Static Optimization Lecture Notes*

Jérôme Bolte

March 10, 2014

1 Analysis, Topology: classical facts

The first notions to be recalled are those of the infimum and the supremum. Let $A \subseteq \mathbb{R}$ be a nonempty set, and take $\mathtt{m}, M \in \mathbb{R}$. Then

$$\mathtt{m} = \inf A \Leftrightarrow \begin{cases} \exists x_n \in A : x_n \to \mathtt{m} \\ \forall y \in A, \text{ if } y_n \to y, \text{ then } y \geqslant \mathtt{m}, \end{cases} \tag{1}$$

and also

$$M = \sup A \Leftrightarrow \begin{cases} \exists x_n \in A : x_n \to M \\ \forall y \in A, \text{ if } y_n \to y, \text{ then } y \leqslant M. \end{cases} \tag{2}$$

The second set of facts is related to vector spaces.

Definition 1. Let $(E, \|\cdot\|)$ be a real normed vector space. A sequence is called a *Cauchy sequence* if

$$\big(\forall \varepsilon > 0\big)\ \big(\exists N \in \mathbb{N}\big) \text{ such that } \big(\forall m, n \geqslant N\big)\ \big(\|x_m - x_n\| < \varepsilon\big).$$

Definition 2. $(E, \|\cdot\|)$ is a Banach space if all Cauchy sequences are convergent.

For example, $\mathbb{R}^n$ is a Banach space with any norm, and so is $C^0([0,1], \mathbb{R})$ with the supremum norm $\|f\|_\infty = \sup_{x \in [0,1]} |f(x)|$.

Definition 3. $(H, \langle \cdot, \cdot \rangle)$ is called a *pre-Hilbert space* if H is a real vector space endowed by a scalar product $\langle \cdot, \cdot \rangle \colon H^2 \to \mathbb{R}$.

Properties of the inner product:

1. Bilinearity: for any $\bar{x} \in H$, the mapping $H \to \mathbb{R}$ defined as $y \mapsto \langle \bar{x}, y \rangle$ is linear, and for any $\bar{y} \in H$ the mapping $x \mapsto \langle x, \bar{y} \rangle$ is linear.
2. Nonnegativity: $\forall x \in H$, $\langle x, x \rangle \geqslant 0$.
3. Separation: $\forall x \in H$,
   $\langle x, x \rangle = 0$ if and only if $x = 0$.

Note that if we set $\|x\| = \sqrt{\langle x, x \rangle}$, then $\|\cdot\|$ is a norm.

Definition 4. Let $(H, \langle \cdot, \cdot \rangle)$ be a pre-Hilbert space. If $(H, \|\cdot\|)$ is a Banach space, then H is called a Hilbert space.

Examples: 1. $H = \mathbb{R}^n$ with the inner product $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ is a Hilbert space, and 2. $H = C^0([0,1], \mathbb{R})$ with the inner product $\langle f, g \rangle = \int_0^1 f(x) g(x)\, \mathrm{d}x$ is a pre-Hilbert space.

Let us now provide a very succinct reminder of topology. Let $(E, \|\cdot\|)$ be a normed space, and let $S, \Omega, F \subseteq E$. $B(x, \varepsilon)$ denotes the open ball of center $x \in E$ and radius $r > 0$.

Definition 5.

1. S is *bounded* if there is an $M \geqslant 0$ such that for all $x \in S$, $\|x\| \leqslant M$.
2. Ω is *open in E* if for any $x \in \Omega$ there is an $\varepsilon > 0$: $B(x, \varepsilon) \subset \Omega$.
3. F is *closed in E* if for each sequence $x_n \in F$ such that $x_n \to x$, the limit x is in F.

Some properties:

Proposition 1.

1. *Ω is open if and only if $E \setminus \Omega$ is closed.*
2. *F is closed if and only if $E \setminus F$ is open.*

Any intersection of closed sets in E is a closed set in E, any finite union of closed sets in E is a closed set. The closure of a set $S \subset E$ is defined as the smallest closed set containing S (which exists by

*These are based on the notes taken and typewritten by Gábor Uhrin and Lois Simandjuntak in the winter semester of 2012–2013. I thank them warmly for their nice help.

using the stability by intersection). One can establish that

$$\overline{S} = \{x \in E : \exists x_n \in S, x_n \to x \text{ as } n \to +\infty, \}.$$

Any union of open sets in E is forms an open set in E, any finite intersection of open sets in E is a open set in E. The biggest open set contained in a set S is called its interior (its existence is ensured by the stability by arbitrary union), it is an open set denoted by int S. We have

$$\text{int } S = \{x \in S : \exists \epsilon > 0, B(x,\epsilon) \subset S\}.$$

Definition 6. A set $K \subseteq E$ is *compact* if for each sequence $(x_n)_{n\in\mathbb{N}}$ with $x_n \in K$, there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ and $x \in K$ such that $x_{n_k} \to x$ as $k \to \infty$.

Theorem 2. *Let $(E, \|\cdot\|)$ be finite dimensional and let $K \subseteq E$. K is compact if and only if it is closed and bounded.*

If E is not finite dimensional, then we can only establish that if K is compact then it is closed and bounded. We even have (one of the famous) Riesz theorem:

Theorem 3 (Riesz)**.** *Let $(E, \|\cdot\|)$ be normed vector space. The following assertions are equivalent*

1. *E is finite dimensional,*
2. *The closed unit ball $\bar{B}(0,1)$ of E is compact,*
3. *The closed bounded subsets of E are exactly the compact subsets of E.*

2 Convex sets

From now on $H, \langle\cdot,\cdot\rangle$ or H systematically denotes a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$.

Definition 7. Let $C \subset H$. C is convex if for all x, y in C the segment $[x,y]$ is entirely contained in C. In other words

$$\lambda x + (1-\lambda)y \in C, \ \forall \lambda \in [0,1].$$

Definition 8. Let $(H, \langle\cdot,\cdot\rangle)$ be a Hilbert-space, $c \in H \setminus \{0\}$ be a non-zero vector, and let $b \in \mathbb{R}$ be a scalar.

- A *hyperplane* is the set of those vectors in H that are orthogonal to c, i.e.

$$P := \{\, x \in H : \langle c, x\rangle = 0 \,\},$$

- A *halfspace* is the set of those vectors in H whose innerproduct with c is smaller-equal b,

$$P_+ := \{\, x \in H : \langle x, c\rangle \leqslant b \,\}.$$

The following object is called a *polyhedron*

$$\{\, x \in \mathbb{R}^n : Ax \leqslant b \,\},\ A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m.$$

Denote a_i, $i = 1, \dots, m$ the rows of A and b_i the i-th element of b. Then $Ax \leqslant b$ translates into $\langle a_i, x\rangle \leqslant b_i$ for all integer $i \in \{1, ...m\}$. A polyhedron is thus a finite intersection of closed half-spaces.

Affines spaces, half-spaces, polyhedra are convex sets.

Elementary operations on sets

Definition 9. Let I be a real interval, $S \subseteq H$. Then

$$I \cdot S := \{\, y \in H : y = \lambda \cdot s \text{ for } \lambda \in I, s \in S \,\},$$

Proposition 4. *Let I be a finite family and J be an infinite family respectively. Consider collections $(C_i)_{i\in I}$, $(D_j)_{j\in J}$. Then*

1. *$\sum_{i\in I} C_i$ is convex* (Minkowski sum)
2. *$\bigcap_{j\in J} D_j$ is convex.*

3 Projection theorem

3.1 Statement and corollaries

Lemma 5 (Parallelogram law)**.** *Take $x, y, z \in H$. Then the following is true:*

$$\|z-x\|^2 + \|z-y\|^2 = 2\left\|z - \frac{x+y}{2}\right\|^2 + \frac{1}{2}\|x-y\|^2. \quad (3)$$

Theorem 6. *Let $(E, \|\cdot\|)$ be a normed space. Then the two following are equivalent:*

1. *The parallelogram law (Eq.* (3)*) holds.*
2. *$\|\cdot\|$ comes from an inner product, i.e. $\exists\langle\cdot,\cdot\rangle$ s.th. $\forall x \in E$: $\|x\| = \sqrt{\langle x, x\rangle}$.*

Theorem 7 (Projection Theorem)**.** *Let $C \subseteq H$ be a* closed, convex, nonempty *set and x a point in H.*

1. *There exists a unique $\bar{x} \in C$ such that*

$$\|x - \bar{x}\| = \inf_{y \in C} \|x - y\|$$
$$\Updownarrow$$
$$\|x - \bar{x}\| \leqslant \|x - y\| \; \forall y \in C$$

   This point is called the projection of x on C and is denoted by $P_C(x)$.
   The mapping $P_C : H \to C$ is the projection mapping.

2. *Variational characterization of the projection:*
   Let $x, z \in H$.

$$z = P_C(x) \Leftrightarrow \begin{cases} z \in C \\ \langle x - z, y - z \rangle \leqslant 0, \; \forall y \in C \end{cases} \tag{4}$$

For example, if we take $C = \mathbb{R}^n_+$, then for any $x \in \mathbb{R}^n$, the projection onto the set C is given by

$$P_C(x) = \begin{pmatrix} \max(0, x_1) \\ \max(0, x_2) \\ \vdots \\ \max(0, x_n) \end{pmatrix}$$

The following is an interesting and useful corollary, if C is a closed (vector) subspace of H.

Corollary 8. *If F is a* closed (vector) subpace *of H, then the characterization takes the following form. Let $x, z \in H$,*

$$z = P_F(x) \Leftrightarrow \begin{cases} z \in F \\ \langle x - z, y \rangle = 0, \; \forall y \in F \end{cases} \tag{4'}$$

Denote and define the orthogonal complement of $S \subseteq H$ with

$$S^\perp := \{\, x \in H : \langle x, s \rangle = 0, \; \forall s \in S \,\}$$

Then we can establish the following

Proposition 9.

1. *$S^\perp$ is a closed subspace of H,*

2. *If F is a closed subspace of H, then we have that*

$$H = F \oplus F^\perp,$$

   i.e. for all $z \in H$ there exists a unique $(x, y) \in F \times F^\perp$ such that $z = x + y$.

3.2 Illustration: Least squares

Theorem 10. *Let $A \in \mathbb{R}^{m \times n}$ be a matrix and $b \in \mathbb{R}^m$ a vector, and consider the equation*

$$Ax = b. \tag{5}$$

To this equation one associates the least-squares problem

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|^2. \tag{6}$$

The latter has at least one solution. Further, $\bar{x}$ is a solution if and only if

$$A^T A \bar{x} = A^T b,$$

where A^T is the transpose matrix of A.

3.3 Other consequences of the projection theorem

Theorem 11 (Riesz representation theorem)**.** *For each continuous linear form $l \colon H \to \mathbb{R}$ there exists a unique vector z in H such that $l(x) = \langle z, x \rangle$ for all $x \in H$.*

Now we turn to separation theorems:

Theorem 12 (Hahn-Banach separation theorem I)**.** *Let C and K be two convex sets such that C is closed in H, K is compact, $C \cap K = \emptyset$. Then there exists $x^* \in H \setminus \{\, 0 \,\}$ such that*

$$\sup_{x \in C} \langle x^*, x \rangle < \inf_{y \in K} \langle x^*, y \rangle.$$

In other words, there exist $\alpha < \beta$ real numbers such that

$$\langle x^*, y \rangle < \alpha < \beta < \langle x^*, y \rangle \; \forall x \in C, y \in K$$

Note that in the second formulation, we need two numbers because if we only take one with $\langle x^*, y \rangle < \delta < \langle x^*, y \rangle$, then it does not preclude the case when $\sup_{x \in C} \langle x^*, x \rangle = \inf_{y \in K} \langle x^*, y \rangle$.

Theorem 13 (Hahn-Banach separation theorem II–weak separation)**.** *Let C and D be two disjoint convex sets, and assume that* $\operatorname{int} C \neq \emptyset$*. Then there exists $x^* \in H \setminus \{\, 0 \,\}$ such that*

$$\sup_{x \in D} \langle x^*, x \rangle \leqslant \inf_{y \in C} \langle x^*, y \rangle.$$

Theorem 14 (Hahn-Banach in finite dimensional spaces)**.** *Assume that H is finite dimensional. Let C and D be two disjoint convex sets. Then there exists $x^* \in H \setminus \{\, 0 \,\}$ such that*

$$\sup_{x \in D} \langle x^*, x \rangle \leqslant \inf_{y \in C} \langle x^*, y \rangle.$$

3.4 Convex cones

3.4.1 Definition and conjugacy

Definition 10. A nonempty set $L \subseteq H$ is a *cone* if $\mathbb{R}_+ \cdot L \subseteq L$, i.e. for all $\lambda \geqslant 0$ and for all $x \in L$, $\lambda x \in L$.

Any union of vector subspaces forms a cone. Note that the zero vector is in L whenever L is nonempty.

Proposition 15. *Let $L \subseteq H$ be a nonempty set. L is a convex cone if and only if $\mathbb{R}_+ L \subseteq L$ and $L + L \subseteq L$*

Examples include: Vector subspaces, $\mathbb{R}^n_+$, the set of symmetric nonnegative semidefinite matrices, the cone generated by a finite number of vectors $v_1, \ldots, v_m \in H$, i.e.

$$\{ \lambda_1 v_1 + \ldots + \lambda_m v_m : \lambda_i \geqslant 0, \forall i = 1, \ldots, m \}.$$

Observe that this cone is the smallest convex cone containing the vectors $(v_i)_{i=1,\ldots,m}$.

Definition 11. Let $S \subseteq H$ be nonempty. The *conjugate of S* is defined as

$$S^* := \{ x \in H : \langle x, y \rangle \leqslant 0, \ \forall y \in S \}.$$

The set S^* is a clearly a closed convex cone and $\left(\overline{S}\right)^* = S^*$.

Note also that $F^* = F^\perp$ for any vector subspace F of H. The conjugacy operation is therefore a kind of generalization of the orthogonality idea from vector spaces to cones.

Theorem 16 (Duality for cones)**.** *Let $S \subseteq H$ be nonempty. S^{**} is the smallest closed convex cone containing S. Thus, if L is a closed convex cone, we have $L^{**} = L$.*

When F is a closed vector space one recovers the famous identity $F^{\perp\perp} = F$.

3.4.2 Normal and tangent cones

Definition 12 (Tangent cone)**.** Let $S \neq \emptyset$ be a closed subset of H and let $x \in S$. The *tangent cone* at x to S is defined by

$$T_s(x) := \{ v \in H : \exists x_k \in S, x_k \to x, \\ \exists \lambda_k \geqslant 0 \text{ with } \lambda_k(x_k - x) \to v \}$$

Remarks: (i) $0 \in T_s(x)$,
(ii) $T_s(x)$ is a closed cone.

Definition 13 (Normal cone)**.** Let $C \neq \emptyset$ be closed and convex. The *normal cone* to C at $x \in C$ is defined by

$$N_c(x) = T_c(x)^*.$$

It is therefore a closed convex cone.

Proposition 17. *Let $C \neq \emptyset$ be a closed convex subset of H. Then*

1. *$T_c(x) = \overline{(\mathbb{R}_+(C - x))}$ for all $x \in C$. Hence $T_c(x)$ is also convex.*
2. *$N_c(x) = \{ w \in H : \langle w, y - x \rangle \leqslant 0 \ \forall y \in C \}$* ([1])

4 Convex, concave functions

4.1 Differential calculus

Let E, F be normed spaces, and let $\emptyset \neq \Omega \subseteq E$ be an open subset of E. *The function $f \colon \Omega \to F$ is differentiable at $\bar{x}$* if there exist $L \colon E \to F$ a linear and **continuous** function and $\varepsilon \colon \Omega - \bar{x} \to F$ with $\varepsilon(0) = 0 = \lim_{h \to 0} \varepsilon(h)$ such that

$$f(\bar{x} + h) = f(\bar{x}) + L(h) + \|h\| \varepsilon(h).$$

An important particular case is when $E = H$ a Hilbert space, $F = \mathbb{R}$ and $f \colon \Omega \to \mathbb{R}$. Then we have that $f'(\bar{x}) \in \mathcal{L}(H, \mathbb{R})$, and by the Riesz representation theorem, there exists a unique $z \in H$ such that

$$f'(\bar{x})(h) = \langle z, h \rangle \ \forall h \in H.$$

This vector z is called the gradient of f at $\bar{x}$ and is denoted by $\nabla f(\bar{x})$. If ∇f is differentiable, we set $\mathrm{d}\nabla f(\bar{x}) = \nabla^2 f(\bar{x}) \in \mathcal{L}(H, H)$.

The second order Taylor expansion of f around $\bar{x}$ is given by

$$f(\bar{x} + h) = \\ f(\bar{x}) + f'(\bar{x})h + \tfrac{1}{2}\langle \nabla^2 f(\bar{x})h, h \rangle + \|h\|^2 \varepsilon(h), \quad (7)$$

where $\varepsilon \colon H \to \mathbb{R}$ is defined on a neighborhood of 0 and

$$\lim_{h \to 0} \varepsilon(h) = \varepsilon(0) = 0.$$

Clairaut-Schwarz theorem asserts that the Hessian is symmetric, that is

$$\langle \nabla^2 f(x)u, v \rangle = \langle u, \nabla^2 f(x)v \rangle, \quad \forall u, v \in H. \quad (8)$$

[1] This formula is particularly useful.

4.2 Convex functions

Recall that $(H, \langle\cdot,\cdot\rangle)$ is a real Hilbert space.

Definition 14 (Convexity)**.** Let $C \subseteq H$ nonempty, convex and let $f: C \to \mathbb{R}$ be a function. We say that f is *convex* if for all $\lambda \in [0,1]$ and for all $x, x' \in C$

$$f(\lambda x + (1-\lambda)x') \leqslant \lambda f(x) + (1-\lambda) f(x'). \quad (9)$$

It is *strictly convex* if the above inequality is strict for all $x \neq x' \in C$ and $\lambda \in]0,1[$.
A function f is *concave* if $-f$ is convex.

Elementary examples include:

- Affine forms: $f(x) = \langle c, x\rangle + d$ where x ranges in H,
- $\|\cdot\|$ and $\|\cdot\|^2$.

Observe that $\|\cdot\|^2$ is strictly convex while affine forms and $\|\cdot\|$ are not.

Definition 15 (Epigraph)**.** Let $S \subseteq H$ be nonempty and let $f: S \to \mathbb{R}$. The *epigraph* of f is

$$\operatorname{epi} f := \{\, (x,\lambda) \in S \times \mathbb{R} : f(x) \leqslant \lambda \,\} \subseteq H \times \mathbb{R}. \quad (10)$$

The *sublevel set* of f with level α is

$$[f \leqslant \alpha] := \{\, x \in H : f(x) \leqslant \alpha \,\}. \quad (11)$$

Proposition 18. *Let $C \subseteq H$ be convex, nonempty and let $f: C \to \mathbb{R}$. Then*

1. *f is convex if and only if its epigraph is convex,*
2. *if f is convex, then $[f \leqslant \alpha]$ is convex for all α (but the converse is not true).*

Proposition 19. *Let $C \subseteq H$ be convex, nonempty, and let $f_i: C \to \mathbb{R}$ be convex for all $i \in I$ with I being an arbitrary index set. Assume that $\sup_{i\in I} f_i(x) < \infty$ for all $x \in C$. Then the function*

$$C \ni x \to \sup_{i\in I} f_i(x)$$

is convex.

Theorem 20 (Continuity for convex functions)**.** *Let C be a nonempty,* **open***, convex set and let $f: C \to \mathbb{R}$ be a convex function. Assume f is bounded from above in the neighborhood of an arbitrary $\bar{x} \in C$. Then f is continuous.*

Corollary 21 (Automatic continuity for convex functions)**.** *Assume that H is finite dimensional, let C be a nonempty, open, convex set and let $f: C \to \mathbb{R}$ be a convex function. Then f is continuous on C .*

Theorem 22. *Let $C \subseteq H$ be nonempty, convex, open and let $f: C \to \mathbb{R}$ be differentiable ($f \in \Delta^1$ for short). Then*

1. *f is convex if and only if*
$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x\rangle \ \forall x, y \in C \quad (12)$$
2. *f is convex if and only if*
$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geqslant 0 \ \forall x, y \in C. \quad (13)$$

Theorem 23. *Let $C \subseteq H$ be nonempty, and $f: C \to \mathbb{R}$, $f \in \Delta^2$. Then*

1. *f is convex if and only if*
$$\langle \nabla^2 f(x)h, h\rangle \geqslant 0 \ \forall x \in C, \forall h \in H,$$
i.e. $\nabla^2 f(x)$ is semidefinite positive.
2. *If $\langle \nabla^2 f(x)h, h\rangle > 0, \forall x \in C$, $\forall h \in H \setminus \{0\}$, then f is strictly convex.*

5 Convex optimization

5.1 Existence and first order conditions

Theorem 24 (Existence of minimizers)**.** *Let $K \subseteq H$ be nonempty, closed, convex and let $f: H \to \mathbb{R}$ be convex and continuous. Assume further that $\exists \alpha \in \mathbb{R}$ such that $[f \leqslant \alpha] \cap K$ is nonempty and bounded.*
Then the problem

$$\inf_{x\in K} f(x) \quad (P)$$

has a solution.

Remarks: When H is finite dimensional the result is a consequence of Weierstrass theorem and on the fact that $[f \leqslant \alpha] \cap K$ is a nonempty compact set.

In the general case the proof is more involved: one can rely on "weak topology arguments" or, as here, on an *ad hoc* version of “weak compactness".

Lemma 25 (Weak compactness and convexity)**.** *Let $(C_n)_{n\in\mathbb{N}}$ be an increasing sequence of nonempty, closed, bounded, convex sets. Then*

$$\bigcap_{n\in\mathbb{N}} C_n \neq \emptyset.$$

Proposition 26 (Uniqueness of a minimizer)**.** *If C is convex and $f: H \to \mathbb{R}$ is* strictly *convex then*

$$\inf_C f$$

has at most one solution.

Theorem 27 (A generalized Fermat's rule). *Let C be a closed, convex, nonempty set and let $f\colon H \to \mathbb{R}$ be a differentiable convex function and consider the problem*

$$(P) \qquad \inf_C f$$

Take $x^* \in H$. Then the following are equivalent:

1. *x^* is a solution to the problem (P),*
2. *$x^* \in C$ and $\nabla f(x^*) + N_C(x^*) \ni 0$, i.e. $\nabla f(x^*) \in -N_C(x^*)$.*
3. *$x^* \in C$ and $\langle \nabla f(x^*), u\rangle \geqslant 0$ for all $u \in T_C(x^*)$.*

For nonconvex problems we can state the following:

Theorem 28. *Let $F \neq \emptyset$ be a closed subset of H. Let further $f\colon H \to \mathbb{R}$ be a Δ^1 function. Assume that x^* is a local minimizer of f over F, i.e.*

$$\exists r > 0 : f(x) \geqslant f(x^*) \quad \forall x \in B(x^*, r) \cap F.$$

Then

$$\langle \nabla f(x^*), u\rangle \geqslant 0 \quad \forall u \in T_F(x^*).$$

Remark 29. Note that in the case of a convex problem a point which is a local minimizer is a global minimizer.

5.2 Second order conditions

Theorem 30. *Let $f\colon H \to \mathbb{R}$ be a Δ^2 function. Let further $C \subseteq H$ be nonempty, closed and convex. Let $x^* \in C$. If*

1. *$\langle \nabla f(x^*), u\rangle \geqslant 0$ for all $u \in T_C(x^*)$,*
2. *$\langle \nabla^2 f(x^*)v, v\rangle \geqslant \alpha\|v\|^2$ for all $v \in T_C(x^*)$ with $\alpha > 0$,*

then x^* is a strict local minimum, i.e. there exists $r > 0$ such that $f(x) > f(x^*)$ for all $x \in B(x^*, r) \cap C$ and $x \neq x^*$.

5.3 Convex minimization

The following result is at the heart of constrained minimization with inequality constraints.

Lemma 31 (Farkas). *Let $a_1, \dots, a_m, b \in H$. The following assertions are equivalent:*

1. *For all $x \in H$,*
$$\Big(\langle a_i, x\rangle \leqslant 0, \forall i \in \{1, \dots, m\}\Big) \Rightarrow \langle b, x\rangle \leqslant 0.$$
2. *There exist $\lambda_1, \dots, \lambda_m \geqslant 0$ such that*
$$b = \sum_{i=1}^m \lambda_i a_i.$$

Its geometric interpretation is clear if we set $P_i := \{\, x \in H : \langle a_i, x\rangle \leqslant 0 \,\}$ and $P := \{\, x \in H : \langle b, x\rangle \leqslant 0 \,\}$. Then assertion 1 can be rewritten as

$$\bigcap_{i=1}^m P_i \subseteq P$$

while 2 can be written as *"b belongs to the closed convex cone generated by the a_i's"*.

It is a duality result in the sense that proving Farkas Lemma is equivalent to proving the identity

$$\left(\bigcap_{i=1}^m P_i\right)^* = \left\{\sum_{i=1}^m \lambda_i a_i : \lambda_i \geqslant 0, \forall i = 1, \dots, m\right\}.$$

Theorem 32 (Normal cone to an intersection of sublevel sets). *Let $g_1, \dots, g_m\colon H \to \mathbb{R}$ be Δ^1 convex functions. We assume that the* Slater *qualification condition holds, i.e. there exists $x_0 \in H$ such that*

$$\text{(Slater Q.C.)} \quad \begin{cases} \bullet\ g_i(x_0) \leqslant 0 \text{ for all } i \\ \text{such that } g_i \text{ is affine,} \\ \bullet\ g_i(x_0) < 0 \text{ otherwise.} \end{cases}$$

Set

$$\begin{aligned} C &:= \{\, x \in H : g_i(x) \leqslant 0, \quad \forall i = 1, \dots, m\} \\ &= \bigcap_{i=1}^m [g_i \leqslant 0]. \end{aligned}$$

Then, for all x in C, one has

$$N_C(x) = \left\{ \sum_{i \in I(x)} \lambda_i \nabla g_i(x) : \lambda_i \geqslant 0 \quad \forall i \in I(x) \right\}$$

where $I(x) := \{\, i \in \{1, \dots, m\} : g_i(x) = 0 \,\}$.

The set $I(x)$ is called the "active set".

6 KKT conditions for convex constrained problems

KKT stands for Karush-Kuhn-Tucker, it is the established name for first order conditions with inequality constraints ([2]). From a conceptual viewpoint it should be understood as a sophisticated Fermat's rule tailored for equality constrained problems.

[2] A more appropriate and meaningful name would be Farkas-Lagrange or Farkas/Lagrange conditions.

Theorem 33 (KKT). *Let $f, f_1, \ldots, f_m \colon H \to \mathbb{R}$ be convex and Δ^1. Consider the problem*

$$\begin{aligned} \inf f(x) \\ f_1(x) \leqslant 0 \\ \vdots \\ f_m(x) \leqslant 0 \end{aligned}$$

Assume further that the Slater qualification condition holds, i.e. there exists $x_0 \in H$ such that $f_i(x_0) \leqslant 0$ if f_i is affine, and $f_i(x_0) < 0$ otherwise. Then x^* solves the problem if and only if

1. *$f_i(x^*) \leqslant 0$ for all $i = 1, \ldots, m$, (Feasibility)*
2. *$\exists \lambda_1, \ldots, \lambda_m \geqslant 0$, (Nonnegativity)*
3. *$\lambda_i f_i(x^*) = 0$, (Complementary slackness)*
4. *$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla f_i(x^*) = 0$, (Lagrange conditions).*

The real numbers $\lambda_1, \ldots, \lambda_m$ are called *the Lagrange multipliers.*

Theorem 34. *Same assumptions as above. Let $c_1, \ldots, c_p \in H$, and $b_1, \ldots, b_p \in \mathbb{R}$. Consider the problem*

$$\begin{aligned} \inf f(x) \\ f_1(x) \leqslant 0 \\ \vdots \\ f_m(x) \leqslant 0 \\ \langle c_1, x \rangle = b_1 \\ \vdots \\ \langle c_p, x \rangle = b_p \end{aligned}$$

Let $x^* \in H$.
The point x^* is a solution to the problem if and only if

1. *$f_i(x^*) \leqslant 0$ for all i and $\langle c_j, x^* \rangle = b_j$ for all j.*

There exists $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_p$ such that

2. *$\lambda_1, \ldots, \lambda_m \geqslant 0$*
3. *$\lambda_i f_i(x^*) = 0$ for all $i = 1, \ldots m$.*
4. *$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla f_i(x^*) + \sum_{j=1}^p \mu_j c_j = 0$.*

Remark 35. (a) As previously $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_p$ are called Lagrange multipliers. Observe that a multiplier associated to equality constraints need not be nonnegative.
(b) The choice of linear equality constraints might seem exagerately demanding: one must yet understand that any other kind of equality constraints could make the constraints set nonconvex.

6.1 Lagrangians and duality theory

The framework: $(H, \langle \cdot, \cdot \rangle)$ is an Hilbert space, $f, f_1, \ldots, f_m \colon H \to \mathbb{R}$ are convex, Δ^1 functions, $c_1, \ldots, c_p \in H, b_1, \ldots, b_p \in \mathbb{R}$. We consider the problem.

$$\begin{aligned} \inf f(x) \\ f_1(x) \leqslant 0 \\ \vdots \\ f_m(x) \leqslant 0 \\ \langle c_1, x \rangle = b_1 \\ \vdots \\ \langle c_p, x \rangle = b_p \end{aligned}$$

The Lagrange condition described previously naturally leads to introduce the Lagrangian:

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p \mu_j(\langle c_j, x \rangle - b_j), \quad (14)$$

where $x \in H$, $\lambda_i \geqslant 0$, $\mu_j \in \mathbb{R}$, λ and μ are the vectors of multipliers.

Now observe that

$$\sup_{\lambda \geqslant 0,\, \mu \in \mathbb{R}^p} \mathcal{L}(x, \lambda, \mu) = \begin{cases} f(x), & \text{if } f_i(x) \leqslant 0 \text{ and } \langle c_j, x \rangle - b_j = 0 \\ +\infty & \text{otherwise.} \end{cases} \quad (15)$$

Thus, the problem becomes

$$V(P) = \inf_{x \in H} \sup_{\lambda \geqslant 0,\, \mu \in \mathbb{R}^p} \mathcal{L}(x, \lambda, \mu). \quad (16)$$

It is natural to exchange the order of sup/inf and to consider the *dual problem*

$$V(P^*) = \sup_{\lambda \geqslant 0,\, \mu \in \mathbb{R}^p} \inf_{x \in H} \mathcal{L}(x, \lambda, \mu), \quad (17)$$

and wonder whether $V(P) = V(P^*)$. We always have

$$V(P) \geqslant V(P^*),$$

the value $V(P) - V(P^*)$ is called the *duality gap*.

We shall see that in most convex instances there is no duality gap, so that the dual problem has the same value as the primal and offers thus an alternative and complementary way of solving the initial problem. For nonconvex problems estimating the duality gap is in general a very involved problem and little is known.

6.1.1 Min-max and saddle points

Let $C \subseteq H$ nonempty, and $D \subseteq F$ nonempty, with F being a Hilbert space. Let

$$\mathcal{L}\colon H \times F \to [-\infty, \infty].$$

Given our concerns we are led to consider

$$(\mathcal{P}) \qquad V(\mathcal{P}) = \inf_{x\in C} \sup_{y\in D} \mathcal{L}(x,y) \tag{18}$$

$$(\mathcal{P}^*) \qquad V(\mathcal{P}^*) = \sup_{x\in C} \inf_{y\in D} \mathcal{L}(x,y). \tag{19}$$

One easily sees that $V(\mathcal{P}^*) \geqslant V(\mathcal{P})$ is always satisfied. The converse inequality is not always true even when C, D are compact sets and $\mathcal{L}$ is smooth. However, as we shall see, convexity is a good means to ensure the coincidence of the above values.

Definition 16 (Saddle points)**.** We say that $(\bar{x}, \bar{y})$ is a saddle point of $\mathcal{L}$ on $C \times D$ if $\mathcal{L}(\bar{x}, \bar{y})$ is finite and

$$\mathcal{L}(\bar{x}, y) \leqslant \mathcal{L}(\bar{x}, \bar{y}) \leqslant \mathcal{L}(x, \bar{y}) \quad \forall (x,y) \in C \times D$$

Proposition 36 (Saddle points and minmax)**.** *A point $(\bar{x}, \bar{y}) \in C \times D$,*

$$\text{is a saddle point iff} \begin{cases} \bar{x} \text{ solves } \inf_{x\in C}\sup_{y\in D} \mathcal{L}(x,y), \\ \bar{y} \text{ solves } \sup_{y\in D}\inf_{x\in C} \mathcal{L}(x,y) \\ V(\mathcal{P}) = V(\mathcal{P}^*) \end{cases}$$

We point out the existence of the following abstract theorem (stated here in a weakened form), which is fundamental in Game Theory:

Theorem 37 (Sion)**.** *Let C, D be nonempty compact convex sets and*

$$\mathcal{L}\colon C \times D \to \mathbb{R}.$$

a continuous convex-concave function[3] *(i.e. $\forall y$, $\mathcal{L}(\cdot, y)$ is convex and $\forall x$, $\mathcal{L}(x, \cdot)$ is concave). Then $\mathcal{L}$ has a saddle point, as a consequence*

$$\min_{x\in C} \max_{y\in D} \mathcal{L}(x,y) = \max_{y\in D} \min_{x\in C} \mathcal{L}(x,y).$$

6.1.2 Convex duality

Theorem 38 (Convex duality)**.**
Let $f, f_1, \ldots, f_m\colon H \to \mathbb{R}$ be convex differentiable functions and let

$$\begin{cases} c_1, \ldots, c_p \in H \\ b_1, \ldots, b_p \in \mathbb{R} \end{cases}$$

Consider the problem

$$(P) \quad \inf f(x)$$

subject to

$$f_i(x) \leqslant 0 \quad \forall i = 1, \ldots, m$$
$$\langle c_j, x\rangle = b_j \quad \forall j = 1, \ldots, p.$$

The dual (P^*) is defined as previously:

$$(P^*) \quad \sup_{\lambda \geqslant 0,\, \mu} \inf_{x\in H} \mathcal{L}(x, \lambda, \mu)$$

where the Lagrangian is given by

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_j(\langle c_j, x\rangle - b_j).$$

Then

1. *$(\bar{x}, \bar{\lambda}, \bar{\mu})$ satisfies the KKT conditions*

$$\text{if and only if} \begin{cases} \bar{x} \text{ is a solution to } (P), \\ (\bar{\lambda}, \bar{\mu}) \text{ is a solution to } (P^*), \\ V(P) = V(P^*). \end{cases}$$

2. *Assume Slater condition holds and that (P) has a solution.*

   Then (P^*) has a solution, given by any of the Lagrange multipliers of (P), and

$$V(P) = V(P^*).$$

6.2 Sensitivity analysis

Theorem 39 (Envelope theorem)**.** *Let H and G be Hilbert spaces. Let further $g\colon H \times G \to \mathbb{R}$ be a Δ^1 function. Take $A \subseteq H$ nonempty. Set*

$$V(y) = \sup_{x\in A} g(x,y) \in (-\infty, \infty]. \tag{20}$$

Fix $y_0 \in G$. Assume that

(i) V is finite around y_0 and the sup *is achieved in $V(y_0)$ at some $x_0 \in A$.*

(ii) V is Δ^1 at y_0.

Then

$$V'(y_0) = \frac{\partial g}{\partial y}(x_0, y_0), \tag{21}$$

where x_0 is such that $g(x_0, y_0) = V(y_0)$.

[3] This could be weakened into convex lsc, concave usc

Theorem 40 (Sensitivity theorem). *Let $f, f_1, \dots, f_m \colon H \to \mathbb{R}$ be continuous, convex functions. Assume the Slater condition and assume also that there is an $r > 0$ such that $\bigcap_{i=1}^{m} [f_i \leqslant r]$ is bounded. For $y \in \mathbb{R}^m$ set*

$$V(y) = \inf_{f_i(y) \leqslant -y_i} f(x).$$

Then

(i) V is well-defined in a neighborhood of 0 and the infimum is achieved.

(ii) If V is differentiable at 0, then

$$\nabla V(0) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$$

is a solution to (P^*), the dual problem of (P).

This last theorem is absolutely fundamental. It may be understood as follows. Assume that the problem has been solved exactly but that there was a feasibility uncertainty for the first constraints, i.e. we only know that $[f_1 \leqslant -\epsilon]$ for a small $\epsilon \in \mathbb{R}$.

The obtained value is denoted V_ϵ while the true value, the one obtained with feasible points, is $V := V(0)$. Natural questions are:

- What is the impact of this small error on the final cost?
- Can we give a numerical estimate of the cost value error $V_\epsilon - V$?

The answer is obtained by computing the first Lagrange multiplier λ_1 (through the resolution of the dual problem, or through KKT) and by applying the result above. Using the definition of the derivative this gives

$$V_\epsilon - V = \epsilon \lambda_1 + o(\epsilon) \simeq \epsilon \lambda_1.$$

This remark justifies the fact that multipliers are sometimes called *shadow prices*. Shadow refers here to the fact that a naive optimizer only sees the primal problem and that small errors he may do on the constraints are priced according to a hidden process. The theorem above shows that this system of prices can actually be fully understood: it is itself the solution of another optimization problem called the dual problem.