : U \rightarrow \overline{\mathbb{R}}$, G convex function $G : V \rightarrow \overline{\mathbb{R}}$ and $A : U \rightarrow V$ 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 $q : Au - p$.

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

References