Honours Independent Study: Convex Analysis & Nonlinear Optimization.

Zhe Chen

August, 2011

Contents

Notes on Convex Analysis

All of the action happens in $\mathbb{R}^n$ (n-dimensional space of real variables). We consider vectors (=points) $x, y, ... \in \mathbb{R}^n$, functions $f, g, ... : \mathbb{R}^n \mapsto \mathbb{R}$ and subsets $A, B, ... \subseteq \mathbb{R}^n$. The space $\mathbb{R}^n$ is equipped with the scalar product $\langle \cdot, \cdot \rangle$: $\forall x, y \in \mathbb{R}^2$ $\langle x, y \rangle = \sum x_k y_k$ and the norm $|x| = \sqrt{\langle x, x \rangle}$.

1 Basic concepts of convex analysis.

1.1 Affine sets and hyperplanes.

Def. 1.1.1 *The set $l_{xy} \equiv \{(1-\theta)\, x + \theta y | \theta \in \mathbb{R}\}$ is called "the line through x and y". The set M is called "affine" iff $l_{xy} \subseteq M$ for $\forall x, y \in M$. The operation on set $x + M = \{x + y | y \in M\}$ is called "translation of M by a". If L is a subspace then the set $x + L$ is called "parallel to L".*

Subspace is an affine set containing the origin. Every affine set is a translation of some subspace.

Def. 1.1.2 *Dimension of the affine set is the dimension of the parallel subspace. Affine sets of dimension $n-1$ are called "hyperplanes". The set $M^{\perp} = \{y | y \perp M\}$ is called "orthogonal complement of M."*

Prop. 1.1.3 *(Hyperplane representation)Hyperplanes are sets of the form $H_{b,\beta} = \{x | \langle x, b \rangle = \beta\}$.*

Proof. Supspaces of dimension $n-1$ are orthogonal complements of vectors. Hyperplanes are translations of such subspaces. $\blacksquare$

Prop. 1.1.4 *Affine sets have the form $A_{B,b} = \{x|Bx = b\}$ where B is a matrix and b is a vector. Consequently, affine sets are interesections of hyperplanes.*

Proof. If M is an affine set then $M = L + a$ for some subspace L. Let $\{b_k\}$ be the basis of $L^{\perp}$ then $L = \left(L^{\perp}\right)^{\perp} = \{x| \langle x, b_k\rangle = 0 \text{ for all } k\}$. We set $B = \{b_k\}$ as a union of columns and $b = Ba$. $\blacksquare$

Intersection of affine sets is an affine set. Hence, we introduce the affine hull as follows.

Def. 1.1.5 *(Affine hull). The affine hull of the set S is $aff\ S \equiv \cap_{\{all\ affine\ A\ s.t.\ S\subseteq A\}} A$.*

1.2 Convex sets and cones.

Def. 1.2.1 *(Convex set). The set $I_{xy} \equiv \{(1-\theta)\,x + \theta y|\theta \in \mathbb{R}\}$ is called "the line segment between x and y". The set C is called "convex" iff $I_{xy} \subseteq C$ for $\forall x, y \in C$. Dimension of the convex set is the dimension of its affine hull.*

Intersection of convex sets is a convex set. Consequently, for any collections of numbers $\{\beta_k\}$ and points $\{b_k\}$ the set $\cap_k \{x| \langle x, b_k\rangle \leq \beta_k\}$ is convex.

Def. 1.2.2 *(Convex hull). The convex hull $conv(S)$ of any set S is the intersection of all convex sets that contain S. If the collection of numbers $\{\lambda_k\}$ is such that $\sum_k \lambda_k = 1$ and $\lambda_k \geq 0$ then the sum $\sum_k \lambda_k b_k$ is called "the convex combination of points $\{b_k\}$".*

Prop. 1.2.3 *The convex hull of set S consists of all convex combinations of all elements of S.*

Def. 1.2.4 *(Convex Hull Cone Relative Interior). The set K is called a "cone" if it is closed with respect to positive scalar mutiplication: $\lambda x \in K$ for $\forall \lambda \geq 0$ and $\forall x \in K$. The convex cone "generated by the set S" and denoted "cone(S)" is the convex hull of all the lines joining all points of S with the origin. Let B denote a unit ball. The "closure" of the set C is the set $cl(C) \equiv \cap_{\varepsilon>0}(C + \varepsilon B)$. The "relative interior" is the set $ri(C) \equiv \{x|\exists\varepsilon > 0, (x + \varepsilon B) \cap aff\ C \subseteq C\}$.*

1.3 Convex functions and epigraphs.

Def. 1.3.1 *(Convex and proper function). The "epigraph" of a function f is the set epi$(f) \equiv \{(x,\mu)\,|x\in\mathbb{R}^n,\mu\geq f(x)\}$, see the picture (1.1). The function f is "convex" iff the set epi(f) is convex. The "effective domain" is the set dom$(f)=\{x|f(x)<+\infty\}$.*

The function is "proper" if the epigraph is nonempty and does not contain a vertical line.

The consideration of the entire notes on convex analysis is restricted to proper functions. Hence, all functions that are said to be convex are also presumed to be proper.

Prop. 1.3.2 *(Main property of convex function). A function f is convex iff $\forall x,y\in\mathbb{R}^n$, $\forall\theta\in[0,1]$ $f(\theta x+(1-\theta)y)\leq\theta f(x)+(1-\theta)f(y)$.*

Prop. 1.3.3 *A smooth function f is convex iff the matrix of second derivatives is non-negatively determined.*

Proof. Fix two points x_0 and x_1 and denote

$$x(\theta)\equiv\theta x_1+(1-\theta)x_0.$$

Let D be the matrix of second derivatives

$$D\equiv\left\|\frac{\partial^2 f}{\partial x_k\partial x_j}(x_0)\right\|_{k=1,\dots,n;\ j=1,\dots,n;}$$

taken at the point x_0. Note that

$$x(\theta)-x_0=\theta(x_1-x_0).$$

We use the Taylor decomposition

$$f(x_1)=f(x_0)+\langle\nabla_x f(x_0),x_1-x_0\rangle+\langle x_1-x_0,D(x_1-x_0)\rangle+o\left(\|x_1-x_0\|^2\right),$$
$$f(x(\theta))=f(x_0)+\theta\langle\nabla_x f(x_0),x_1-x_0\rangle+\theta^2\langle x_1-x_0,D(x_1-x_0)\rangle+o\left(\|x_1-x_0\|^2\right).$$

If we assume that the function is convex then we have, by the proposition (1.3.2),

$$\begin{aligned}&f(x(\theta))\leq(1-\theta)f(x_0)-\theta f(x_1)\Rightarrow\\&\qquad f(x_0)+\theta\langle\nabla_x f(x_0),x_1-x_0\rangle+\theta^2\langle x_1-x_0,D(x_1-x_0)\rangle+o\left(\|x_1-x_0\|^2\right)\\&\qquad\leq(1-\theta)f(x_0)\\&\qquad-\theta\left(f(x_0)+\langle\nabla_x f(x_0),x_1-x_0\rangle+\langle x_1-x_0,D(x_1-x_0)\rangle+o\left(\|x_1-x_0\|^2\right)\right).\end{aligned}$$

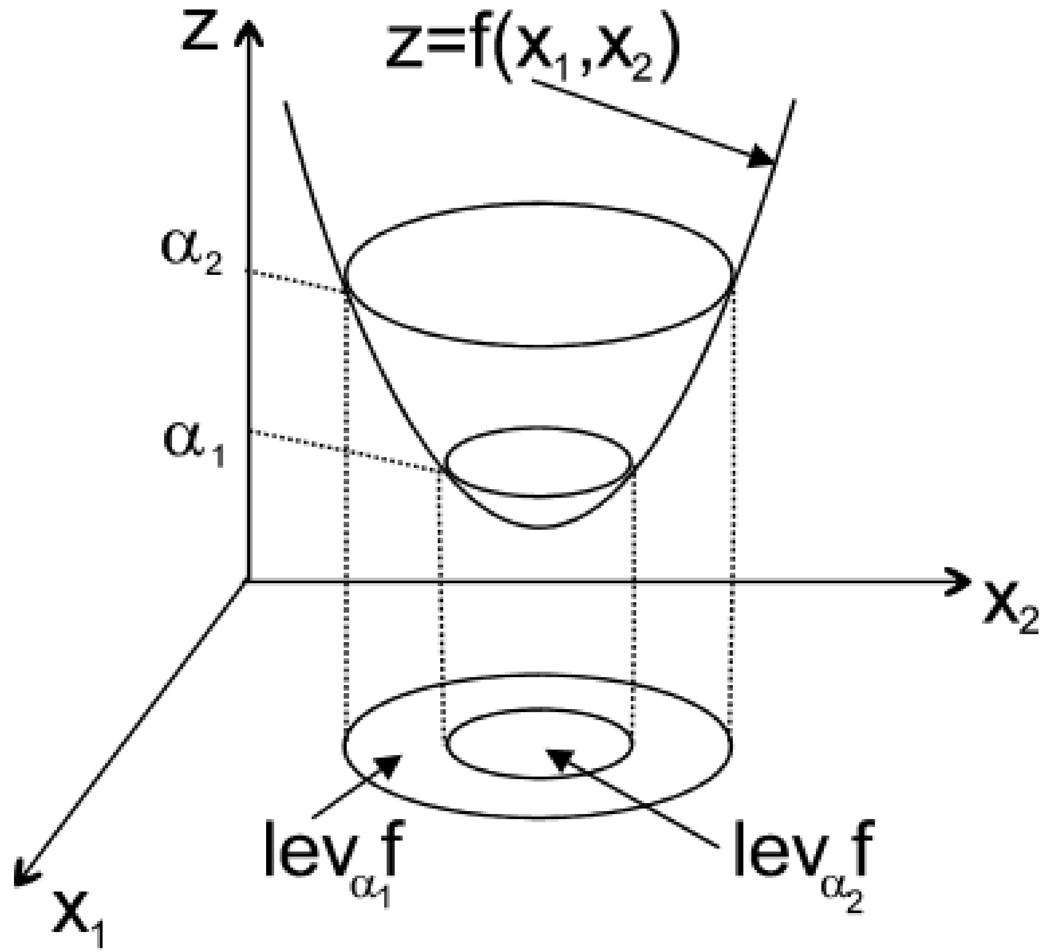

Figure 1.1: Convex function f acting from $\mathbb{R}^2$ to $\mathbb{R}$. Level sets $lev_\alpha(f)$.

Hence,

$$\theta^2 \langle x_1 - x_0, D(x_1 - x_0) \rangle \leq \theta \langle x_1 - x_0, D(x_1 - x_0) \rangle$$

for $\theta \in (0,1)$. Therefore,

$$0 \leq \langle y, Dy \rangle$$

for any $y = x_1 - x_0$. ■

Prop. 1.3.4 *(Preservation of convexity).*

1. If $\{f_i\}_{i=1,...,m}, f_i : \mathcal{R}^n \mapsto (-\infty, +\infty]$ are convex functions and $\{\lambda_i\}$ are positive real numbers then $\sum_i \lambda_i g_i$ is convex.

2. If $f : \mathcal{R}^n \mapsto (-\infty, +\infty]$ is a convex function and A is a matrix then $g(x) \equiv f(Ax)$ is convex.

3. If $\{f_i\}_{i \in I}, f_i : \mathcal{R}^n \mapsto (-\infty, +\infty]$ are convex functions and I is an arbitrary index set then $g(x) \equiv \sup_{i \in I} f_i(x)$ is convex.

2 Caratheodory's theorem.

Thm. 2.1 *Let X be a non-empty subset of $\mathcal{R}^n$.*

1. For every $x \in cone(X)$ there are linearly independent vectors $\{x_1, ..., x_m\}$, $x_i \in X$, $i = 1, ..., m$ such that

$$x = \sum_i \alpha_i x_i,$$

for some $\alpha_i \in \mathcal{R}$, $\alpha_i > 0$ and finite $m > 0$.

2. For every $x \in conv(X)$ there are vectors $\{x_1, ..., x_m\}$, $x_i \in X$, $i = 1, ..., m$ such that

$$x = \sum_i \lambda_i x_i,$$

for some $\lambda_i \in \mathcal{R}$, $\lambda_i > 0$, $\sum_i \lambda_i = 1$ and the vectors $\{x_2 - x_1, ..., x_m - x_1\}$ are lineraly independent.

Remark 2.2 *The Caratheodory theorem does not state that $\{x_1, ..., x_m\}$ might serve as a fixed basis. Indeed, on the picture (1.2) if the set X is open then for any pair of vectors x_1 and x_2 from X a point $x^* \in cone(X)$ may be found outside of the area span by the positively linear combinations of x_1 and x_2.*

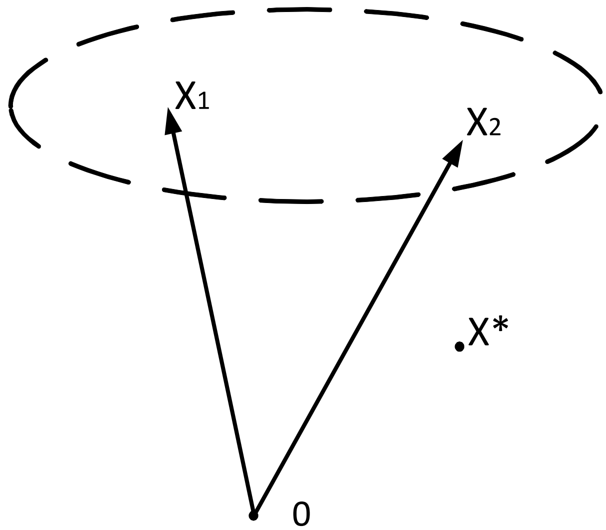

Figure 1.2: Caratheodory theorem remark.

Proof. 1. The definition of $cone\,(X)$ provides that there are some vectors $\{x_i\}\,, x_i \in X$ such that

$$x = \sum_i \alpha_i x_i.$$

If such vectors are linearly dependent then there are numbers β_i

$$0 = \sum_i \beta_i x_i.$$

We take a linear combination of two equalities

$$x = \sum_i (\alpha_i - \gamma\beta_i)\, x_i$$

and note that for at least one i the β_i is positive. Hence, a γ exists such that all $(\alpha_i - \gamma\beta_i)$ are non negative and $(\alpha_{i_0} - \gamma\beta_{i_0}) = 0$ for at least one index i_0. Hence, we decreased the number of terms in the sum. We continue this process until $\{x_i\}$ are linearly independent. ■

Proof. 2. The definition of $conv\,(X)$ provides that there are some vectors

$\{x_i\}\,,\ x_i \in X$ such that

$$x = \sum_i \lambda_i x_i,$$
$$1 = \sum_i \lambda_i,$$
$$\lambda_i \geq 0.$$

We consider $(x,1) \in R^{n+1}$ and restate the above conditions as

$$(x,1) = \sum_i \lambda_i (x_i,1)\,,$$
$$\lambda_i \geq 0.$$

Therefore, $(x,1) \in cone(\{(y,1)\,|y \in X\})$. The first part of the theorem applies and the vectors $(x_i,1)$ may be assumed linearly independent. Hence, no all non-zero $\{\beta_i\}\,,\beta_i \in \mathcal{R}$ exist such that

$$\sum_i \beta_i (x_i,1) = 0.$$

Equivalently,

$$\sum_{i\neq 1} \beta_i x_i + \beta_1 x_1 = 0,$$
$$\sum_{i\neq 1} \beta_i + \beta_1 = 0.$$

We express the β_1 from the second equation and substitute it into the first. We obtain the following consequence

$$\sum_{i\neq 1} \beta_i (x_i - x_1) = 0.$$

We conclude that no all non-zero $\{\beta_i\}$ exist such that the above is true. Hence, the $\{x_i - x_1\}$ are linearly independent. $\blacksquare$

3 Relative interior.

Prop. 3.1 *Let C be a nonempty convex set consisting of more then one point.*

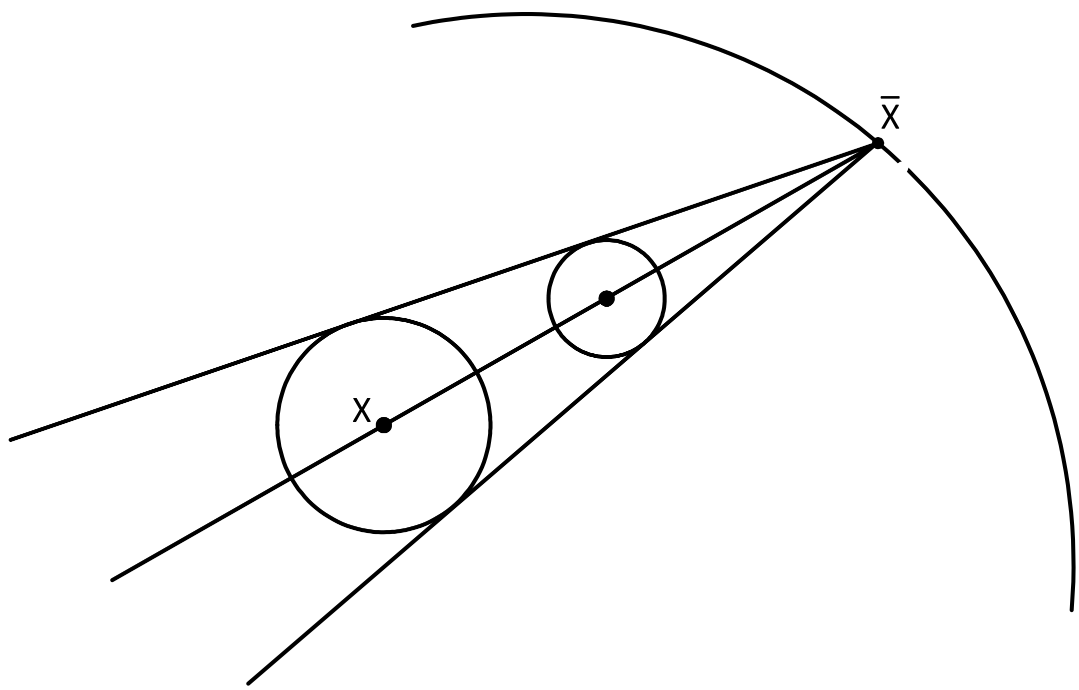

Figure 1.3: Relative interior.

a. (Non emptiness of relative interior). The $ri\,(C)$ *is not empty and* $aff\,(C) = aff\,(ri\,(C))$*. If* $m = \dim\,(aff(C)$ *then there are vectors* $x_0, ..., x_m \in ri\,(C)$ *such that* $x_1 - x_0, ..., x_m - x_0$ *span the subspace parallel to* $aff\,(C)$*.*

b. (Line segment principle). If $x \in ri\,(C)$ *and* $\bar{x} \in cl\,(C)$ *then all point of the line connecting* x *and* $\bar{x}$*, except possibly the* $\bar{x}$*, belong to* $ri\,(C)$*.*

Let a, $a \in C$ be a vector such that $aff\,(C - a)$ is a subspace. One can choose vectors $\{y_k\}$ such that $y_k \in C - a$, y_k are linearly independent and the linear span of $\{y_k\}$ is $aff\,(C - a)$. All convex combinations of $\{y_k\}$ belong to C and also belong to $ri\,(C)$. Hence, the $ri\,(C)$ is not empty. We construct $\{x_k\}$ as claimed in (a) by taking $x_k = y_k + a$. Consequently, $aff\,(C) = aff\,(ri\,(C))$.

The statement (b) is evident from the picture (1.3).

4 Recession cone.

Def. 4.1 *A vector y is a direction of recession of the set C iff for $\forall x \in C$, $\forall \lambda > 0$ we have $x + \lambda y \in C$.*

Directions of recession of a set C constitute a cone that we denote R_C. We introduce the notation

$$L_C = \{y|\ y \in R_C,\ \ -y \in R_C\}.$$

The L_C, if not empty, constitutes a subspace. We call it a "linearity space" of C.

Prop. 4.2 *(Main properties of direction of recession) Let C be a closed convex set.*

1. The vector y is a direction of recession if C contains $\{\lambda y + x|\lambda > 0\}$ for at least one $x \in C$.

2. C is either compact or has a direction of recession.

To see that the closedness is necessary consider the set $C_0 = \{(x, y)\,|x \in (0, 1)\,, y \geq 0\}$, see the figure (1.4). The only candidate for the direction of recession is $(0, 1)$. However, the point $(0, 0) \in C$ translates outside of C along $(0, 1)$.

Proof. 1. The statement (1) follows from the construction on the picture (1.5). We start from the point x and a direction of recession y. We take any point $\bar{x}$ and show that $\{\lambda y + \bar{x}|\lambda > 0\} \subset C$ as follows.

For small enough sphere around $\bar{x}$ if we take $\{z_n\}$, $z_n \to \infty$ then $\{a_n\}$ must be in C. Then the limit $A = \lim_n a_n$ is in C. Hence, $\{\lambda y + \bar{x}\} \subset C$ for small enough $\lambda > 0$.

We conclude that $\{\lambda y + \bar{x}|\lambda > 0\} \subset C$ for all λ by contradiction. If there is a finite $\bar{\lambda} = \arg\sup_\lambda \{\lambda y + \bar{x} \in C|\lambda > 0\}$ then we step back $x_0 = \left(\bar{\lambda} - \varepsilon\right) y + \bar{x}$ for small enough ε and build an 2ε-sphere around x_0 as in the first part of this proof.

2. Take a point $y_0 \in C$ and assume existence of $x_n \in C$ such that $x_n \to \infty$. A limit point of $\left\{\frac{x_n - y_0}{\|x_n - y_0\|}\right\}$ is a direction of recession. ■

Prop. 4.3 *(Recession cone of intersection). Let X and Y be closed convex sets and $X \cap Y \neq \varnothing$. Then $R_{X\cap Y} = R_X \cap R_Y$.*

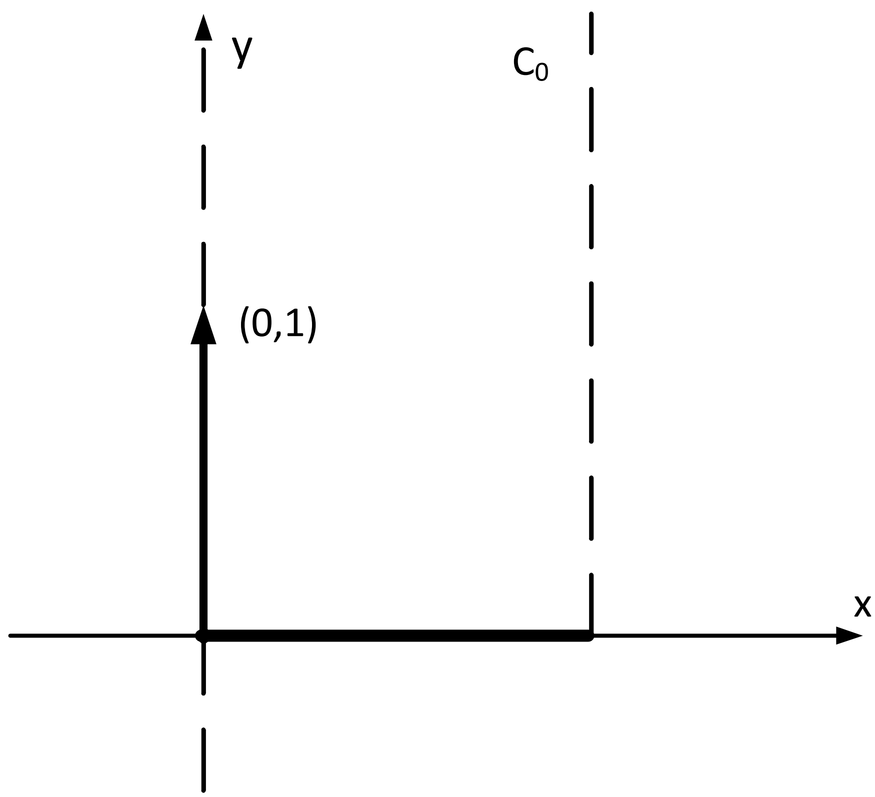

Figure 1.4: Closedness and recession

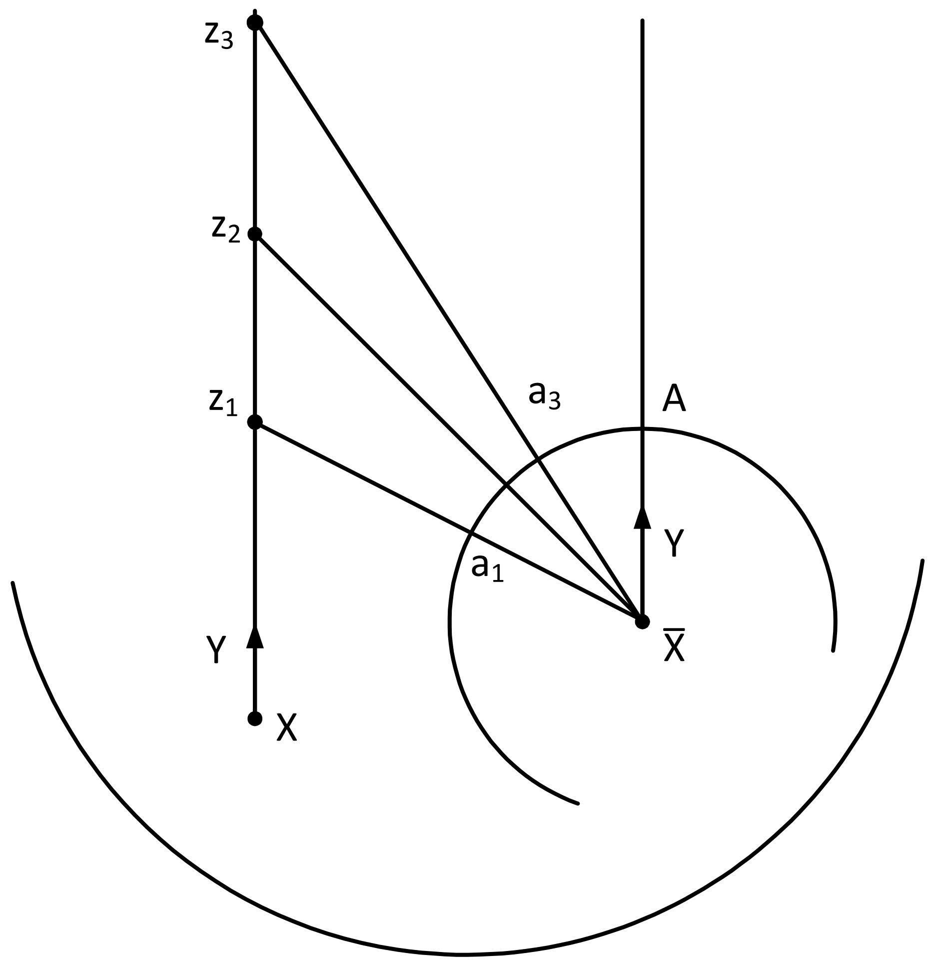

Figure 1.5: Direction of recession.

To see that the requirement $X \cap Y \neq \varnothing$ is necessary consider the sets $cl\,(C_0)$ and $cl\,(C_0) + (2,0)$ for $C_0 = \{(x,y)\,|x \in (0,1)\,, y \geq 0\}$, see the figure (1.4). These do not intersect but have a common direction of recession.

To see that the closedness is necessary consider C_0 and $ri\,(C_0)$. The intersection is $ri\,(C_0)$. It has a direction of recession $(0,1)$. The C_0 has no direction of recession.

Proof. The statement (4.3) follows from (4.2 -1) and the definitions.
■

Prop. 4.4 *(Recession cone of inverse image). Let C be a nonempty closed convex subset of $\mathcal{R}^n$, let A be a n by m matrix and let W be a nonempty convex compact subset of $\mathcal{R}^m$. Let the set*

$$V = \{x \in C | Ax \in W\}$$

be nonempty. Then

$$R_V = R_C \cap N\,(A)\,.$$

Proof. By definition of V we have $V = A^{-1}\,[W] \cap C$. The sets $A^{-1}\,[W]$ and C are convex and closed. Hence, the proposition (4.3) applies. It remains to note that $R_{A^{-1}(W)} = N\,(A)$. ■

Note that the compactness of W is important. In absence of compactness we cannot state that the $A^{-1}\,[W]$ is closed and we cannot state that $R_{A^{-1}(W)} = N\,(A)$.

Prop. 4.5 *(Decomposition of a convex set). For any subspace S contained in L_C for a non-empty convex set $C \subset \mathcal{R}^m$ we have*

$$C = S + \left(C \cap S^{\perp}\right).$$

Proof. Let S be a subspace contained in L_C. For any $x \in \mathcal{R}^m$ the affine set $x + S$ intersects $S^{\perp}$. If $x \in C$ then $x + S \subset C$. Hence, the intersection $(x + S) \cap \left(C \cap S^{\perp}\right)$ is not empty. Then $x = y + z$ for some $y \in S$ and $z \in (x + S) \cap \left(C \cap S^{\perp}\right)$. ■

5 Intersection of nested convex sets.

Intersection of nested closed compact sets is not empty.

Intersection of nested unbounded closed convex sets may be empty. Consider $C_k = \{(x,y)\,|y \geq k\}$, $\cap_k C_k = \varnothing$. The sets C_k escape to infinity along the common direction of recession $(0,1)$. However, if all directions of recession are included in a common linearity space then this cannot happen as stated in the proposition below.

Prop. 5.1 *(Principal intersection result). Let $\{C_k\}_{k=1,\ldots,\infty}$ be a sequence of nonempty closed convex sets, $C_k \subset \mathcal{R}^m$. Let R_k and L_k be the recession cone and linearity space of C_k and let*

$$R = \cap_k R_k,$$
$$L = \cap_k L_k.$$

If $C_{k+1} \subset C_k$ and $R = L$ then the intersection $\cap_k C_k$ is nonempty and

$$\cap_k C_k = L + \tilde{C}$$

for some nonempty compact set $\tilde{C}$.

Proof. Starting from some k the L_k has to stop decreasing because it is a space of finite dimension. After such k we have $L_k = L$. Let us restrict attention to such k.

Starting from some k we must have

$$R_k \cap L^{\perp} = \varnothing.$$

Indeed, if this is not so then for every k there is $y_k \in R_k \cap L^{\perp}$. We consider $z_k = \frac{y}{\|z_k\|} \in R_k \cap L^{\perp}$. The sets R_k are closed and nested. Hence, a limit point of such sequence $\{z_k\}$ has to be in R. This contradicts the condition $R = L$.

We now apply the result (4.5).

$$C_k = L + \left(C_k \cap L^{\perp}\right).$$

Hence,

$$\cap_k C_k = L + \cap_k \left(C_k \cap L^{\perp}\right).$$

Starting from some k the set $\left(C_k \cap L^{\perp}\right)$ has no direction of recession. Hence, the sets $C_k \cap L^{\perp}$ are nested and compact. We conclude $\tilde{C} = \cap_k \left(C_k \cap L^{\perp}\right) \neq \varnothing$. ■

Prop. 5.2 *(Linear intersection result). Let $\{C_k\}$ be a sequence of closed convex subsets of $\mathcal{R}^n$.*

Let the set X be given by the relationships

$$X = \{x | \langle a_j, x\rangle \leq b_j, \ j = 1, ..., r\}$$

where $a_j \in \mathcal{R}^n$ and $b_j \in \mathcal{R}$.

Assume that

1. $C_{k+1} \subset C_k$ for all k.

2. $X \cap C_k \neq \varnothing$ for all k.

3. $R_X \cap R \subset L$, where $R = \cup_k R_k$, $L = \cup_k L_k$, $R_k \equiv R_{C_k}$, $L_k \equiv L_{C_k}$.

Then $X \cap (\cap_k C_k) \neq \varnothing$.

To see that the X has to be linear conider $X = \{(x, y) | x > 0, xy \geq 1\}$ and $C_k = \left\{(x, y) | x \in \left[0, \frac{1}{k}\right]\right\}$. Such X and C_k fail only the linearity requirement and the conclusion of the theorem.

Proof. If $R_X \cap R = L_X \cap L$ then the statement is a consequence of the (5.1). We exclude such case from further consideration.

We consider the case when $R_X \cap R \neq L_X \cap L$. Since always $L_X \subset R_X, L \subset R$ and $R_X \cap R \subset L$ then there has to be a $y \in R_X \cap R$ that does not belong to L_X.

Let us take a sequence $\{x_k\}$ such that $x_k \in X \cap C_k$. Since the sets C_k are nested it is enough to prove the statement for some subsequence.

For any k we form the sum $x_k - \lambda y$, $\lambda \in \mathcal{R}$. Since $y \in R_X \cap R$ and $-y \notin R_X$ then for some $\bar{\lambda}_k$ the $x_k - \bar{\lambda}_k y = \bar{x}_k$ lies on the boundary of X. Hence, $\langle a_{j_k}, \bar{x}_k\rangle = b_{j_k}$. The X is given by a finite number of linear conditions. Hence, there is some j_0 such that $\langle a_{j_0}, \bar{x}_k\rangle = b_{j_0}$ for infinite number of k. We restrict our attention to such subsequence.

The set $\bar{X} = X \cap \{x | \langle a_{j_0}, x\rangle = b_{j_0}\}$ satisfies the conditions of the proposition within the subspace $\{x | \langle a_{j_0}, x\rangle = b_{j_0}\}$ and $aff\left(\bar{X}\right)$ is one dimension smaller then $aff\,(X)$. Therefore, we proceed by induction in the number of dimensions of $aff\,(X)$. The proposition is true for dimension 0 (X is a point). Then we assume that it is true for $\dim aff\,(X) = l$ and prove it for $\dim aff\,(X) = l+1$ using the construction above. Indeed, since the proposition holds for $\bar{X}$ then the intersection $\bar{X} \cap (\cap C_k)$ is not empty for the chosen subindexing of k. But $\bar{X} \subset X$ hence $X \cap (\cap_k C_k) \neq \varnothing$. ■

Prop. 5.3 *(Quadratic intersection result). Let $\{C_k\}$ be a sequence of subsets of $\mathcal{R}^n$ given by*

$$C_k = \{x|\, \langle x, Qx\rangle + \langle a, x\rangle + b \le w_k\}\,,$$

where Q is a symmetric positive semidefinite matrix, a is a vector, b is a scalar and w_k is a non-increasing sequence of real numbers converging to 0.

Let X be a subset of $\mathcal{R}^n$ of the form

$$X = \{x|\, \langle x, Q_j x\rangle + \langle a_j, x\rangle + b_j \le 0,\ \ j = 1, ..., r\}$$

where the Q_j are positive semidefinite matrixes.

Let $X \cap C_k$ be nonempty for all k.

Then the intersection $X \cap (\cap_k C_k)$ is nonempty.

Proof. The elements of recession cones and linear spaces of C_k are given by

$$R = \{y|Qy = 0,\ \ \langle a, y\rangle \le 0\}\,,$$
$$L = \{y|Qy = 0,\ \ \langle a, y\rangle = 0\}$$

and are k-independent.

If $R_X \cap R = L_X \cap L$ then the statement follows from the (5.1). Hence, we consider the situation $L_X \cap L \subset R_X \cap R$ and $R_X \cap R \ne \varnothing$. If there is a $y \in R_X \cap R$ then $a^T y \le 0$ and for any $x \in X,\ \lambda > 0$ we have $x + \lambda y \in X$. Note,

$$\langle x + \lambda y, Q\,(x + \lambda y)\rangle + \langle a, x + \lambda y\rangle + b = \langle x, Qx\rangle + \langle a, x\rangle + \lambda\,\langle a, y\rangle + b.$$

If $\langle a, y\rangle < 0$ then for a sufficiently large λ the $x + \lambda y$ lies in all C_k and in X and we are done.

Therefore, it remains to consider a situation when for any $y \in R_X \cap R$ we have $\langle a, y\rangle = 0 \Rightarrow y \in L$ but $y \notin L_X \Rightarrow -y \notin R_X$.

The recession cone of X is given by

$$R_X = \{y|Q_j y = 0,\ \ \langle a_j, y\rangle \le 0,\ j = 1, ..., r\}\,.$$

Hence, we are considering the case when for any $y \in R_X \cap R$ we have $\langle a, y\rangle = 0$ and $\langle a_j, y\rangle < 0$ for some j.

We now proceed by induction in the number of conditions r. For $r = 0$ the case that we are considering is excluded. Hence, we assume that the proposition holds for $\bar{r}$ and proceed to prove that it hold for $\bar{r}+1$. We are interested only in adding an equation with $\langle a_{\bar{r}+1}, y \rangle < 0$ because the all the equations with $\langle a_{\bar{r}+1}, y \rangle = 0$ may be arranged to be in the beginning of the induction and, hence, fall into the $R_X = L_X$ category.

A step of the induction in $\bar{r}$ proceeds in the following stages.

1. Assume that the statement holds for $\bar{r}$.

2. Let X be the $\bar{r}+1$-equations set.

3. Let $\bar{X}$ be the set holding $\bar{r}$ equations of X. The exclusion of a condition from X makes the $\bar{X}$ a bigger set. Hence, the conditions of the statement holds for the set $\bar{X}$ and $\bar{X} \cap (\cap_k C_k)$ is not empty.

4. We take a point $\bar{x} \in \bar{X} \cap (\cap_k C_k)$ and a direction $\bar{y} \in R_X \cap R$. In our case $a_j^T \bar{y} < 0$ for the one additional equation. Hence, we can construct $x = \bar{x} + \lambda \bar{y} \in X \cap (\cap_k C_k)$ by taking a sufficiently large λ. ■

6 Preservation of closeness under linear transformation.

The set $C = \{(x, y) \,|x > 0, xy \geq 1\}$ is a closed convex set. The projection on the x-axis is a linear transformation. The image of C under such transformation is open.

Prop. 6.1 *(Preservation of closeness result). Let C be a nonempty subset of $\mathcal{R}^n$ and let A be an $m \times n$ matrix.*

1. If $R_C \cap N(A) \subset L_C$ then the set AC is closed.

2. Let X be a nonempty subset of $\mathcal{R}^n$ given by linear constraints

$$X = \{x | \langle a_j, x \rangle \leq b_j, j = 1, ..., r\}.$$

If $R_X \cap R_C \cap N(A) \subset L_C$ then the set $A(C \cap X)$ is closed.

3. Let C is given by the quadratic constaints

$$C = \{x | \langle x, Q_j x \rangle + \langle a_j, x \rangle + b_j \leq 0, j = 1, ...r\}$$

where the Q_j are positive semidefinite matrices. Then the set AC is closed.

Proof. (1). Let $z \in cl\,(AC) \Rightarrow \forall \varepsilon > 0\ B\,(z,\varepsilon) \cap AC \neq \varnothing$, where the $B\,(z,\varepsilon)$ is the ball around z of radius ε. The sets $C_{\varepsilon_k} = A^{-1}\,[B\,(z,\varepsilon_k) \cap AC]$ are nested if $\varepsilon_k \downarrow 0$. It is suffice to prove that $\cap_{\varepsilon_k} C_{\varepsilon_k}$ is not empty for any sequence $\{\varepsilon_k\}\,, \varepsilon_k > 0, \varepsilon_k \downarrow 0$.

We have

$$C_\varepsilon = A^{-1}\,[B\,(z,\varepsilon) \cap AC] = \{x \in C \mid Ax \in B\,(z,\varepsilon)\}\,.$$

Therefore, by the proposition (4.4),

$$\begin{aligned} R_{C_\varepsilon} &= R_C \cap N\,(A)\,, \\ L_{C_\varepsilon} &= L_C \cap N\,(A)\,. \end{aligned}$$

Consequently, in the context of the proposition (5.1) for $\{C_{\varepsilon_k}\}$,

$$\begin{aligned} R &= R_C \cap N\,(A)\,, \\ L &= L_C \cap N\,(A)\,. \end{aligned}$$

Since, generally

$$L_C \subset R_C,$$

to accomplish the condition $R = L$ of the (5.1) it is enough to have

$$R_C \cap N\,(A) \subset L_C$$

as required by the theorem. ■

Proof. (2). Let $z \in cl\,(A\,[C \cap X])$. We introduce the sets $C_{\varepsilon_k} = A^{-1}\,[B\,(z,\varepsilon_k) \cap A\,[C \cap X]]$ for $\varepsilon_k \downarrow 0$ and aim to prove that the intersection $\cap_{\varepsilon_k} C_{\varepsilon_k}$ is not empty.

We have

$$C_\varepsilon = A^{-1}\,\{B\,(z,\varepsilon) \cap A\,[C \cap X]\} = \{x \in C \cap X \mid Ax \in B\,(z,\varepsilon)\}\,.$$

By the propositions (4.3) and (4.4)

$$\begin{aligned} R_{C_\varepsilon} &= R_C \cap R_X \cap N\,(A)\,, \\ L_{C_\varepsilon} &= L_C \cap L_X \cap N\,(A)\,. \end{aligned}$$

Consequently, in the context of the proposition (5.1) for $\{C_{\varepsilon_k}\}$,

$$\begin{aligned} R &= R_C \cap R_X \cap N\,(A)\,, \\ L &= L_C \cap L_X \cap N\,(A)\,. \end{aligned}$$

Since, generally

$$L_C \cap L_X \subset R_C \cap R_X,$$

to accomplish the condition $R = L$ of the (5.1) it is enough to have

$$R_X \cap R_C \cap N(A) \subset L_C.$$

■

Proof. (3). Let $z \in cl(AC)$. We introduce the sets $C_{\varepsilon_k} = A^{-1}[B(z, \varepsilon_k) \cap AC]$ for $\varepsilon_k \downarrow 0$ and aim to prove that the intersection $\cap_{\varepsilon_k} C_{\varepsilon_k}$ is not empty.

We have

$$\begin{aligned} C_\varepsilon &= A^{-1}\{B(z, \varepsilon) \cap AC\} \\ &= \{x \in C \mid Ax \in B(z, \varepsilon)\} \\ &= C \cap \left\{x \middle| \ \|Ax - z\|^2 \leq \varepsilon\right\} \\ &= C \cap \left\{x \middle| \ \left\langle x, A^T Ax\right\rangle - 2\langle z, Ax\rangle + \langle z, z\rangle \leq \varepsilon\right\}. \end{aligned}$$

We now apply the proposition (5.3) to conclude the proof. ■

7 Weierstrass Theorem.

A continuous function attains its minimum on a compact set. Such statement is the simplest version of the Weierstrass theorem. In this section we prove an extended version. We need some preliminary results and definitions.

Def. 7.1 *(Limit points) Let $\{x_k\}$ be a sequence of real numbers.*

1. Let $y_m = \sup\{x_k | k \geq m\}$, $z_m = \inf\{x_k | k \geq m\}$. We introduce the notation

$$\limsup_{k \to \infty} x_k \equiv \lim_{m \to \infty} y_m,$$
$$\liminf_{k \to \infty} x_k \equiv \lim_{m \to \infty} z_m.$$

If $\{x_k\}$ is not bounded from above then we write $\limsup_{k \to \infty} x_k = \infty$.

If $\{x_k\}$ is not bounded from below then we write $\liminf_{k \to \infty} x_k = -\infty$.

2. The point x_0 is a limit point of the sequence $\{x_k\}$ is there is an infinite number of points from $\{x_k\}$ in an ε-neibourhood of x_0 for any $\varepsilon > 0$.

Prop. 7.2 *The* $\limsup_{k\to\infty} x_k$ *is the greatest limit point of the sequence* $\{x_k\}$*. The* $\liminf_{k\to\infty} x_k$ *is the smallest limit point of the sequence* $\{x_k\}$*.*

Def. 7.3 *A function* $f : \mathcal{R}^n \mapsto [-\infty, +\infty]$ *is proper if its epigraph is nonempty and does not contain a vertical line. The function* f *is closed if the epi* (f) *is a closed set.*

Def. 7.4 *A function* f *is lower semicontinuous if for any* x *and* $\{x_k\}, x_k \to x$ *we have*

$$f(x) \leq \limsup_{k\to\infty} f(x_k).$$

Prop. 7.5 *(Closeness and lower semicontinuity). Let* f *be a function* $f : \mathcal{R}^n \mapsto [-\infty, +\infty]$*. The following statements are equivalent:*

1. *The level sets* $V_f(\gamma) = \{x|\ f(x) \leq \gamma\}$ *are closed for every* $\gamma \in \mathcal{R}$*.*
2. *The function* f *is lower semicontinuous.*
3. *The epi* (f) *is a closed set.*

Proof. (1) implies (2). Since the level sets are closed we have that for any sequence $\{x_k\}$ and vector x such that $x_n \to x$ and $f(x_k) \leq \gamma$ we must also have $f(x) \leq \gamma$. Assume that (2) is not true. Then there exists a y and $\{y_k\}$ such that $y_k \to y$ and $\limsup_k f(x_k) < \alpha < f(y)$ for some scalar α. This constitutes a contradition with the noted consequence of closeness of the level sets.

The rest may be proved with similar means. ■

Prop. 7.6 *(Weierstrass theorem). Let* $f : \mathcal{R}^n \mapsto (-\infty, +\infty]$ *be a closed proper function. If any of the below three conditions holds then the set* $\arg\min_{x\in\mathcal{R}^n} f(x)$ *is nonempty and compact.*

1. *dom* (f) *is bounded.*
2. *There exists an* $\gamma \in \mathcal{R}$ *such that the level set* $V_f(\gamma) = \{x|\ f(x) \leq \gamma\}$ *is nonempty and bounded.*
3. *If* $\|x_k\| \to \infty$ *then* $f(x_k) \to \infty$*.*

Proof. 1. Let $\{x_k\}$ be a sequence such that $f(x_k) \to \inf f$. Since *dom* (f) is bounded the sequence x_k has a limit point x^*. By proposition (7.5) the f is lower semicontinuous. Hence, $f(x^*) = \inf f$. Therefore,

$\arg\min f$ is nonempty. The $\arg\min f$ is an intersection of level sets. Hence, the compactness of $\arg\min f$ follows from the boundedness of $dom\,(f)$ and closeness of the level sets.

The (2) proves similarly to (1).

The (3) implies (2). ■

8 Local minima of convex function.

Prop. 8.1 *(Local minima of convex function). Let X be a convex subset of $\mathcal{R}^n$ and let $f : \mathcal{R}^n \mapsto (-\infty, +\infty]$ be a proper convex function. Then a local minimum is also a global minimum.*

Proof. If x_0 is a local minimum and x_1 is a different point and a global minimum then f takes values smaller then $f(x_0)$ at all points on the line $(x_0, x_1]$ because of the convexity. ■

9 Projection on convex set.

Prop. 9.1 *(Projection theorem). Let C be a nonempty closed convex set.*

1. For any $x \in \mathcal{R}^n$ there exists a unique vector

$$P_C(x) = \arg\min_{z \in C} \|z - x\|$$

called the projection of x on C.

2. The $P_C(x)$ could be defined as the only vector with the property

$$(y - P_C(x))^T (x - P_C(x)) \leq 0, \ \forall y \in C.$$

If the C is affine and S is a subspace parallel to C then the above may be replaced with

$$(x - P_C(x)) \in S^{\perp}.$$

3. The function $P_C(x)$ is continuous and nonexpansive:

$$\|P_C(x) - P_C(y)\| \leq \|x - y\|.$$

4. The distance function

$$d(x, C) = \min_{z \in C} \|x - z\|$$

is convex.

Proof. (1) follows from the theorem (7.6).

(2) We use notation $x_0 \equiv P_C(x)$. Clearly, x_0 has to lie on the boundary of C. Also, the x_0 has to satisfy the condition

$$\frac{\partial}{\partial \varepsilon} ||x_0 + \varepsilon z - x|| \geq 0$$

where the z is taken among all directions such that $x_0 + \varepsilon z$ remain in C for small $\varepsilon > 0$. The differentiation reveals that

$$\langle z, x_0 - x \rangle \geq 0.$$

For any $y \in C$ the difference $y - x_0$ is a valid z. Hence, the (2) follows.

(3) Since $P_C(x) \in C$ we can write from (2)

$$\begin{aligned}
&\langle P_C(y) - P_C(x), x - P_C(x) \rangle \leq 0, \\
&\langle P_C(x) - P_C(y), y - P_C(y) \rangle \leq 0.
\end{aligned}$$

We add the above and obtain

$$2 \langle P_C(y) - P_C(x), x - y + P_C(y) - P_C(x) \rangle \leq 0.$$

Hence,

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

(4) follows from (3) and definition of convexity. ■

10 Existence of solution of convex optimization problem.

Prop. 10.1 *(Directions of recession). Let $f : \mathcal{R}^n \mapsto (-\infty, +\infty]$ be a closed proper convex function.*

1. All nonempty level sets $V_f(\gamma)$ have the same recession cone given by

$$R_f \equiv R_{V_f} = \left\{ y | (y, 0) \in R_{epi(f)} \right\}.$$

2. If one nonempty level set is compact then all the level sets are compact.

Proof. Given a direction y and a point x the function $g(\lambda) \equiv f(x+\lambda y)$ is either nonincreasing or increasing starting from some large enough λ. If it is nonincreasing then y is in $R_{V_f(\gamma)}$ for any γ. The rest follows from the proposition (4.2). $\blacksquare$

Prop. 10.2 *(Basic existence result). Let X be a closed convex subset of R^n and let $f : R^n \mapsto (-\infty, +\infty]$ be a closed proper convex function such that $X \cap dom(f) \neq \varnothing$. The set $\arg\min_{x\in X} f(x)$ is nonempty and compact if and only if the X and f have no common directions of recession.*

If X and f has no common direction of recession then the minimum cannot escape to infinity. Such intuition may be formalized into a proof by considering intersections of the nested compact convex sets $C_k \equiv X \cap V_f(\gamma_k)$ with the sequence $\{\gamma_k\}$ converging to the $\inf_{x\in X} f(x)$. The following proposition is a consequence of the same observation and the propositions (5.1),(5.2) and (5.3).

Prop. 10.3 *(Unbounded existence result). Let X be a closed convex subset of $\mathcal{R}^n$ and let $f : \mathcal{R}^n \mapsto (-\infty, +\infty]$ be a closed proper convex function such that $X \cap dom(f) \neq \varnothing$. The set $\arg\inf_{x\in X} f(x)$ is nonempty if any of the following conditions hold.*

1. *$R_X \cap R_f = L_X \cap L_f$.*
2. *$R_X \cap R_f \subseteq L_f$ and X is given by the linear constraints*

$$X = \{x| \ \langle a_j, x\rangle \leq b_j, \ j = 1, ..., r\}$$

for some a_j, b_j.

3. *$\inf_{x\in X} f(x) > -\infty$ and f, X are of the form*

$$f(x) = \langle x, Qx\rangle + \langle c, x\rangle,$$
$$X = \{x| \langle x, Q_j x\rangle + \langle a_j, x\rangle + b_j \leq 0, \ j = 1, ..., r\}$$

where the Q, Q_j are positive semidefinite matrixes.

Remark 10.4 *The convex function f is constant on the subspace L_f.*

11 Partial minimization of convex functions.

Prop. 11.1 *(Convexity of partial minimum). Let $F : \mathcal{R}^{n+m} \mapsto [-\infty, +\infty]$ be a convex function. Then the function f given by*

$$f(x) = \inf_{z \in \mathcal{R}^m} F(x, z)$$

is convex.

The proof of the above proposition is a direct verification based on definitions.

The study of closeness of the partial minimum is based on the following observation.

Suppose the level set $V_f(\gamma) = \{x | \ f(x) \leq \gamma\}$ is nonempty for some γ. Let $\{\gamma_k\}$ be a sequence such that $\gamma_k \downarrow \gamma$. Then

$$\{x | \ f(x) \leq \gamma\} = \cap_{k=1}^{\infty} \{x | \ \exists (x, z) : \ F(x, z) \leq \gamma_k\}.$$

The set $\{(x, z) \mid F(x, z) \leq \gamma_k\}$ is closed if $F(x, z)$ is closed. The set $\{x | \ \exists (x, z) : \ F(x, z) \leq \gamma_k\}$ is a projection of $\{(x, z) \mid F(x, z) \leq \gamma_k\}$. Its closeness may be studied by means of the proposition (6.1). The intersection preserves the closeness. Hence, we arrive to the following proposition.

Prop. 11.2 *(Partial minimization result). Let $F : \mathcal{R}^{n+m} \mapsto [-\infty, +\infty]$ be a closed proper convex function. Then the function $f(x) = \inf_{z \in \mathcal{R}^m} F(x, z)$ is closed, convex and proper if any of the following conditions hold.*

0. There exist $\bar{x} \in \mathcal{R}^n$ and $\bar{\gamma} \in \mathcal{R}$ such that $U_F(\bar{x}, \bar{\gamma}) \equiv \{z | \ F(\bar{x}, z) \leq \bar{\gamma}\}$ is nonempty and compact.

1. There exist $\bar{x} \in \mathcal{R}^n$ and $\bar{\gamma} \in \mathcal{R}$ such that $U_F(\bar{x}, \bar{\gamma})$ is nonempty and $L_{U_F(\bar{x}, \bar{\gamma})} = R_{U_F(\bar{x}, \bar{\gamma})}$.

2. $F(x, z) = \left\{ \begin{array}{c} \bar{F}(x, z), (x, z) \in C, \\ +\infty, \ \text{otherwise}. \end{array} \right\}$, where the C is given by the linearity constraints

$$C = \{(x, z) \mid \langle a, (x, z) \rangle \leq b_j, \ j = 1, ..., r\},$$

and there exists $\bar{x}$ such that

$$R_{F(\bar{x}, \cdot)} = L_{F(\bar{x}, \cdot)}.$$

3. $F(x,z) = \left\{ \begin{array}{c} \bar{F}(x,z), (x,z) \in C, \\ +\infty, \text{ otherwise.} \end{array} \right\}$, *where the C is given by the quadratic constraints*

$$C = \{(x,z) \,|\, \langle (x,z), Q_j(x,z) \rangle + \langle a_j, (x,z) \rangle + b_j \le 0, \; j = 1, ..., r\},$$

where the Q_j are positive semidefinite and there exists $\bar{x}$ such that

$$-\infty < f(\bar{x}) < +\infty.$$

12 Hyperplanes and separation.

Def. 12.1 *Hyperplane in $\mathcal{R}^n$ is a set of the form*

$$H_{a,b} = \{x| \; \langle a, x \rangle = b\}.$$

The a is called the "normal vector". The sets

$$\{x| \; \langle a, x \rangle \ge b\}, \; \{x| \; \langle a, x \rangle \le b\}$$

are called "closed halfspaces" associated with $H_{a,b}$.

The two sets C_1 and C_2 are separated by $H_{a,b}$ if either

$$\langle a, x_1 \rangle \le b \le \langle a, x_2 \rangle, \; \forall x_1 \in C_1, \forall x_2 \in C_2$$

or

$$\langle a, x_1 \rangle \ge b \ge \langle a, x_2 \rangle, \; \forall x_1 \in C_1, \forall x_2 \in C_2.$$

The two sets C_1 and C_2 are strictly separated by $H_{a,b}$ if the above inequalities are strict.

A hyperplane $H_{a,b}$ may be represented as

$$H_{a,b} = \bar{x} + \{x| \; \langle a, x \rangle = 0\},$$
$$b = \langle a, \bar{x} \rangle$$

for any fixed $\bar{x} \in H_{a,b}$.

Prop. 12.2 *(Supporting hyperplane theorem). Let C be a nonempty convex subset of $\mathcal{R}^n$ and $\bar{x} \in \mathcal{R}^n$. If $\bar{x}$ does not belong to interior of C then there is a hyperplane that passes through $\bar{x}$ and contains C in one of its closed halfspaces:*

$$\begin{gathered}\exists H_{a,b}: \ \bar{x} \in H_{a,b}, \\ \langle a, \bar{x} \rangle \leq \langle a, x \rangle, \ \forall x \in C.\end{gathered}$$

Proof. If $\bar{x} \notin cl(C)$ then we obtain the normal vector by projecting on C:

$$a = P_C(\bar{x}) - \bar{x}.$$

The b may be obtained from the requirement that the $H_{a,b} \equiv H(x)$ pass trough $\bar{x}$.

If $\bar{x} \in cl(C)$ by does not belong to interior of C then there is a sequence $\{x_k\}$ such that $x_k \to \bar{x}$ and $x_k \notin cl(C)$. We utilise the construction from the case $\bar{x} \notin cl(C)$ to obtain a sequence $H(x_k) \equiv H_{a_k,b_k}$. The $\{a_k\}$ may be normalized to unity. The $\{a_k\}$ then has a limit point. Such limit point delivers the sought out hyperplane because of the proposition 9.1-2. ■

Prop. 12.3 *(Separating hyperplane theorem). If the C_1, C_2 are two nonempty disjoint convex sets then there is a hyperplane that separates them.*

Proof. Apply the proposition (12.2) to the set $C = C_1 - C_2$ and $\bar{x} = 0$. ■

Two nonempty convex disjoint sets C_1, C_2 are not necessarily strictly separated. For example, $C_1 = \{(x,y) \,| x \leq 0\}$, $C_2 = \{(x,y) \,| x > 0, y > 0, xy \geq 1\}$ do not have a strictly separating hyperplane.

Prop. 12.4 *(Strict hyperplane separation 1). Let C_1 and C_2 are two nonempty convex disjoint sets. If $C_1 - C_2$ is closed then there is a strictly separating hyperplane.*

Proof. Let $P_{C_1 - C_2}(0) = x_1 + x_2$, $x_1 \in C_1$, $x_2 \in C_2$. Set

$$a = \frac{x_2 - x_1}{2}, \ \bar{x} = \frac{x_2 + x_1}{2}, \ b = a^T \bar{x}.$$

By the closedness, $a \neq 0$. The $H_{a,b}$ strictly separates C_1, C_2. ■

Let C_1 and C_2 be two disjoint closed convex subsets of $\mathcal{R}^n$. To investigate the conditions for $C_1 - C_2$ to be closed we introduce the subset $C = C_1 \times C_2$ of $\mathcal{R}^{2n}$, note that the transformation $A\,(x_1, x_2) \mapsto x_1 - x_2$ is linear and seek to apply the proposition (6.1). We note that C is closed and convex,

$$R_C = R_{C_1} \times R_{C_2}$$

and

$$\begin{aligned} N\,(A) &= \{(x_1, x_2) \mid x_1 - x_2 = 0\} \\ &= \{(x_1, x_2) \mid x_1 = x_2\}\,. \end{aligned}$$

The condition $R_C \cap N\,(A) \subset L_C$ of the proposition (6.1) becomes

$$R_{C_1} \cap R_{C_2} \subset L_{C_1} \cap L_{C_2}.$$

We arrive to the following additional sufficient conditions for strict separation.

Prop. 12.5 *(Strict hyperplane separation 2). Let C_1 and C_2 are two nonempty convex disjoint sets. There is a strictly separating hyperplane if any of the following conditions holds.*

1. C_1 is closed and C_2 is compact.

2. C_1, C_2 are closed and $R_{C_1} \cap R_{C_2} = L_{C_1} \cap L_{C_2}$.

3. C_1 is closed, C_2 is given by the linearity constraints $C_2 = \{x \mid \langle a_j, x\rangle \leq b_j, j = 1, ..., r\}$ and $R_{C_1} \cap R_{C_2} \subset L_{C_1}$.

4. C_1 and C_2 are given by quadratic constraints

$$C_i = \{x \mid \ \langle x, Q_{ij}x\rangle + \langle a_{ij}, x\rangle + b_{ij} \leq 0, j = 1, ...r\}$$

where the Q_{ij} are positive semidefinite matrices.

Prop. 12.6 *(Intersection of halfspaces). The closure of convex hull of a set C is the intersection of all closed halfspaces that contain C.*

Proof. If there is a point in C that is not contained in the intersection of halfplanes then we arrive to contradition by using the theorem (12.2). ■

Def. 12.7 *The subsets C_1, C_2 of $\mathcal{R}^n$ are properly separated by a hyperplane $H_{a,b}$ if the following conditions are true*

$$\sup_{x_1 \in C_1} \langle a, x_1 \rangle \leq \inf_{x_2 \in C_2} \langle a, x_2 \rangle ,$$
$$\inf_{x_1 \in C_1} \langle a, x_1 \rangle < \sup_{x_2 \in C_2} \langle a, x_2 \rangle .$$

Let l be a line $l = \{x | \ x = x_0 + \lambda a, \ \lambda \in \mathcal{R}\}$ for some x_0. The definition of the proper separation requires that $P_l (C_1 \cap C_2)$ is a single point or nothing and $P_l (C_1 \cup C_2)$ consists of more then one point.

The sets $C_1 = \{(x,y) | x \in [0,2], y = 0\}$ and $C_2 = \{(x,y) | x \in [1,3], y = 0\}$ may not be properly separated.

The sets $C_1 = \{(x,y) | x \in [0,2], y = 0\}$ and $C_2 = \{(x,y) | x \in [0,1], y \in [0,1]\}$ are properly separated by the x-axis $H_{(1,0),0}$.

Prop. 12.8 *(Proper separation 1) Let C is a subset of $\mathcal{R}^n$ and $x \in \mathcal{R}^n$. There is a properly separating hyperplane for C and $\{x\}$ if*

$$x \notin ri(C).$$

Proof. If $x \notin aff(C)$ then $\{x\}$ and $aff(C)$ are strictly separated by the proposition 12.5-2.

If $x \in aff(C)$ then we translate $aff(C)$ into a subspace S and apply the 12.5-2 within $aff(C)$ to obtain some separating plane $\bar{H}$ then extend it to a hyperplane by $H = \bar{H} + S^{\perp}$. ∎

Prop. 12.9 *(Proper separation 2) The two subsets C_1, C_2 of $\mathcal{R}^n$ are properly separated if*

$$ri(C_1) \cap ri(C_2) = \varnothing.$$

Proof. Apply the proposition (12.8) to $C = C_1 - C_2$ and $x = 0$. ∎

13 Nonvertical separation.

Given a space $\mathcal{R}^{n+1}$ we can separate the last variable $(x,y) \in \mathcal{R}^{n+1}, y \in \mathcal{R}$ and call a hyperplane vertical if its normal vector is of the form $(x,0)$. A set $\{(x_0, \lambda) | \lambda \in \mathcal{R}\}$ for a fixed $x_0 \in \mathcal{R}^n$ is called a vertical line.

Prop. 13.1 *(Nonvertical separation). Let C be a nonempty convex subset of $\mathcal{R}^{n+1}$that contains no vertical lines. Then:*

1. The C is contained in a halfspace of a nonvertical hyperplane.

2. If $x \notin cl\,(C)$ then there is a nonvertical hyperplane that separates C and x.

Proof. 1. By contradition and proposition (12.6), if all halfspaces that surround C come from vertical hyperplanes then C must have a vertical line.

2. Consider $P_C(x) - x = a$. If a is not of the form $(\mu, 0)$ then we are done. If its is of the form $(\mu, 0)$ then we use a perturbation on the figure (1.6). First, take any hyperplane from the part (1) of the statement. There are no points of C in the part of the space below the broken plane (A,O,D). We perform a slight $\varepsilon-$perturbation the hyperplane (A,B) into that area while maintaining separation from the point x. ■

14 Minimal common and maximal crossing points.

Let M be a subset of $\mathcal{R}^{n+1}$. The M may have common points with the $(n+1)$-th coordinate axis. We introduce the quantity

$$w^*(M) = \inf_{(0,w)\in M} w.$$

A normal vector to a nonvertical hyperplane may be normalized to a form $(\mu, 1)$. A nonvertical hyperplane that crosses the $(n+1)$-th coordinate axis at the point $(0, \xi)$ and has a normal vector $(\mu, 1)$ has the representation

$$H(\mu, \xi) = \{(u, w) \mid w + \langle \mu, u \rangle = \xi\}.$$

Indeed,

$$\begin{aligned} H(\mu, \xi) &= \{(u, w) \mid \langle (u, w) - (0, \xi), (\mu, 1) \rangle = 0\} \\ &= \{(u, w) \mid \langle (u, w - \xi), (\mu, 1) \rangle = 0\} \\ &= \{(u, w) \mid \langle \mu, u \rangle + w - \xi = 0\}. \end{aligned}$$

The set M is contained in the upper half plane of $H(\mu, \xi)$ iff

$$\xi \le w + \langle \mu, u \rangle, \ \forall (u, w) \in M.$$

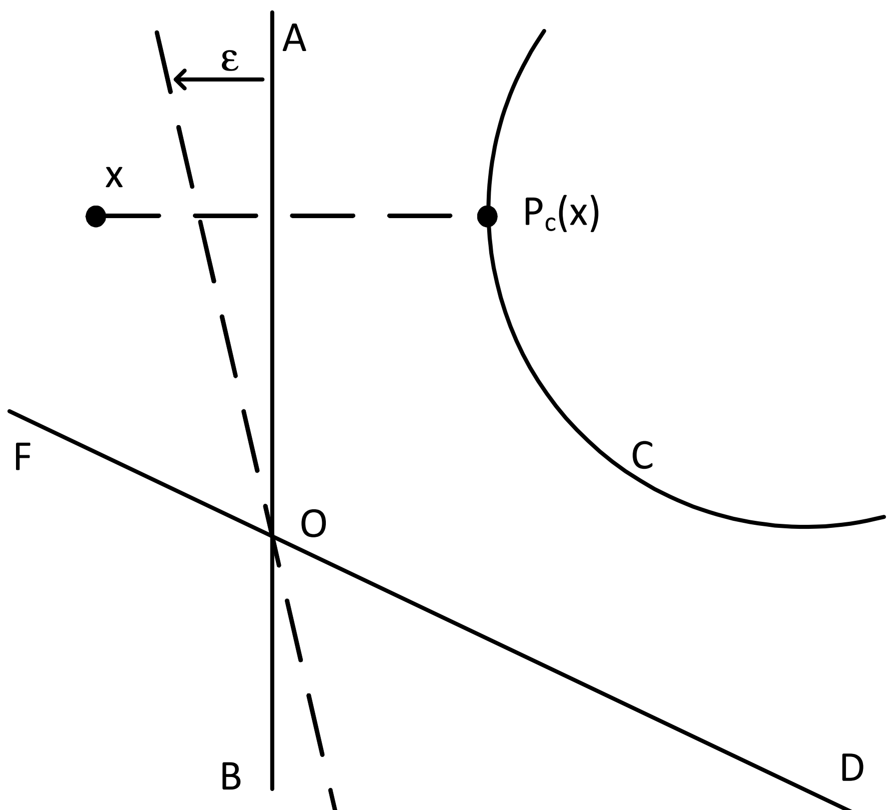

Figure 1.6: Nonvertical separation.

Hence, the quantity

$$q\left(\mu, M\right) = \inf_{(u,w)\in M} \left\{w + \langle \mu, u\rangle\right\}$$

is the maximum $(n+1)$-th axis crossing level for all hyperplances that contain the set M in the upper half space and have the normal vector $(\mu, 1)$.

The $q\left(\cdot, M\right)$ is a concave function.

We introduce the quantity

$$q^{*}\left(M\right) = \sup_{\mu\in\mathcal{R}^{n}} q\left(\mu, M\right).$$

Prop. 14.1 *(Weak duality theorem). Let M be a subset of $\mathcal{R}^{n+1}$. Then*

$$q^{*}\left(M\right) \leq w^{*}\left(M\right).$$

Proof. $q\left(\mu, M\right) = \inf_{(u,w)\in M}\left\{w + \langle \mu, u\rangle\right\} \leq \inf_{(0,w)\in M}\left\{w\right\} = w^{*}\left(M\right)$. ■

We investigate the conditions for the equality

$$q^{*}\left(M\right) = w^{*}\left(M\right).$$

Observe that by definition of these quantitites all that is needed is existence of a supporting hyperplane at a point $(0, w^{*})$. The pictures (1.7)-(1.9) show basic examples when this may or may not happen.

Prop. 14.2 *(Crossing theorem 1). Let M be a subset of $\mathcal{R}^{n+1}$. Assume the following:*

1. *M and $(n+1)$-th axis have nonempty intersection and $w^{*} \neq \infty$.*
2. *The set*

$$\bar{M} = \left\{(u, w)\,|\exists \bar{w} : \bar{w} \leq w, (u, \bar{w}) \in M\right\}$$

is convex.

Then $q^{*}\left(M\right) = w^{*}\left(M\right)$ if and only if for any sequence $\left\{(u_k, w_k)\right\} \subset M$ such that $u_k \to 0$ we have

$$w^{*} \leq \lim\inf_{k\to\infty} w_k.$$

Proof. By definition of w^{*}, $w^{*} \in cl\left(M\right) \subset cl\left(\bar{M}\right)$.

$\bar{M}$ contains no vertical lines. Indeed, if it does then by the proposition (4.2) one may infinitely go along the vector $(u_0, -1)$ inside $\bar{M}$ starting from any $(u_0, w) \in M$. This contradicts the condition 1.

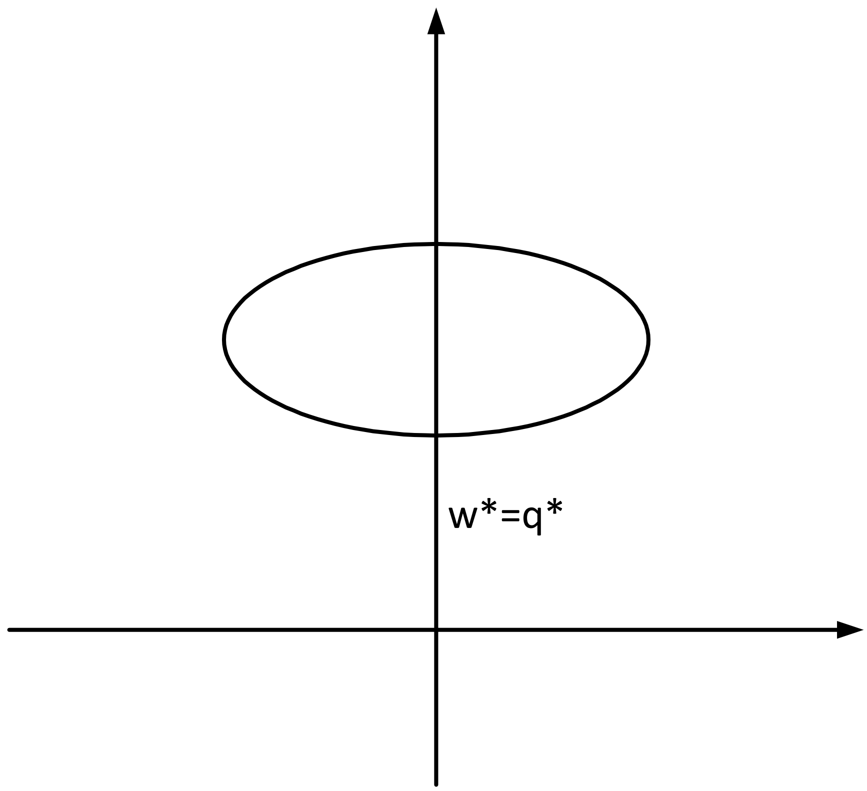

Figure 1.7: Crossing points figure 1

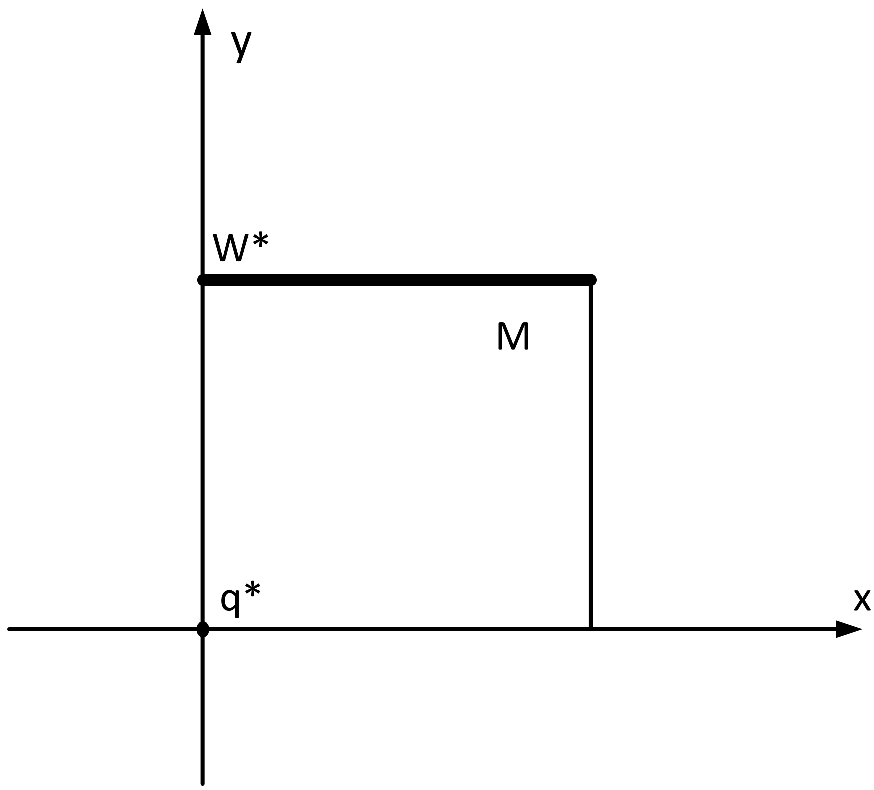

Figure 1.8: Crossing points figure 2. The upper boundary is included in the set. The other boundaries are excluded.

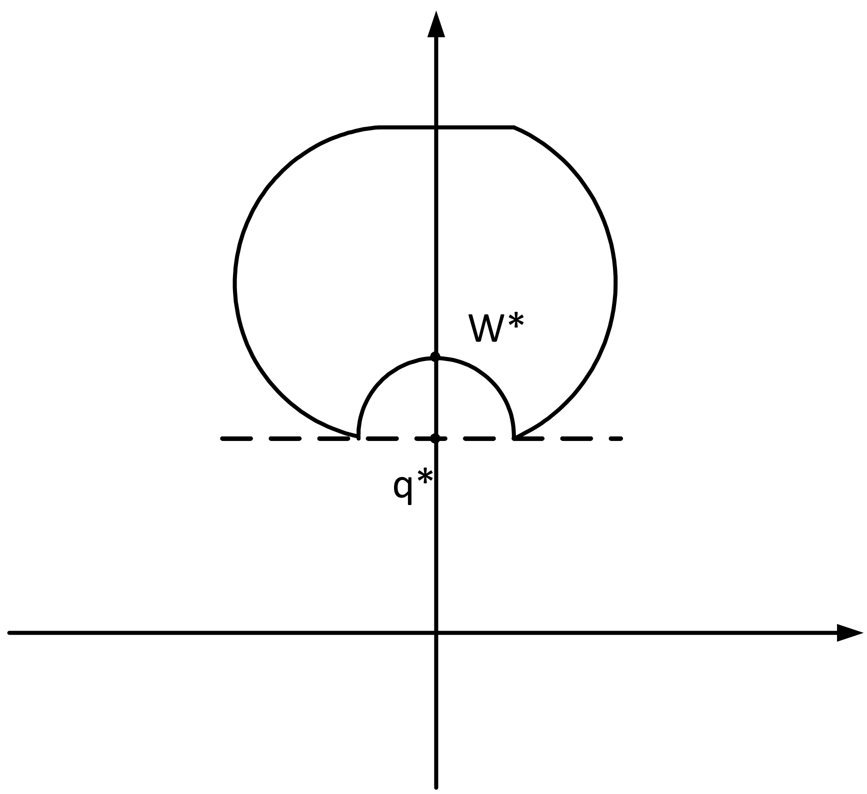

Figure 1.9: Crossing points figure 3

We have $(0, w^* - \varepsilon) \notin cl\left(\bar{M}\right)$ for any small positive ε. Indeed, on the contrary, if $(0, w^* - \varepsilon) \in cl\left(\bar{M}\right)$ then by definition of the closure one can construct the sequence $\{(u_k, w_k)\} \subset M$ that violates $w^* \leq \liminf_{k\to\infty} w_k$.

Therefore, by the proposition (13.1), there is a nonvertical separation of $\bar{M}$ from $(0, w^* - \varepsilon)$ for any small positive ε. The $(n+1)$-th axis crossing point for such separating hyperplane must be between $(0, w^*)$ and $(0, w^* - \varepsilon)$. Hence, $q^* = w^*$. ■

Prop. 14.3 *(Crossing theorem 2). Let M be a subset of $\mathcal{R}^{n+1}$. Assume the following:*

1. M and $(n+1)$-th axis have nonempty intersection and $w^* \neq \infty$.

2. The set

$$\bar{M} = \{(u, w) \,|\exists \bar{w} : \bar{w} \leq w, (u, \bar{w}) \in M\}$$

is convex.

3. $0 \in ri\,(D)$, where the set D is defined by

$$D = \left\{u \middle| \; \exists w \in \mathcal{R} : (u, w) \in \bar{M}\right\}.$$

Then $q^*(M) = w^*(M)$ and the solution set $Q^* = \{\mu |\; q(\mu) = q^*\}$ has the form

$$Q^* = (aff\,(D))^{\perp}_{\mathcal{R}^n} + \tilde{Q}$$

where the set $\tilde{Q}$ is nonempty convex and compact and $(aff\,(D))^{\perp}_{\mathcal{R}^n}$ is the orthogonal complement of $aff\,(D)$ relative to the plane of the first n coordinates $\{(u, 0)\,|u \in \mathcal{R}^n\}$.

Proof. By the proposition (12.8) there is a separating hyperplane H for the point $(0, w^*)$ and set $\bar{M}$. Such hyperplane cannot be vertical. Indeed, if it is vertical then the point $(0, w^*)$ projects on the plane $\{(u, 0)\,|u \in \mathcal{R}^n\}$ along the H onto the origin 0. Indeed, the segment

$$\left[(0, w^*), P_{\{(u,0)|u\in\mathcal{R}^n\}}\left((0, w^*)\right)\right] \equiv \left[(0, w^*), (0, 0)\right]$$

would belong to H. But then the condition $0 \in ri\,(D)$ is violated because it would belong to the boundary of D. Therefore the H is nonvertical, $q^*(M) = w^*(M)$ and Q^* is nonempty.

We next claim that $L_{Q^*} = (aff\,(D))^{\perp}_{\mathcal{R}^n}$. Indeed, by construction of D, if it has an orthogonal complement in $\{(u, 0)\,|u \in \mathcal{R}^n\}$ then we can rotate

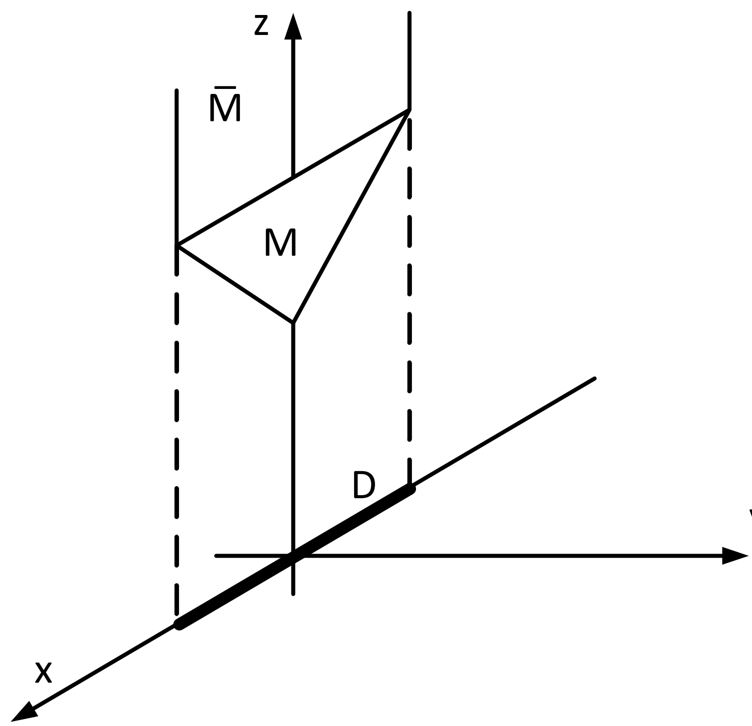

Figure 1.10: Crossing theorem 2 figure

coordinate system to make $(aff\,(D))^{\perp}_{\mathcal{R}^n}$ a coordinate subspace and then remove the coordinates that span the $(aff\,(D))^{\perp}_{\mathcal{R}^n}$ from the consideration (see the picture (1.10)).

In addition, $R_{Q^*} \cap aff\,(D) = \varnothing$. To see this, consider any hyperplane H, corresponding normal μ that delivers $q^*(M)$ and the perturbation $\mu + \lambda\eta$, $\eta \in aff\,(D)$, $\lambda \in \mathcal{R}$. If $\eta \in R_{Q^*}$ then $\mu + \lambda\eta$ can be made arbitrarily close to horizontal and H would be close to vertical by taking large enough $\pm\lambda$. Hence, such η can be in R_{Q^*} only if $ri\,(D) = \{0\}$. If $ri\,(D) = \{0\}$ then the statement is trivially true. We exclude such case from consideration.

We conclude that $L_{Q^*} = R_{Q^*} = (aff\,(D))^{\perp}_{\mathcal{R}^n}$.

We next apply the proposition (4.5) within the $\mathcal{R}^n$

$$Q^* = L_{Q^*} + \left(Q^* \cap (L_{Q^*})^{\perp}_{\mathcal{R}^n}\right)$$

with $L_{Q^*} = (aff\,(D))^{\perp}_{\mathcal{R}^n}$. The Q^* and $aff\,(D)$ have no common direction of recession as we already established. Hence,

$$Q^* = (aff\,(D))^{\perp}_{\mathcal{R}^n} + \tilde{Q}$$

for some convex and nonempty $\tilde{Q}$. The $\tilde{Q}$ is compact by 4.2-2. $\blacksquare$

15 Minimax theory.

Let ϕ be a function $\phi : X \times Z \mapsto \mathcal{R}$ where the X and Z are subsets of $\mathcal{R}^n$ and $\mathcal{R}^m$ respectively. We always have

$$\inf_{x\in X} \phi(x, z_0) \leq \inf_{x\in X} \sup_{z\in Z} \phi(x, z).$$

Therefore,

$$\sup_{z\in Z} \inf_{x\in X} \phi(x, z) \leq \inf_{x\in X} \sup_{z\in Z} \phi(x, z).$$

In this section we investigate the conditions for

$$\sup_{z\in Z} \inf_{x\in X} \phi(x, z) = \inf_{x\in X} \sup_{z\in Z} \phi(x, z) \tag{1.1}$$

and attainment of the sup and inf.

Def. 15.1 *The pair $(x^*, z^*) \in X \times Z$ is called a saddle point of ϕ iff*

$$\phi(x^*, z) \le \phi(x^*, z^*) \le \phi(x, z^*)$$

for $\forall x \in X, \ \forall z \in Z$.

Prop. 15.2 *(Saddle point's defining property). The pair (x^*, z^*) is a saddle point iff the relationship (1.1) holds and*

$$x^* \in \arg\inf_{x \in X} \left\{ \sup_{z \in Z} \phi(x, z) \right\},$$
$$z^* \in \arg\sup_{z \in Z} \left\{ \inf_{x \in X} \phi(x, z) \right\}.$$

We introduce the function $p : \mathcal{R}^m \mapsto [-\infty, +\infty]$ given by

$$p(u) = \inf_{x \in X} \sup_{z \in Z} \{\phi(x, z) - \langle u, z \rangle\}. \tag{1.2}$$

Prop. 15.3 *(Minimax lemma 1).Assume that $\phi(\cdot, z)$ is convex for each $z \in Z$. Then the function p is convex.*

Proof. The statement is a consequence of the propositions (13.4) and (11.1). ■

We will be using results of the section 14. Following that section we define

$$M = epi(p),$$
$$w^*(M) = p(0) = \inf_{x \in X} \sup_{z \in Z} \phi(x, z),$$
$$q(\mu, M) = \inf_{(u,w) \in M} \{w + \langle \mu, u \rangle\},$$
$$q^*(M) = \sup_{\mu} q(\mu, M).$$

Prop. 15.4 *(Minimax lemma 2). Let $\phi : X \times Z \mapsto (-\infty, +\infty)$ and $-\phi(x, \cdot)$ is closed and convex for every $x \in X$. Then*

1. $q(\mu) = \left\{ \begin{array}{c} \inf_{x \in X} \phi(x, \mu), \ \mu \in Z, \\ -\infty, \ \mu \notin Z. \end{array} \right\}$

2. $q^*(M) = w^*(M)$ *iff the relationship (1.1) holds.*

Proof. By definitions we have

$$
\begin{aligned}
q(\mu) &= \inf_{(u,w)\in epi(p)} \{w + \langle \mu, u\rangle\} \qquad (1.3)\\
&= \inf_{u\in dom(p)=\mathcal{R}^n} \{p(u) + \langle \mu, u\rangle\}\\
&= \inf_{u\in\mathcal{R}^n} \left\{ \inf_{x\in X} \sup_{z\in Z} \{\phi(x,z) - \langle u, z\rangle\} + \langle \mu, u\rangle \right\}\\
&= \inf_{u\in\mathcal{R}^n} \inf_{x\in X} \sup_{z\in Z} \{\phi(x,z) + \langle u, \mu - z\rangle\}.
\end{aligned}
$$

Since $\mu \in Z$ we nonincrease the last quantity by choosing $z = \mu$ among the $\sup_{z\in Z}$ values. We obtain

$$q(\mu) \geq \inf_{x\in X} \phi(x,\mu).$$

Next, we prove that $q(\mu) \leq \inf_{x\in X} \phi(x,\mu)$ when $\mu \in Z$.

Take any small $\varepsilon > 0$ and fix $z_0 \in Z$. Since the function $-\phi(x,\cdot)$ is convex then there is a separating hyperplane

$$H(\eta_x, c_x) = \{(z,w) \,|\, w + \langle \eta_x, z\rangle = c_x\}$$

between the point $(z_0, -\phi(x,z_0) - \varepsilon)$ and $epi(-\phi(x,\cdot))$. Hence, the point $(z_0, -\phi(x,z_0) - \varepsilon)$ lies below $H(\eta_x, c_x)$:

$$-\phi(x,z_0) - \varepsilon + \langle \eta_x, z_0\rangle < c_x$$

and the $epi(-\phi(x,\cdot))$ lies above $H(\eta_x, c_x)$:

$$\forall z \in Z,\ \ -\phi(x,z) + \langle \eta_x, z\rangle > c_x.$$

We combinte both inequalities into the statement

$$-\phi(x,z_0) - \varepsilon + \langle \eta_x, z_0\rangle < -\phi(x,z) + \langle \eta_x, z\rangle$$

where we claim existence of such η_x for any $z \in Z$. We transform the inequality as follows

$$\phi(x,z) + \langle \eta_x, z_0 - z\rangle < \phi(x,z_0) + \varepsilon.$$

We intend to combine this result with the expression (1.3) above. Hence, we set $z_0 = \mu$ and perform the operation $\inf_{\eta_x\in\mathcal{R}^n} \inf_{x\in X} \sup_{z\in Z}$. We obtain

$$q(\mu) < \inf_{x\in X} \phi(x,\mu) + \varepsilon.$$

Hence,

$$q(\mu) \le \inf_{x \in X} \phi(x, \mu), \mu \in Z.$$

Next, we prove that $q(\mu) = -\infty$ when $\mu \notin Z$. Indeed, if $\bar{\mu} \notin Z$ then for any $x \in X$ and any $\bar{w} \in \mathcal{R}$ the point $(\bar{\mu}, \bar{w})$ lies away from the epigraph of the convex function $-\phi(x, \cdot)$ of $z \in Z$. Hence, there is always a nonvertical hyperplane

$$H(\eta_x, c_x) = \{(z, w) \mid w + \langle \eta_x, z \rangle = c_x\}$$

that separates any $(\bar{\mu}, \bar{w}), \mu \notin Z$ from $epi(-\phi(x, \cdot))$ and the $epi(-\phi(x, \cdot))$ lies in the upper half plane.

$$\begin{gathered} \bar{w} + \langle \eta_x, \bar{\mu} \rangle \le c_x, \\ -\phi(x, z) + \langle \eta_x, z \rangle \ge c_x, \ \forall z \in Z. \end{gathered}$$

Hence,

$$-\bar{w} \ge \phi(x, z) + \langle \eta_x, \bar{\mu} - z \rangle$$

where we claim existence of such η_x and the statement holds for fixed $\bar{\mu} \notin Z$ and $\bar{w}$ and any $z \in Z$. Again, we apply the operation $\inf_{\eta_x \in \mathcal{R}^n} \inf_{x \in X} \sup_{z \in Z}$. Then the RHS becomes $q(\mu)$ and the LHS may be let to $-\infty$. We conclude that

$$q(\mu) = -\infty, \ \mu \notin Z.$$

With the representation

$$q(\mu) = \left\{ \begin{array}{c} \inf_{x \in X} \phi(x, \mu), \ \mu \in Z, \\ -\infty, \ \mu \notin Z. \end{array} \right\}$$

proven we remark that

$$q^*(M) = \sup_{\mu \in \mathcal{R}^n} q(\mu) = \sup_{\mu \in Z} \inf_{x \in X} \phi(x, \mu)$$

and

$$w^*(M) = p(0) = \left(\inf_{x \in X} \sup_{z \in Z} \{\phi(x, z) - \langle u, z \rangle\} \right)_{u=0} = \inf_{x \in X} \sup_{z \in Z} \phi(x, z).$$

Therefore the statement (2) of the proposition follows. ■

Prop. 15.5 *(Minimax theorem). Let X and Z be nonempty convex subsets of $\mathcal{R}^n$ and $\mathcal{R}^m$ respectively and let ϕ be a function $\phi : X \times Z \longmapsto \mathcal{R}$ such that*

1. For every $x \in X$ the function $-\phi(x, \cdot) : Z \longmapsto \mathcal{R}$ is convex and closed,

2. For every $z \in Z$ the function $\phi(\cdot, z) : X \longmapsto \mathcal{R}$ is convex,

3. $\inf_{x\in X} \sup_{z\in Z} \phi(x, z) < \infty$.

Then

1.The minimax equality

$$\sup_{z\in Z} \inf_{x\in X} \phi(x, z) = \inf_{x\in X} \sup_{z\in Z} \phi(x, z)$$

holds iff the function p given by the formula (1.2) is lower semicontinuous at 0.

2. If $0 \in ri\,(dom\,(p))$ then the minimax equality holds and the suppremum over Z in $\sup_{z\in Z} \inf_{x\in X} \phi(x, z)$ is finite and is attained. Furthemore,

$$\arg\sup_{z\in Z} \left[\inf_{x\in X} \phi(x, z) \right] \text{ is compact } \Leftrightarrow 0 \in ri\,(dom\,(p)).$$

Proof. The statement follows from propositions (15.3)-(15.4) and (14.2)-(14.3) applied to the epigraph of p. ■

16 Saddle point theory.

Let X and Z be nonempty convex subsets of $\mathcal{R}^n$ and $\mathcal{R}^m$ respectively and let ϕ be a function $\phi : X \times Z \longmapsto \mathcal{R}$. We introduce the following notations.

$$p(u) = \inf_{x\in\mathcal{R}^n} F(x, u),$$

$$F(x, u) = \left\{ \begin{array}{c} \sup_{z\in Z} [\phi(x, z) - \langle u, z \rangle], \; x \in X, \\ \infty, \; x \notin X, \end{array} \right\},$$

$$t_z(x) = \left\{ \begin{array}{c} \phi(x, z), x \in X, \\ \infty, \; x \notin X \end{array} \right\},$$

$$r_x(z) = \left\{ \begin{array}{c} -\phi(x, z), z \in Z, \\ \infty, \; x \notin X \end{array} \right\},$$

$$t(x) = \sup_{z\in Z} t_z(x), \; x \in \mathcal{R}^n,$$
$$r(z) = \sup_{x\in X} r_x(z), \; z \in \mathcal{R}^m.$$

The following statements are consequences of the propositions (15.5) and (11.2).

Prop. 16.1 *(Saddle point result 1) Assume that*

1. $\forall z \in Z$ the function t_z is convex and closed,

2. $\forall x \in X$ the function r_x is convex and closed,

3. $\inf_{x\in X} \sup_{z\in Z} \phi(x,z) < \infty$.

Then the minimax equality

$$\sup_{z\in Z} \inf_{x\in X} \phi(x,z) = \inf_{x\in X} \sup_{z\in Z} \phi(x,z)$$

holds and $X^* = \arg\inf_{x\in X}\left[\sup_{z\in Z} \phi(x,z)\right]$ is nonempty under any of the following conditions.

0. The level sets of the function t are compact.

1. The recession cone and the constancy space of the function t are equal.

2. The function $F(x,u)$ has the form

$$F(x,u) = \left\{ \begin{array}{c} \bar{F}(x,u), (x,u) \in C, \\ \infty, (x,u) \notin C \end{array} \right\}$$

with $\bar{F}$ being a closed proper convex function and set C being given by the linear constraints

$$C = \{(x,u) \,|Ax + Bu \leq b\}$$

and $R_C \subset L_{\bar{F}}$.

3. $-\infty < \inf_{x\in X} \sup_{z\in Z} \phi(x,z)$,

$$\phi(x,z) = \langle x, Qx\rangle + \langle c, x\rangle + \langle z, Mx\rangle - \langle z, Rz\rangle - \langle d, z\rangle,$$

where Q, R are symmetric matrices, Q is positive semidefinite, R is positive definite,

$$Z = \mathcal{R}^m,$$
$$X = \{x| \; \langle x, Q_j x\rangle + \langle a_j, x\rangle + b_j \leq 0, \; j = 1, ..., r\},$$

where the Q_j are positive semidefinite matrixes.

In addition, if (0) holds then X^* is compact.

Prop. 16.2 *(Saddle point result 2). Assume that*

1. $\forall z \in Z$ *the function* t_z *is convex and closed,*

2. $\forall x \in X$ *the function* r_x *is convex and closed,*

3. Either $\inf_{x\in X} \sup_{z\in Z} \phi(x,z) < \infty$ *or* $-\infty < \sup_{z\in Z} \inf_{x\in X} \phi(x,z)$.

Then

(a) If the the level sets of functions r *and* t *are compact then the set of saddle points of* ϕ *is nonempty and compact.*

(b) If $R_r = L_r$ *and* $R_t = L_t$ *then the set of saddle points of* ϕ *is nonempty.*

Prop. 16.3 *(Saddle point theorem). Assume that*

1. $\forall z \in Z$ *the function* t_z *is convex and closed,*

2. $\forall x \in X$ *the function* r_x *is convex and closed,*

Then the set of saddle points of ϕ *is nonempty and compact if any of the following conditions are satisfied*

1. X *and* Y *are compact.*

2. Z *is compact and* $\{x|\ x\in X,\ \phi(x,\bar{z}) \leq \gamma\}$ *is nonempty and compact for some* $\bar{z}\in Z$ *and* γ.

3. X *is compact and* $\{z|\ z\in Z,\ \phi(\bar{x},z) \geq \gamma\}$ *is nonempty and compact for some* $\bar{x}\in X$ *and* γ.

4. $\{x|\ x\in X,\ \phi(x,\bar{z}) \leq \gamma\}$ *and* $\{z|\ z\in Z,\ \phi(\bar{x},z) \geq \gamma\}$ *are nonempty and compact for some* $\bar{z}\in Z$, $\bar{x}\in X$ *and* γ.

17 Polar cones.

Prop. 17.1 *(Polar cone definition). For a nonempty set* C *we define the polar cone* C^**:*

$$C^* = \{y|\ \langle y,x\rangle \leq 0, \forall x\in C\}.$$

The following statement is a direct consequence of the definitions.

Prop. 17.2 *(Polar cone properties). For any nonempty set* C*, we have*

1. C^* *is a closed convex set.*

2. $C^* = (cl(C))^* = (conv(C))^* = (cone(C))^*$.

3. If $C \subset M$ *for some set* M *then* $M^* \subset C^*$.

Prop. 17.3 *(Polar cone theorem). For any nonempty cone C we have*

$$C^{**} = cl\left(conv\left(C\right)\right).$$

If C is closed and convex then $C^{**} = C$.

Proof. First, we show that for any nonempty C we have $C \subset C^{**}$. Indeed, by the definitions, for a fixed $x \in C$

$$x \in C \Rightarrow \forall y \in C^{*}, \langle x, y\rangle \leq 0$$

Therefore, $x \in C^{**} = \{z| \ \langle z, y\rangle \leq 0, \forall y \in C^{*}\}$.

Next, we prove that for a closed nonempty C, we have $C^{**} \subset C$.

Let $x \in C^{**}$. Since C is closed, there exists the projection $P_C\left(x\right)$. Let us translate the coordinate system so that $P_C\left(x\right) = 0$. Then by the proposition (9.1)-2 we have

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

Hence,

$$C \subset \{x\}^{*} \Rightarrow \{x\}^{**} \subset C^{*}.$$

We already established that

$$\{x\} \subset \{x\}^{**}.$$

Therefore,

$$x \in C^{*}$$

but also

$$x \in C^{**}.$$

Hence, for a nonempty set $M \equiv C^{*}$ (empty M is a trivial case) we have

$$x \in M \cap M^{*}.$$

By the definition of polar cone, we always have $M \cap M^{*} = \{0\}$ for a nonempty M. Hence, $x = 0 = P_C\left(x\right) \in C$.

Finally, we prove that $C^{**} = cl\left(conv\left(C\right)\right)$. By the proposition (17.2), we have

$$C^{*} = \left(cl\left(conv\left(C\right)\right)\right)^{*}.$$

Therefore,

$$C^{**} = \left(cl\left(conv\left(C\right)\right)\right)^{**}.$$

We already proved that $\left(cl\left(conv\left(C\right)\right)\right)^{**} = cl\left(conv\left(C\right)\right)$. ■

18 Polyhedral cones.

Def. 18.1 *A cone C is polyhedral if it has the form*

$$C = \{x | \ \langle a_j, x \rangle \leq 0, \ j = 1, ..., r\}.$$

A cone C is finitely generated if it has the form

$$C = cone(\{a_1, ..., a_r\}) = \left\{ x | \ x = \sum_{j=1}^{r} \mu_j a_j, \ \mu \geq 0, \ j = 1, ..., r \right\}$$

where $a_1, ..., a_r \in \mathcal{R}^n$.

Prop. 18.2 *(Polar polyhedral cone). Let $a_1, ..., a_r \in \mathcal{R}^n$. Then*

$$C = cone(\{a_1, ..., a_r\})$$

is closed and

$$C^* = \{y | \ \langle a_j, y \rangle \leq 0, \ j = 1, ..., r\}.$$

Proof. First we prove that $C^* = \{y| \ \langle a_j, y \rangle \leq 0, \ j = 1, ..., r\}$. Indeed, by the definition of polar cone

$$\begin{aligned} C^* &= \{y | \ \langle x, y \rangle \leq 0, \ \forall x \in C\} \\ &= \left\{ y | \ \sum_{j=1}^{r} \mu_j \langle a_j, y \rangle \leq 0, \ \forall \mu_j \geq 0 \right\} \\ &= \{y | \ \langle a_j, y \rangle \leq 0\}. \end{aligned}$$

Next, we prove that C is close by induction in r.

For $r = 1$ it is closed.

We assume that $C_r = cone(\{a_1, ..., a_r\})$ is closed and prove that $C_{r+1} = cone(\{a_1, ..., a_r, a_{r+1}\})$ is closed. Without loss of generality we assume $\|a_j\| = 1, \forall j$.

Take any sequence $\{x_k\}, x_k \in C_{r+1}, x_k \to x_0$. We aim to prove that $x_0 \in C_{r+1}$. We have

$$x_k = y_k + \lambda_k a_{r+1}, \ y_k \in C_r, \ \lambda_k \geq 0.$$

The λ_k must be a bounded sequence. Hence, we take a subsequence converging to some limit point and restrict consideration to such subsequence:

$$\lambda_k - \lambda_0 \to 0, \ \lambda_0 \geq 0.$$

We have

$$x_k = y_k + \lambda_0 a_{r+1} + (\lambda_k - \lambda_0)\, a_{r+1} \to x_0.$$

Therefore, y_k must be convergent:

$$y_k \to y_0$$

and $y_0 \in C_r$ by the induction hypothesis. Hence,

$$\begin{aligned} x_0 &= y_0 + \lambda_0 a_{r+1}, \ y_0 \in C_r, \ \lambda_0 \geq 0 \\ &\Rightarrow x_0 \in C_{r+1}. \end{aligned}$$

■

Prop. 18.3 *(Farkas lemma). Let*

$$P = \{y| \ \langle y, e_i \rangle = 0, i = 1, ..., m, \ \langle y, a_j \rangle \leq 0, \ j = 1, ..., r\},$$

$$C = \left\{ x| \ x \in \mathcal{R}^n, \ x = \sum_{i=1}^{m} \lambda_i e_i + \sum_{j=1}^{r} \mu_j a_j, \ \mu_j \geq 0, \ \lambda_i \in \mathcal{R} \right\}$$

where $e_1, ..., e_m, a_1, ..., a_r \in \mathcal{R}^n$*.*

Then

$$P^* = C.$$

Proof. Note that

$$\begin{aligned} C &= \left\{ x| \ x \in \mathcal{R}^n, \ x = \sum_{i=1}^{m} \lambda_i^+ e_i + \sum_{i=1}^{m} \lambda_i^- (-e_i) + \sum_{j=1}^{r} \mu_j a_j, \ \mu_j \geq 0, \ \lambda_i^+ \geq 0, \lambda_i^- \geq 0 \right\} \\ &= cone\left(\{e_1, -e_1, ..., e_m, -e_m, a_1, ..., a_r\}\right), \end{aligned}$$

$$\begin{aligned} P &= \{y| \ \langle b_k, y \rangle \leq 0, k = 1, ..., K\}, \\ \{b_k\} &= \{e_1, -e_1, ..., e_m, -e_m, a_1, ..., a_r\}. \end{aligned}$$

Therefore, by the proposition (18.2),

$$C^{*}=P$$

and C is closed. Hence,

$$C=C^{**}=P^{*}.$$

■

Prop. 18.4 *(Minkowski-Weyl theorem). A cone is polyhedral if and only if it is finitely generated.*

Proof. Suppose $C\subset\mathcal{R}^{n}$ is a finitely generated cone

$$C=cone\left(\{a_{1},...,a_{r}\}\right).$$

We prove that there exist vectors $\{b_{j}\}$ such that

$$C=\{y|\ \langle y,b_{j}\rangle\leq0,\text{ for some set of indexes }j\}.$$

Let H be a linear span of $\{a_{1},...,a_{r}\}$, and $k\equiv\dim H$. We introduce $\{e_{p}\}_{p=1,...,k}$ to be the orthogonal basis of H. Hence we have defined the linear transformations Λ and A as follows

$$\begin{aligned}\Lambda&=\|\lambda_{jp}\|,\\ a_{j}&=\sum_{p=1}^{k}\lambda_{jp}e_{p},\\ \vartheta&=\|\zeta_{pj}\|,\\ e_{p}&=\sum_{j=1}^{k}\zeta_{pj}a_{j}.\end{aligned}$$

The transformation ϑ is known as "orthogonalization". Some of its columns have all zero elements because $\{a_{j}\}$ might be linearly dependent.

We have

$$\begin{aligned}
C_r &= \left\{ x| \ x = \sum_{j=1}^{r} \mu_j a_j, \mu_j \geq 0 \right\} \\
&= \left\{ x| \ x = \sum_{j=1}^{r} \mu_j \sum_{p=1}^{k} \lambda_{jp} e_p, \mu_j \geq 0 \right\} \\
&= \left\{ x| \ x = \sum_{p=1}^{k} e_p \sum_{j=1}^{r} \mu_j \lambda_{jp}, \ \mu_j \geq 0 \right\} \\
&= \left\{ y| \ \ \langle y, e_p \rangle = \sum_{j=1}^{r} \mu_j \lambda_{jp}, \ \mu_j \geq 0 \right\}.
\end{aligned}$$

Let

$$\kappa_p = \sum_{j=1}^{r} \mu_j \lambda_{jp},$$

$$\begin{pmatrix} \kappa_1 \\ ... \\ \kappa_k \end{pmatrix} = \Lambda^T \begin{pmatrix} \mu_1 \\ ... \\ \mu_r \end{pmatrix}.$$

We introduce the vectors $\{z_p\}$:

$$\begin{pmatrix} z_1 \\ ... \\ z_r \end{pmatrix} = \vartheta^T \begin{pmatrix} e_1 \\ ... \\ e_k \end{pmatrix}$$

then

$$y^T z_p = y^T \left[\vartheta^T \begin{pmatrix} e_1 \\ ... \\ e_k \end{pmatrix} \right]_p = \left[\vartheta^T \begin{pmatrix} y^T e_1 \\ ... \\ y^T e_k \end{pmatrix} \right]_p$$

$$= \left[\vartheta^T \begin{pmatrix} \kappa_1 \\ ... \\ \kappa_k \end{pmatrix} \right]_p = \left[\vartheta^T \Lambda^T \begin{pmatrix} \mu_1 \\ ... \\ \mu_r \end{pmatrix} \right]_p = \mu_p.$$

Therefore,

$$C_r = \{y|\ \langle y, z_p\rangle \geq 0,\ \text{for } p \text{ such that } z_p \neq 0\}.$$

■

Def. 18.5 *A set P is a polyhedral set if it is nonempty and has the form*

$$P = \{x|\ \langle a_j, x\rangle \leq b_j,\ j = 1, ..., r\}.$$

Prop. 18.6 *(Minkowski-Weyl representation). A set P is polyhedral iff*

$$P = conv(\{v_1, ..., v_m\}) + cone(\{a_1, ..., a_r\})$$

for some $\{v_i\}_{i=1,...,m}, \{a_j\}_{j=1,...,r}$.

Proof. Note that the inequality

$$\langle a_j, x\rangle \leq b_j$$

may be represented as

$$\langle (a_j, -b_j), (x, 1)\rangle \leq 0.$$

Based on this observation we aim to apply the proposition (18.4). Set of the form $\{(y, 1)\}$ is not cone. We consider

$$\tilde{P} = \{(y, w) \,|\, \langle (a_j, -b_j), (y, w)\rangle \leq 0,\ w \geq 0\}.$$

Observe that $P = \left\{x|\ (x, 1) \in \tilde{P}\right\}$. By the proposition (18.4), we have

$$\tilde{P} = cone(\{\tilde{z}_1, ..., \tilde{z}_r\}).$$

We introduce the notation

$$\begin{aligned} &\tilde{z}_j = (u_j, w_j) \in \mathcal{R}^{n+1},\ u_j \in \mathcal{R}^n, w_j \in \mathcal{R}, \\ &J^0 = \{j|\ w_j = 0\},\ J^+ = \{j|\ w_j > 0\}. \end{aligned}$$

We have

$$\begin{aligned} P &= \left\{x|\ (x, 1) \in \tilde{P}\right\} \\ &= \left\{x|\ x = \sum_{j \in J^+ \cup J^0} \mu_j u_j,\ \sum_{j \in J^+} \mu_j = 1, \mu_j \geq 0\right\} \\ &= conv\left(\{u_j\}_{j \in J^+}\right) + cone\left(\{u_j\}_{j \in J^0}\right). \end{aligned}$$

■

Def. 18.7 *A function $f : \mathcal{R}^n \mapsto (-\infty, +\infty]$ is polyhedral if $epi\,(f)$ is polyhedral.*

The following proposition is a direct consequence of the definition.

Prop. 18.8 *(Polyhedral function). Let $f : \mathcal{R}^n \mapsto (-\infty, +\infty]$ be a convex function. Then f is polyhedral if and only if $dom\,(f)$ is polyhedral and*

$$f\,(x) = \max_{j=1,\ldots,m} \{\langle a_j, x\rangle + b_j\}\,.$$

19 Extreme points.

Def. 19.1 *For a nonempty convex set C the point x is an extreme point if there is no two points $u, w \in C$ such that $x \in (u, w)$. We denote $ep\,(C)$ the set of all extreme points.*

Prop. 19.2 *(Krein-Milman theorem). Let C be a nonempty convex set. Then*

1. For a hyperplane H that contains C in one of its closed halfspaces

$$ep\,(C \cap H) \subset ep\,(C)\,.$$

2. If C is closed then

$$L_C = \varnothing \Leftrightarrow ep\,(C) \neq \varnothing.$$

3. If C is compact then

$$C = conv\,(ep\,(C))\,.$$

Proof. (1). Assume the contrary: $\exists x :\ x \in ep\,(C \cap H),\ x \notin ep\,(C)$. Then there must be $u, w \in C :\ x \in (u, w)$. There are three cases:

a. $u, w \notin H$. Since $x \in (u, w)$ and $x \in H$ this means that H does not contain C in one of its halfspaces.

b. $u, w \in H$. Then $x \notin ep\,(C \cap H)$.

c. $u \in H, w \notin H$. Impossible because $x \in (u, w)$ and $x \in H$. ■

Proof. (2). If $L_C \neq \varnothing$ then any candidate x to be in $ep\,(C)$ may be translated in both directions along any $y \in L_C$ while remaining in C. Hence, $L_C \neq \varnothing \Rightarrow ep\,(C) = \varnothing$. We have $ep\,(C) \neq \varnothing \Rightarrow L_C = \varnothing$ for the same reason.

If $L_C = \varnothing$ then for any point $x \in C$ there is a direction y such that the line $\{z | z = x + \lambda y, \lambda \geq 0\}$ hits the relative boundary of C at some point x_0. By proposition (12.8) there is a properly separating hyperplane H at that point. By closedness of C the set $H \cap C$ is not empty and $L_{H \cap C} = \varnothing$. We reduced the dimensionality of our proof. Because of the part (1) of the proposition we can complete this proof by induction in the number of dimensions. ■

Proof. (3). We prove the statement by induction in the number of dimensions of $\mathcal{R}^n$. For $n = 1$ the statement is trivial. Assume that it is true in $\mathcal{R}^{n-1}$. Let $C \in \mathcal{R}^n$and $x \in C$, $x \notin ep(C)$. There is a line that passes through x such that $x \in [x_1, x_2]$ and $x_1, x_2 \in \partial\ ri(C)$. There are properly seprating hyperplanes H_1 and H_2 at points x_1 and x_2. By applying the statement in $\mathcal{R}^{n-1}$, $x_i \in conv(ep(C \cap H_i))$, $i = 1, 2$. Then by (1), $x_i \in conv(ep(C))$, i=1,2. Hence, $x \in conv(ep(C))$. ■

Prop. 19.3 *(Extreme points of polyhedral set 1). Let P be a polyhedral set. According to the proposition (18.6)*

$$P = conv(\{v_1, ..., v_m\}) + cone(\{a_1, ..., a_r\}).$$

We have

$$ep(P) \subset \{v_1, ..., v_m\}.$$

Proof. According to the proposition (18.6) any point $x \in P$ has the representation $x = y + z$, $y \in conv(\{v_1, ..., v_m\})$, $z \in cone(\{a_1, ..., a_r\})$. An extreme point $x^* \in ep(P)$ may not have a non zero z-part because it would contradict the definition of the extreme point. The x^* also cannot be a convex combination of $\{v_1, ..., v_m\}$. Therefore, the only possibility is $x^* \in \{v_1, ..., v_m\}$. ■

Prop. 19.4 *(Extreme points of polyhedral set 2). Let P be a polyhedral subset of $\mathcal{R}^n$. Then*

1. Let P has the form

$$P = \{x | \ \langle a_j, x \rangle \leq b_j, \ j = 1, ..., r\}$$

and denote

$$A_v \equiv \{a_j | \ \langle a_j, v \rangle = b_j, \ j \in \{1, ..., r\}\}$$

then

$$v \in ep(P) \Leftrightarrow \dim A_v = n \text{ and } v \in P.$$

2. Let P has the form

$$\begin{aligned} P &= \{x|\ x \geq 0,\ \ \langle a_j, x\rangle = b_j,\ \ j = 1, ..., r\} \\ &= \{x|\ x \geq 0,\ Ax = b\}, \\ a_j &\equiv (a_{jk})_{k=1,...,n} \in \mathcal{R}^n,\ \ j = 1, ..., r, \\ A &= ||a_{jk}||\,,\ \ b = (b_j) \end{aligned}$$

and denote

$$B_v = \|a_{jk}\|_{j=1,...,r;}^{k \in \{k|\ \ v_k \neq 0, k=1,...,n\}}$$

then $v \in ep\,(P)$ iff B_v has the maximal rank (all columns are linearly independent) and $v \in P$.

3. Let P has the form

$$\begin{aligned} P &= \{x|\ \ c \leq x \leq d,\ \ \langle a_j, x\rangle = b_j,\ \ j = 1, ..., r\} \\ &= \{x|\ \ c \leq x \leq d,\ \ Ax = b\},\ \ c, d \in \mathcal{R}^n, \\ a_j &\equiv (a_{jk})_{k=1,...,n} \in \mathcal{R}^n,\ \ j = 1, ..., r, \\ A &= ||a_{jk}||\,,\ \ b = (b_j) \end{aligned}$$

and denote

$$C_v = \|a_{jk}\|_{j=1,...,r;}^{k \in \{k|\ \ v_k \in (c_k, d_k), k=1,...,n\}}$$

then $v \in ep\,(P)$ iff C_v has the maximal rank (all columns are linearly independent) and $v \in P$.

Proof. (1). We state that $v \in ep\,(P)$ iff for any direction vector $y \in \mathcal{R}^n, \|y\| = 1$ such that

$$v + ty \in P$$

for some $t \neq 0$ and any small $\varepsilon > 0$ one of the conditions

$$v \pm \varepsilon y \in P$$

is violated. If $\dim A_v = n$ then no y can be orthogonal to all $a_j \in A_v$ and then one of the conditions $a_j^T\,(v \pm \varepsilon y) \leq b_j$ is violated. Hence,

$$\dim A_v = n \Rightarrow v \in ep\,(P)\,.$$

Conversely, if there is a y that is orthogonal to all A_v then for such y

$$\langle a_j, v + ty\rangle \leq b_j$$

if $v \in P$. Hence, $v \notin ep(P)$. $\blacksquare$

Proof. (2). We apply the part (1) of the proposition. In context of the part (1) the P is represented by

$$P = \{x| \ \langle a_j, x\rangle \leq b, \ \ -\langle a_j, x\rangle \leq -b_j, \ j = 1, ..., r, \ \ -e_k x \leq 0, \ k = 1, ..., n\}$$

where the $\{e_k\}$ is the coordinate basis. Therefore, the A_v for such situation has the form

$$A_v = \{a_j | \langle a_j, v\rangle = b_j, \ j \in \{1, ..., r\}\} \cup \{e_k | v_k = 0\}.$$

Note that $v \in P$ is given, hence, the condition $\langle a_j, v\rangle = b_j$ above is not restrictive. The set $\{e_k | v_k = 0\}$ contains linearly independent vectors. Let $k = \#\{e_k | v_k = 0\}$. We cannot state that according to (1), for $v \in ep(P)$ we need to have at least $n-k$ independent vectors among $\{a_j | \langle a_j, v\rangle = b_j, \ j \in \{1, ..., r\}\}$. Indeed, some of the a_j might be in the linear span of the $\{e_k | v_k = 0\}$. Hence, we need to exclude the projection on $\{e_k | v_k = 0\}$. For $v \in ep(P)$ we need to have

$$\dim\left\{a_j - P_{\{e_k|(v_k)=0\}} a_j\right\} = n - k$$

where the $P_{\{e_k\}}$ is the projection. The original matrix $||a_{jk}||$ has n columns in total. To establish the last equality it is enough to form a matrix from the columns $\{a_j\}$, remove the k columns that correspond to $(v_k) = 0$ and check that the remaining matrix has the maximal rank $n - k$. $\blacksquare$

Proof. The proof of (3) is the same as the proof of (2). $\blacksquare$

Prop. 19.5 *Let C be a closed convex set with at least one extreme point. A convex function $f : C \to \mathcal{R}$ that attains a maximum over C attains the maximum at some extreme point of C.*

Proof. Let S be a segment $S = \{x | x = a + \theta(b - a), \ \theta \in (0, 1)\}$. Note that S is open. A convex function that attains its maximum at S is constant on S. Such statement follows directly from the definition of convexity.

The proof is based on the above statement and the theorem (12.8). Let x^* be a point where the maximum is attained. By the above statement either f is constant on C or $x^* \notin ri(C)$. In the former case we are done. In the latter case there is a properly separating hyperplane H. Since $x^* \in C$ we have $x^* \in H \cap C$. If $H \cap C = \{x^*\}$ then we are done: the x^* is an extreme point. Otherwise we observe that we reduce the dimension of the proof by switching the consideration from C to $H \cap C$. $\blacksquare$

20 Directional derivative and subdifferential.

Prop. 20.1 *(Nondecreasing ratio). Let I be an interval of $\mathcal{R}$ and $f(x)$ is a convex function on I. The function*

$$h(x,y) = \frac{f(y) - f(x)}{y - x}$$

is nondecreasing in each argument.

Proof. Observe that $h(x,y) = h(y,x)$. Hence, we assume $x \leq y$ without loss of generality. We aim to show that $h(z,x) - h(y,x) \geq 0$ for $x \leq y \leq z$. There exists a θ such that $y = \theta x + (1-\theta) z$. We use such θ and the definition of convexity to calculate

$$\begin{aligned} h(z,x) - h(y,x) &= \frac{f(z) - f(x)}{z - x} - \frac{f(y) - f(x)}{y - x} \\ &= \frac{f(z) - f(x)}{z - x} - \frac{f(\theta x + (1-\theta) z) - f(x)}{\theta x + (1-\theta) z - x} \\ &\geq \frac{f(z) - f(x)}{z - x} - \frac{\theta f(x) + (1-\theta) f(z) - f(x)}{\theta x + (1-\theta) z - x} \\ &= \frac{f(z) - f(x)}{z - x} - \frac{(\theta - 1) f(x) + (1-\theta) f(z)}{(\theta - 1) x + (1-\theta) z} \\ &= 0. \end{aligned}$$

■

Def. 20.2 *(Left and right derivatives). Let f be a convex function on the interval $I \subseteq \mathcal{R}$. The left and right derivatives f^-, f^+ of f are defined by*

$$\begin{aligned} f^+(x) &= \inf_{\varepsilon > 0} \frac{f(x+\varepsilon) - f(x)}{\varepsilon}, \\ f^-(x) &= \sup_{\varepsilon > 0} \frac{f(x) - f(x-\varepsilon)}{\varepsilon}. \end{aligned}$$

Prop. 20.3 *(Properties of left and right derivative). Let I be an interval $I \subseteq \mathcal{R}$ and let f be a convex function on I.*

1. $f^-(x) \leq f^+(x)$, $x \in I$.

2. If $x \in ri\,(I)$ then $f^+(x)$ and $f^-(x)$ are finite.
3. If $x, y \in ri\,(I)$ and $x \le y$ then $f^+(x) \le f^-(y)$.
4. The functions f^+, f^- are nondecreasing.

Proof. The statements are consequences of the proposition (20.1). ■

Def. 20.4 *(Directional derivative). For a function $f : \mathcal{R}^n \to \mathcal{R}$ the directional derivative is defined by*

$$f'(x; y) = \inf_{\varepsilon > 0} \frac{f(x + \varepsilon y) - f(y)}{\varepsilon}.$$

Let f be a convex function $f : \mathcal{R}^n \to \mathcal{R}$. We use the notation $(x, z) \in \mathcal{R}^{n+1}$, $x \in \mathcal{R}^n$, $z \in \mathcal{R}$. Fix $x_0, d \in \mathcal{R}^n$. A hyperplane H that passes through the point $(x_0, f(x_0))$ and has the normal vector $(-d, 1)$ is given by the relationship

$$H = \{(x, z) \mid \langle (x_0, f(x_0)) - (x, z), (-d, 1) \rangle = 0\}.$$

Equivalently,

$$H = \{(x, z) \mid \langle d, x - x_0 \rangle + f(x_0) = z\}.$$

The $epi\,(f)$ lies above H iff

$$\text{for } \forall y \in \mathcal{R}^n, \ \ \langle d, y - x_0 \rangle + f(x_0) \le f(y)$$

or

$$\forall y \in \mathcal{R}^n, \ \ f(x_0) - \langle d, x_0 \rangle \le f(y) - \langle d, y \rangle. \tag{1.4}$$

Def. 20.5 *(Subgradient and subdifferential). The vector $d \in \mathcal{R}^n$ is a subgradient to the function f at $(x_0, f(x_0))$ iff the relationship (1.4) holds. The set of all subgradients at x_0 is called subdifferential at x_0 and denoted $\partial f(x_0)$.*

Prop. 20.6 *(Existence of subdifferential). Let $f : \mathcal{R}^n \to \mathcal{R}$ be a convex function. For any $x_0 \in \mathcal{R}^n$ the $\partial f(x_0)$ is nonempty, convex and compact set.*

Proof. We match the conditions of the present proposition with the setup of the proposition (14.3) as follows

$$M = \{(u,w) \mid f(x_0+u) \leq w\},$$
$$D = \mathcal{R}^n,$$
$$0 \in D.$$

Hence, according to the proposition (14.3)

$$w^* = q^*$$

where

$$w^* = \inf_{(0,w)\in M} w = f(x_0)$$

and q^* is the maximal crossing point of the hyperplanes $H(\mu, f(x_0)) = \{(u,w) \mid w + \langle \mu, u\rangle = f(x_0)\}$ such that the M lies above the hyperplane $H(\mu, f(x_0))$. Hence, there is a μ such that

$$\forall (u,w) \in M,\ w \geq f(x_0) - \langle \mu, u\rangle$$

or

$$\forall u \in \mathcal{R}^n, \quad f(x_0+u) \geq f(x_0) - \langle \mu, u\rangle.$$

Set $y = x_0 + u$ then

$$\forall y \in \mathcal{R}^n, \quad f(y) \geq f(x_0) - \langle \mu, y - x_0\rangle.$$

Hence, $-\mu$ is a subgradient. The rest of the conclusions follow from the conclusions of the proposition (14.3) and $D = \mathcal{R}^n$. ■

The following statements are verified with similar techniques.

Prop. 20.7 *(Properties of subgradient).*

1. Let $f : \mathcal{R}^n \rightarrow \mathcal{R}$ be a convex function. For any $x \in \mathcal{R}^n$ and any $y \in \mathcal{R}^n$ we have

$$f'(x;y) = \max_{d\in\partial f(x)} \langle y, d\rangle.$$

2. For convex functions $f_1, f_2 : \mathcal{R}^n \rightarrow \mathcal{R}$

$$\partial(f_1+f_2)(x) = \partial f_1(x) + \partial f_2.$$

3. For a $m \times n$ matrix A

$$\partial f(Ax) = A^T \partial f(Ax).$$

4. Let g be a smooth function $\mathcal{R} \to \mathcal{R}$ and $F(x) = g(f(x))$ then

$$F'(x; y) = \nabla g(f(x)) f'(x; y).$$

If g is convex and nondecreasing then

$$\partial F(x) = \nabla g(f(x)) \partial f(x).$$

Prop. 20.8 *Let $f : \mathcal{R}^n \to (-\infty, +\infty]$ be a proper convex function then*

$$\forall x \in ri(dom(f)) : \quad \partial f(x) = S^\perp + G$$

where the S is a subspace parallel to $aff(dom(f))$ and G is a nonempty compact set. Furthemore, $\partial f(x)$ is nonmepty and compact iff x is in interior of $dom(f)$.

Proof. The proof of the proposition (20.5) applies almost without changes. ∎

21 Feasible direction cone, tangent cone and normal cone.

Def. 21.1 *Let X be a subset of $\mathcal{R}^n$ and x be a point in X.*

(Feasible direction cone). The feasuble direction cone $F_X(x)$ of X at x is defined as follows.

$$F_X(x) = \{y |\ \exists \varepsilon_0 > 0 \ s.t.\ \forall \varepsilon \in (0, \varepsilon_0),\ x + \varepsilon y \in X\}.$$

(Tangent cone). The tangent cone $T_X(x)$ of X at x is defined as follows

$$T_X(x) = \{0\} \cup \left\{ y |\ y \neq 0, \exists \{x_k\} \subset X \ s.t.\ x_k \neq x \text{ and } x_k \to x,\ \frac{x_k - x}{\|x_k - x\|} \to \frac{y}{\|y\|} \right\}.$$

(Normal cone). The normal cone $N_X(x)$ of X at x is defined as follows

$$N_X(x) = \{z |\ \exists \{x_k\} \subset X, \{z_k\} \ \ s.t.\ z_k \in T_X(x_k)^*, x_k \to x, z_k \to z\}.$$

(Regularity of a set). By definition, the X is regular at x if

$$N_X(x) = T_X(x)^*.$$

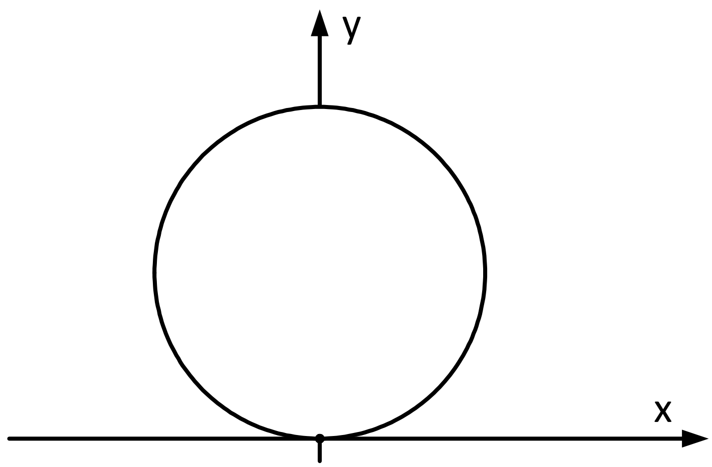

Figure 1.11: Tangent cone figure 1

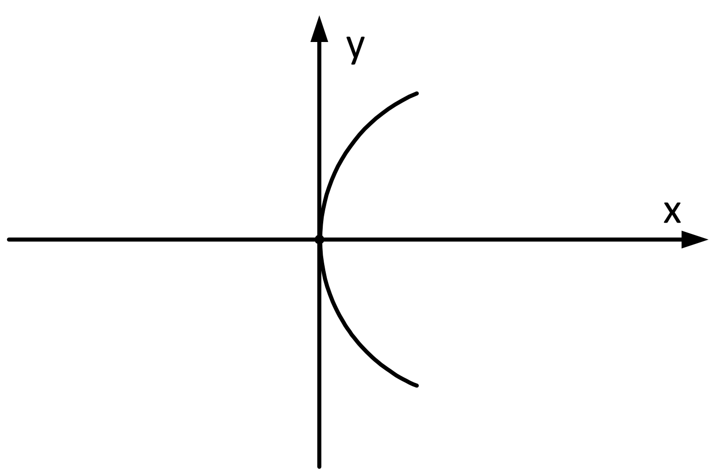

Figure 1.12: Tangent cone figure 2

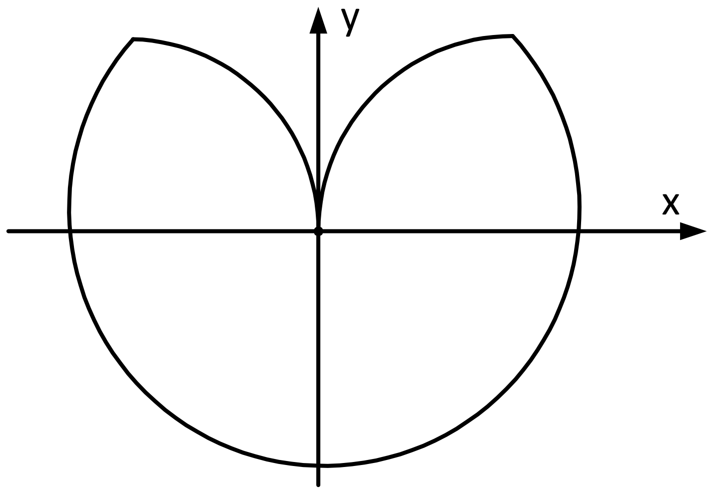

Figure 1.13: Normal cone figure 1

On the figure (1.11) the X is the closed area bounded by the circle, the x is the origin, the $F_X(x) = \{(x, y) | y > 0\}$ and $T_X(x) = \{(x, y) | y \geq 0\}$.

On the figure (1.12) the X is the curved line, the x is the origin, the $F_X(x) = \{(0,0)\}$ and $T_X(x) = \{(0, y) | y \in (-\infty, +\infty)\}$.

On the figure (1.13) the X is the closed area bounded by the curved shape, the x is the origin, $F_X(x) = \mathcal{R}^2 \setminus \{(0, y) | y \in (0, +\infty)\}$, $T_X(x) = \mathcal{R}^2$ and $N_X(x) = \{(x, 0) | x \in (-\infty, +\infty)\}$. To see that $N_X(x) = \{(x, 0) | x \in (-\infty, +\infty)\}$ note that the condition $z_k \in T_X(x_k)^*$ of the definition (21.1) requires that $\{x_k\}$ approach $x = (0,0)$ along the boundary of X. For any other choice of $\{x_k\}$ we have $T_X(x) = \mathcal{R}^2$ and $T_X(x)^* = \{(0,0)\}$.

Prop. 21.2 *(Tangent cone 2). Let X be a subset of $\mathcal{R}^n$ and $x \in X$. Then*

$$T_X(x) = \left\{ y | \exists \{x_k\} \subset X, \exists \{\alpha_k\} \subset (0, +\infty) \;\; s.t. \; \alpha_k \to 0, \; \frac{x_k - x}{\alpha_k} \to y \right\}.$$

Proof. Let $y \in T_X(x)$ then according to the definition (21.1) there is a sequence $\{x_k\} \subset X$ s.t. $x_k \to x$ and $\frac{x_k - x}{\|x_k - x\|} \to \frac{y}{\|y\|}$. We set $\alpha_k = \frac{\|x_k - x\|}{\|y\|}$.

Conversely, let $\{\alpha_k\}$ be the sequence as stated in the proposition then

$$\frac{x_k - x}{\alpha_k} \to y \Rightarrow x_k \to x$$

and

$$\frac{x_k - x}{\|x_k - x\|} = \frac{(x_k - x)/\alpha_k}{\|x_k - x\|/\alpha_k} \to \frac{y}{\|y\|}.$$

■

Prop. 21.3 *(Tangent cone 3). Let X be a subset of $\mathcal{R}^n$ and $x \in X$.*

1. $T_X(x)$ is a closed cone.

2. $cl(F_X(x)) \subseteq T_X(x)$.

3. If X is convex then $F_X(x)$ and $T_X(x)$ are convex and $cl(F_X(x)) = T_X(x)$.

Proof. (1). Consider $\{y_k\}, y_k \in T_X(x)$ such that $y_k \to y$. We aim to show that $y \in T_X(x)$. We exclude non essential case $y = 0$.

By definition of $T_X(x)$ there are sequences $\{x_{pk}\}$, $x_{pk} \to x$ and $\frac{x_{pk} - x}{\|x_{pk} - x\|} \to \frac{y_k}{\|y_k\|}$ as $p \to \infty$.

There exists an increasing function $m(\cdot)$ s.t. $\|y_{m(k)} - y\| \le \frac{1}{k}$. We can also find a function $p(\cdot)$ such that $p(k) > \max(p(k-1), m(k))$, $\|x_{p(k)m(k)} - x\| \le \frac{1}{k}$ and $\left\| \frac{x_{p(k)m(k)} - x}{\|x_{p(k)m(k)} - x\|} - \frac{y_{m(k)}}{\|y_{m(k)}\|} \right\| \le \frac{1}{k}$. The sequence $\tilde{x}_k \equiv x_{p(k)m(k)}$ is the sequence that we need to show that $y \in T_X(x)$ in context of the definition (21.1). ■

Proof. (2). $F_X(x) \subseteq T_X(x)$ by definitions and by (1) the $T_X(x)$ is closed. ■

Proof. (3). Since X is convex all the feasible directions $\bar{x} \in F_X(x)$ are of the form $\alpha(\bar{x} - x)$, $\alpha > 0$. Hence, $F_X(x)$ is convex. By the proposition (21.2) the $T_X(x)$ consists of y that are limit points sequences of such feasible directions $\frac{x_k - x}{\alpha_k}$. Hence, $T_X(x) \subseteq cl(F_X(x))$. Therefore, in combination with (2), the $T_X(x) = cl(F_X(x))$ follows and $T_X(x)$ is convex. ■

Prop. 21.4 *(Tangent cone 4). Let X be a nonempty convex subset of $\mathcal{R}^n$ and $x \in X$.*

1. $z \in T_X(x)^* \Leftrightarrow \forall \bar{x} \in X: \quad \langle z, \bar{x} - x \rangle \le 0$.

2. X is regular for all $x \in X$: $T_X(x)^* = N_X(x)$.

3. $T_X(x) = N_X(x)^*$.

Proof. Since X is convex, any feasible direction $y \in F_X(x)$ is of the form $\alpha(\bar{x} - x)$, $\alpha > 0$. Hence, (1) follows from the proposition (21.3)-3 and the definition (17.1).

The (2) follows from (1) and the definition (21.1).

The (3) is a consequence of the proposition (17.3), (2) and the proposition (21.3)-1,3. ■

22 Optimality conditions.

Prop. 22.1 *(Minimum of a smooth function). Let $f : \mathcal{R}^n \rightarrow \mathcal{R}$ be a smooth function and let x^* be a minimum of f over the subset X of $\mathcal{R}^n$. Then*

$$\forall y \in T_X(x^*) \quad : \quad \langle \nabla f(x^*), y \rangle \geq 0.$$

Equivalently,

$$-\nabla f(x^*) \in T_X(x^*)^*.$$

If X is convex then

$$\forall x \in X \quad : \quad \langle \nabla f(x^*), x - x^* \rangle \geq 0.$$

If $X = \mathcal{R}^n$ then

$$\nabla f(x^*) = 0.$$

Proof. Let $y \in T_X(x^*)$, $y \neq 0$ then there exists $\{x_k\}, x_k \in X$ such that

$$x_k \rightarrow x,$$
$$\frac{x_k - x^*}{\|x_k - x^*\|} \rightarrow \frac{y}{\|y\|}.$$

By smoothness of f we have

$$f(x_k) = f(x^*) + \langle \nabla f(x^*), x_k - x^* \rangle + o(\|x_k - x^*\|).$$

Hence,

$$0 \leq \frac{f(x_k) - f(x^*)}{\|x_k - x^*\|} = \left\langle \nabla f(x^*), \frac{x_k - x^*}{\|x_k - x^*\|} \right\rangle + \frac{o(\|x_k - x^*\|)}{\|x_k - x^*\|}.$$

We pass the above to the limit and obtain

$$0 \leq \langle \nabla f(x^*), y \rangle.$$

The rest of the proposition follows from the proposition (21.4)-1. ■

Prop. 22.2 *(Minimum of a convex function). Let $f:\mathcal{R}^n\to\mathcal{R}$ be a convex function and let X be a convex subset of $\mathcal{R}^n$. Then*

$$x^*\in\arg\min_X f(x)\ \Leftrightarrow\ \exists d\in\partial f(x^*)\ \ s.t.\ \forall x\in X:\ \ \langle d,x-x^*\rangle\geq 0.$$

Equivalently,

$$x^*\in\arg\min_X f(x)\ \Leftrightarrow\ 0\in\partial f(x^*)+T_X(x^*)^*.$$

Proof. Assume $d\in\partial f(x^*)$ and $\langle d,x-x^*\rangle\geq 0$ for any $x\in X$. Then by the definition (20.5) $f(x)-f(x^*)\geq\langle d,x-x^*\rangle\geq 0$ and thus $x^*\in\arg\min_X f(x)$.

Conversely, let $x^*\in\arg\min_X f(x)$. Then $f'(x^*;x-x^*)\geq 0$ for any $x\in X$. According to the proposition (20.7)-1

$$f'(x^*;x-x^*)=\sup_{d\in\partial f(x^*)}\langle d,x-x^*\rangle.$$

According to the proposition (20.6) the $\sup_{d\in\partial f(x^*)}$ is taken over a compact set. Also, the $\langle d,x-x^*\rangle$ is a continuous function of d. Hence, the $\sup_{d\in\partial f(x^*)}\langle d,x-x^*\rangle$ is achieved at some d^*. Such d^* has the property

$$\forall x\in X:\ \ \langle d^*,x-x^*\rangle\geq 0.$$

The second part of the proposition $0\in\partial f(x^*)+T_X(x^*)^*$ is evident because the statement $\langle d,x-x^*\rangle\geq 0$ may be rewritten as $-d\in T_X(x^*)^*$ according to the definition (17.1). ■

Prop. 22.3 *(Local minimum of a sum). Let $f_1:\mathcal{R}^n\to\mathcal{R}$ be a convex function, $f_2:\mathcal{R}^n\to\mathcal{R}$ be a smooth function, X be a subset of $\mathcal{R}^n$, x^* be a local minimum of $f=f_1+f_2$ and let $T_X(x^*)$ be convex. Then*

$$-\nabla f_2(x^*)\in\partial f_1(x^*)+T_X(x^*)^*.$$

Proof. The proof is a repetition of the proofs for the propositions (22.1) and (22.2). ■

The figure (1.14) illustrates the condition $-\nabla f(x^*)\in T_X(x^*)^*$. The painted triangle is the constraint set X. The elipses are the level curves of a function $f(x)$ with the internal elipse is the level curve with the smallest value. The slightly transparent triangle is the set $x^*+T_X(x^*)^*$. The arrow is

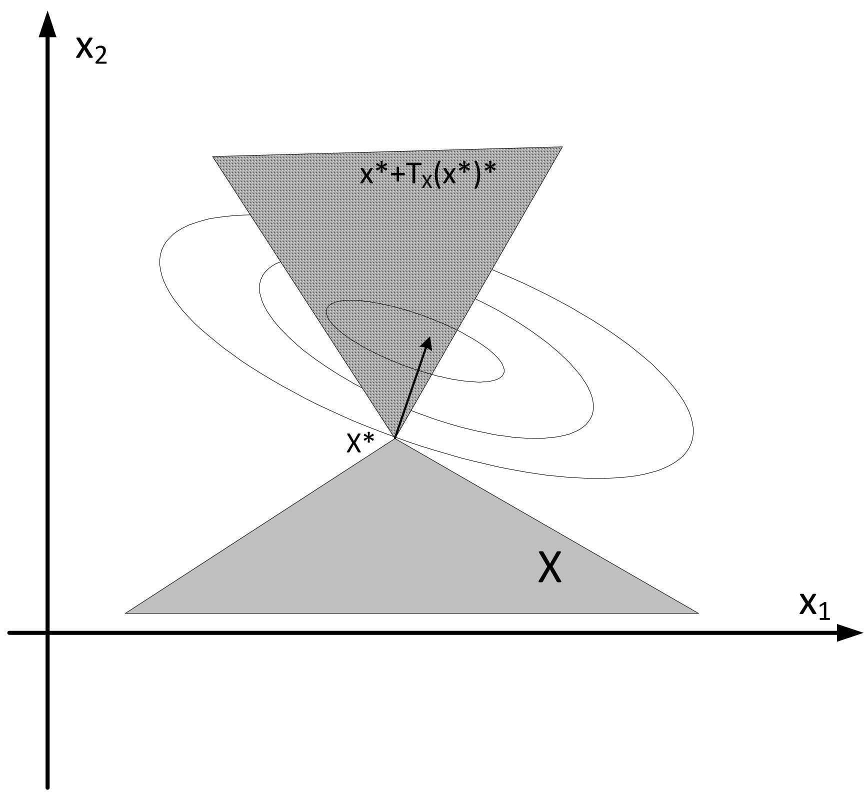

Figure 1.14: Optimality for smooth function figure 1

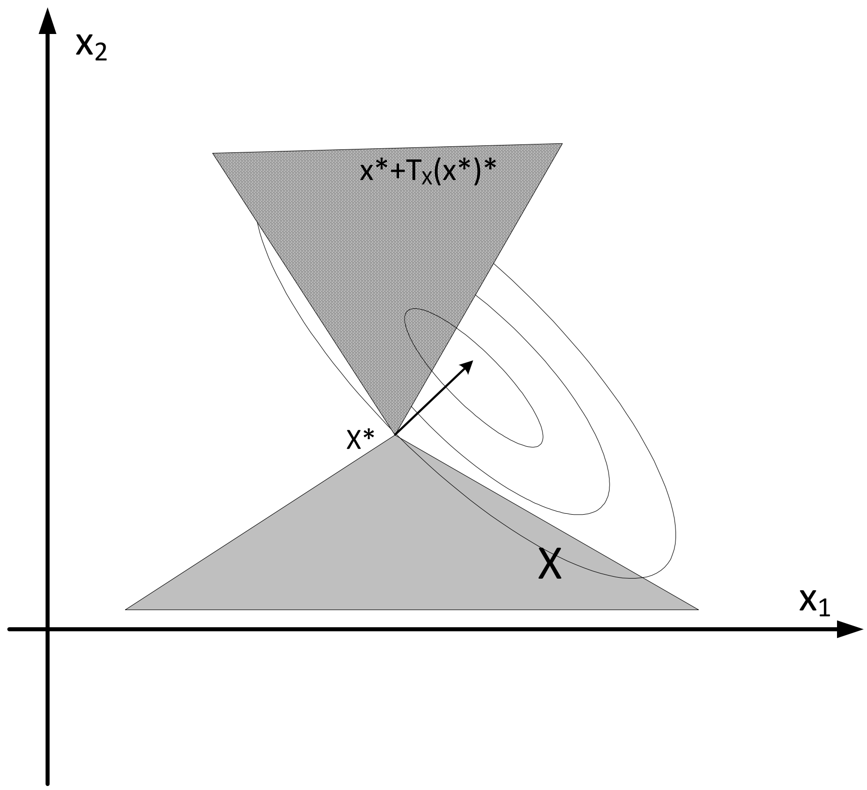

Figure 1.15: Optimality for smooth function figure 2

the vector $-\nabla f(x^*)$. The $\nabla f(x^*)$ is orthogonal to the level curve that passes through x^* and points to the direction of increase of f. The $-\nabla f(x^*)$ points in direction of decrease. Because the $-\nabla f(x^*)$ lies within the $T_X(x^*)^*$ the point x^* minimises f over X. The alternative situation is presented on the picture (1.15). Here, $-\nabla f(x^*)$ lies outside of the $T_X(x^*)^*$. In addition the $-\nabla f(x^*)$ must be orthogonal to the level curve. Therefore, the level curve must cross into X thus preventing x^* from being the minimum.

23 Lagrange multipliers for equality constraints.

We are considering the following problem.

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } h_i(x) = 0,\ i = 1, ..., m, \end{aligned} \tag{1.5}$$

where the f and h_i are smooth functions $\mathcal{R}^n \to \mathcal{R}$.

Prop. 23.1 *(Existence of Lagrange multipliers for equality constraints). Let x^* be a local minimum of the problem (1.5) and*

$$X = \{x|\ h_i(x) = 0,\ i = 1, ..., m\},$$
$$T_X(x^*) = \{y|\ \langle \nabla h_i(x^*), y\rangle = 0,\ i = 1, ..., m\}$$

then there are scalars $\{\lambda_i^*\}_{i=1,...,m}$ such that

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) = 0.$$

Proof. The condition

$$T_X(x^*) = \{y|\ \langle \nabla h_i(x^*), y\rangle = 0,\ i = 1, ..., m\}$$

implies

$$T_X(x^*) = N\left(A^T\right),$$

where the $n \times m$ matrix A consists of the columns $\{h_i(x^*)\}$:

$$A = \|h_i(x^*)\|_{i=1,...,m}.$$

Hence,

$$\left\langle x, A^T y\right\rangle = 0,\ \forall x \in \mathcal{R}^m,\ \forall y \in N\left(A^T\right) = T_X(x^*).$$

Equivalently,

$$\langle z, y\rangle = 0,\ \forall z \in R(A),\ \forall y \in N\left(A^T\right) = T_X(x^*).$$

Therefore,

$$R(A) \subseteq T_X(x^*)^*.$$

We next show that

$$T_X(x^*)^* \subseteq R(A).$$

Indeed, we already established that $T_X(x^*)$ is a subspace, hence, $y \in T_X(x^*) \Rightarrow -y \in T_X(x^*)$. Therefore, if $\langle z, y\rangle \leq 0, \forall y \in T_X(x^*)$ then $\langle z, y\rangle = 0, \forall y \in T_X(x^*)$. But

$$T_X(x^*)^* = \{z | \langle z, y\rangle \leq 0, \forall y \in T_X(x^*)\}$$

and we proved already that

$$R(A) = \{z | \langle z, y\rangle = 0, \forall y \in T_X(x^*)\}.$$

Therefore,

$$R(A) = T_X(x^*)^*.$$

According to the proposition (22.1),

$$-\nabla f(x^*) \in T_X(x^*)^*$$

hence,

$$-\nabla f(x^*) \in \text{Linear hull of } \{\nabla h_i(x^*)\}_{i=1,\dots,m}$$

and the conclusion of the proposition follows. ■

The condition

$$T_X(x^*) = \{y | \ \langle \nabla h_i(x^*), y\rangle = 0, \ i = 1, ..., m\}$$

states that the $T_X(x^*)$ consists of directions tangent to the level surfaces of h_i crossing the x^*. For example,

$$h_i(x) = \langle a_i, x\rangle + b_i, \ i = 1, ..., m \Rightarrow$$
$$T_X(x) = \{y | \ \langle a_i, y\rangle = 0, \ i = 1, ..., m\}.$$

24 Fritz John optimality conditions.

Prob. 24.1 *(Smooth optimization problem). We consider the following problem*

$$\text{minimize } f(x)$$
$$\text{subject to } x \in C,$$
$$C = X \cap \{x | \ h_i(x) = 0, \ i = 1, ..., m\}$$
$$\cap \{x | \ g_j(x) \leq 0, \ j = 1, ..., r\}$$

where the f, h_i, g_j *are smooth functions* $\mathcal{R}^n \to \mathcal{R}$ *and* X *is a nonempty closed set.*

Prop. 24.2 *(Fritz John conditions). Let* x^* *be a local minimum of the problem (24.1). Then there exist quantities* $\mu_0^*, \lambda^* \equiv \{\lambda_i^*\}_{i=1,...,m}, \mu^* \equiv \{\mu_j^*\}_{j=1,...,r}$ *such that*

1.

$$
\begin{aligned}
&-\nabla_x L(x^*, \mu_0^*, \lambda^*, \mu^*) \in N_X(x^*),\\
&\qquad L(x, \mu_0, \lambda, \mu) \equiv \mu_0^* f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle,\\
&\qquad\qquad h(x) \equiv \{h_i(x)\}_{i=1,...,m},\ g(x) \equiv \{g_j(x)\}_{j=1,...,r}.
\end{aligned}
$$

2.

$$
\begin{aligned}
&\mu_j^* \geq 0,\ j = 1, ..., r,\\
&\mu_0^* \geq 0.
\end{aligned}
$$

3.

$$
(\mu_0^*)^2 + \|\lambda^*\|^2 + \|\mu^*\|^2 \neq 0.
$$

4. Let

$$
I \equiv \{i|\ \lambda_i^* \neq 0\},\ J = \{j|\ \mu_j^* \neq 0\}.
$$

If

$$
I \cup J \neq \varnothing
$$

then there exists $\{x_k\} \subset X$ *such that*

$$
\begin{aligned}
&x_k \to x^*,\\
&\forall k : f(x_k) < f(x^*),\\
&\forall i \in I : \lambda_i^* h_i(x_k) > 0,\\
&\forall j \in J : \mu_j^* g_j(x_k) > 0,\\
&\forall i \notin I : |h_i(x_k)| = o(w(x_k)),\\
&\forall j \notin J : g_j^+(x_k) = o(w(x_k))
\end{aligned}
$$

where $g_j^+(x) = \max(0, g_j(x))$ *and*

$$
w(x) = \min\left\{\min_{i \in I} |h_i(x)|, \min_{j \in J} g_j^+(x)\right\}.
$$

Proof. Let

$$F_k(x) \equiv f(x) + \frac{k}{2}\|h(x)\|^2 + \frac{k}{2}\left\|g^+(x)\right\|^2 + \frac{1}{2}\|x - x^*\|^2$$

for $k = 1, 2, ...$

Consider the problems

$$\begin{gathered}\text{minimize } F_k(x) \\ \text{subject to } x \in X \cap S_\varepsilon\end{gathered} \tag{1.6}$$

where

$$S_\varepsilon = \{x \mid \|x - x^*\| \le \varepsilon\}$$

and $\varepsilon > 0$ is such that

$$\forall x \in S_\varepsilon \cap C: \; f(x^*) \le f(x).$$

By the classic version of the Weierstrass theorem there exists a solution x_k of the problem (1.6) for every k . In particular,

$$F_k(x_k) \le F_k(x^*).$$

Note that $x^* \in C \Rightarrow h(x^*) = 0$ and $g^+(x^*) = 0$. Hence, we rewrite the last inequality as

$$f(x_k) + \frac{k}{2}\|h(x_k)\|^2 + \frac{k}{2}\left\|g^+(x_k)\right\|^2 + \frac{1}{2}\|x_k - x^*\|^2 \le f(x^*). \tag{1.7}$$

By construction, $\{x_k\}$ is a bounded sequence. Therefore, is has one or more limit points $\bar{x}$.

The f is smooth, hence, $f(x_k)$ is bounded. Therefore,

$$\|h(x_k)\| \to 0, \;\; \left\|g^+(x_k)\right\| \to 0$$

because otherwise $\frac{k}{2}\|h(x_k)\|^2$ and $\frac{k}{2}\|g^+(x_k)\|^2$ cannot be bounded by the $f(x^*)$.

Thus, all the limit points $\bar{x}$ are feasible:

$$\bar{x} \in C.$$

Therefore, by the construction of S_ε and ε,

$$f(x^*) \le f(\bar{x}). \tag{1.8}$$

By passing to the limit the inequality (1.7) and combining with (1.8) we conclude

$$\bar{x} = x^*$$

for every limit point $\bar{x}$. Thus x_k is convergent and

$$x_k \to x^*.$$

According to the proposition (22.1)

$$-\nabla_x F(x_k) \in T_{X\cap S_\varepsilon}(x_k)^*.$$

By convergence $x_k \to x^*$, for large enough k the x_k is inside S_ε, hence

$$T_{X\cap S_\varepsilon}(x_k)^* = T_X(x_k)^*.$$

We restrict our attention to such k.

We calculate

$$\nabla_x F(x_k) = \nabla_x f(x_k) + k\langle h(x_k), \nabla_x h(x_k)\rangle + k\left\langle g^+(x_k), \nabla_x g^+(x_k)\right\rangle + x_k - x^* \tag{1.9}$$

and introduce the notation

$$\begin{aligned}
\delta_k &= \sqrt{1 + k^2\|h(x_k)\|^2 + k^2\|g^+(x_k)\|^2},\\
\mu_{0,k} &= \frac{1}{\delta_k},\\
\lambda_k &= \frac{k}{\delta_k}h(x_k),\\
\mu_k &= \frac{k}{\delta_k}g^+(x_k).
\end{aligned}$$

Note that the sequence of k

$$\{\mu_{0,k}, \lambda_k, \mu_k\}$$

is bounded:

$$\mu_{0,k}^2 + \|\lambda_k\|^2 + \|\mu_k\|^2 = 1.$$

Hence, it has a limit point

$$\left\{\mu_{0,k}^*, \lambda_k^*, \mu_k^*\right\}.$$

By dividing (1.9) with δ_k we obtain

$$\mu_{0,k} f(x_k) + \langle \lambda_k, \nabla_x h(x_k) \rangle + \left\langle \mu_k, \nabla_x g^+(x_k) \right\rangle + \frac{1}{\delta_k}(x_k - x^*) \in T_X(x_k)^* .$$

We pass the last relationship to the limit $k \to \infty$ and arrive to

$$\mu_0^* f(x^*) + \langle \lambda^*, \nabla_x h(x^*) \rangle + \left\langle \mu^*, \nabla_x g^+(x^*) \right\rangle \in N_X(x^*),$$

(compare with the definition (21.1)).

To see that the statement (4) holds, note that by construction of λ_k^*, μ_k^*, $\lambda_k = \frac{k}{\delta_k} h(x_k)$, if $i \in I$ then $\lambda_i^* h_i(x_k) > 0$ for large enough k. If $i \notin I$ then the i-th component of λ_k: $(\lambda_k)_i = \frac{k}{\delta_k} h_i(x_k)$ has to vanish as $k \to \infty$. Hence, if $i \notin I$ then $h_i(x_k)$ vanishes quicker than any of the $h_i(x_k)$ for $i \notin I$. The consideration for g_j is identical. ■

25 Pseudonormality.

We use the notation of the problem (24.1).

Def. 25.1 *(Pseudonormality). The vector $x^* \in C$ is called "pseudo-normal" if one cannot find the vectors λ,μ and a sequence $\{x_k\} \subset X$ such that*

1. $-\langle \lambda, \nabla_x h(x^*) \rangle - \langle \mu, \nabla_x g(x^*) \rangle \in N_X(x^*)$,
2. $\mu \geq 0$, $g_j(x^*) \neq 0 \Rightarrow \mu_j = 0$ *and* $g_j(x^*) = 0 \Rightarrow \mu_j \neq 0$.
3. $x_k \to x^*$ *and* $\langle \lambda, h(x_k) \rangle + \langle \mu, g(x_k) \rangle > 0,\ \forall k$.

Note that (1) implies that the proposition (24.2) cannot take place with $\mu_0^* = 0$. The conditions (2),(3) imply that the components of λ, μ are "informative" in the sense that the set $I \cup J$ of the proposition (24.2) is nonempty and the non-zero components of λ, μ mark those conditions $h_i(x) = 0$ and $g_j(x) \leq 0$ that are "active" (x_k of the proposition (24.2)'s proof violates these conditions and the x^* lies on the boundary set by such conditions).

We introduce the notation

$$A(x^*) = \{j |\ g_j(x^*) = 0\} .$$

The condition 2 of the above definition may be equivalently written as

$$\mu \geq 0 \text{ and } A(x^*) = \{j |\ \mu_j \neq 0\} .$$

Prop. 25.2 *(Constraint qualification 1). If $X = \mathcal{R}^n$, $x^* \in C$ and the vectors $\{\nabla h_i(x^*)\}_{i=1,...,m} \cup \{\nabla g_j(x^*)\}_{j \in A(x^*)}$ are linearly independent then the vector x^* is pseudonormal.*

Proof. Since $X = \mathcal{R}^n$ we have $N_X(x^*) = \{0\}$. Hence, the conditions 1 and 2 of the definition (25.1), if true, would imply the linear dependence $\{\nabla h_i(x^*)\}_{i=1,...,m} \cup \{\nabla g_j(x^*)\}_{j \in A(x^*)}$. Therefore, such λ and μ, as in the definition (25.1), cannot exist. ∎

Prop. 25.3 *(Constraint qualification 2). If $X = \mathcal{R}^n$, $x^* \in C$, $A(x^*) \neq \varnothing$ and there exists a $y \in \mathcal{R}^n$ such that*

$$\begin{aligned}
&\langle y, \nabla_x h_i(x^*)\rangle_x = 0,\ i = 1, ..., m;\\
&\langle y, \nabla_x g_j(x^*)\rangle_x < 0,\ j \in A(x^*)
\end{aligned}$$

then the vector x^* is pseudonormal.

Here the x-sign after the brackets $\langle\ \cdot\ \rangle_x$ indicates that the summation of the scalar product is applied to the components of the gradient ∇_x.

Proof. In the condition 1 of the definition (25.1) the LHS

$$-\langle \lambda, \nabla_x h(x^*)\rangle - \langle \mu, \nabla_x g(x^*)\rangle$$

is a vector of components of the gradient ∇_x. The scalar product applies to i and j indexes of $\{h_i(x)\}_{i=1,...,m}$ and $\{g_j(x)\}_{j=1,...,r}$. We appy the scalar product with respect to components of the gradient ∇_x and write the following consequence of the condition 1:

$$\left\langle y, \sum_{i=1}^{m} \lambda_i \nabla_x h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla_x g_j(x^*) \right\rangle_x = 0.$$

Here we used that $N_X(x^*) = \{0\}$ for $X = \mathcal{R}^n$. We rearrange the terms as follows

$$\sum_{i=1}^{m} \lambda_i \langle y, \nabla_x h_i(x^*)\rangle_x + \sum_{j=1}^{r} \mu_j \langle y, \nabla_x g_j(x^*)\rangle_x = 0.$$

Therefore, the λ and μ as in the definition (25.1) cannot exist because the first sum $\sum_{i=1}^{m} \lambda_i \langle y, \nabla_x h_i(x^*)\rangle_x$ is zero by the condition $\langle y, \nabla_x h_i(x^*)\rangle_x = 0$ of the proposition and the second sum is negative by the condition $\langle y, \nabla_x g_j(x^*)\rangle_x < 0,\ j \in A(x^*)$ of the proposition and the condition 2 of the definition (25.1). ∎

Prop. 25.4 *(Constraint qualification 3). If $X = \mathcal{R}^n$, $x^* \in C$, the functions $h_i(x)$ are affine and the functions $g_j(x)$ are concave then the vector x^* is pseudonormal.*

Proof. By the conditions on h_i and g_i we have

$$\begin{aligned} h_i(x) &= h_i(x^*) + \langle \nabla_x h_i(x^*), x - x^* \rangle_x, \\ g_i(x) &\leq g_i(x^*) + \langle \nabla_x g_i(x^*), x - x^* \rangle_x \end{aligned}$$

for any $x \in \mathcal{R}^n$. Therefore, for any λ and μ

$$\begin{aligned} \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_i g_i(x) \leq \sum_{i=1}^{m} \lambda_i h_i(x^*) + \sum_{j=1}^{r} \mu_j g_j(x^*) \\ + \left\langle \sum_{i=1}^{m} \lambda_i \nabla_x h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla_x g_j(x^*), x - x^* \right\rangle. \end{aligned}$$

By the inclusion $x \in C$, the first sum is zero and the second sum is non-positive. Hence, if λ and μ satisfy the condition 1 of the definition (25.1):

$$\sum_{i=1}^{m} \lambda_i \nabla_x h_i(x^*) + \sum_{j=1}^{r} \mu_j \nabla_x g_j(x^*) \in N_{\mathcal{R}^n}(x^*) = \{0\}$$

then the condition 3 of the definition (25.1) must fail. ■

Prop. 25.5 *(Constraint qualification 4). Let $X = \mathcal{R}^n$, $x^* \in C$, the x^* is pseudonormal for the set*

$$\bar{C} = \{x \mid h_i(x) = 0, i = 1, ..., m; \ g_j(x) < 0, j = \bar{r} + 1, ..., r\},$$

and for some $\bar{r} < r$. Furthemore, there exists a $y \in \mathcal{R}^n$ such that

$$\begin{aligned} &\langle y, \nabla_x h_i(x^*) \rangle_x = 0, \ i = 1, ..., m; \\ &\langle y, \nabla_x g_j(x^*) \rangle_x \leq 0, \ j \in A(x^*), \\ &\langle y, \nabla_x g_j(x^*) \rangle_x < 0, \ j \in \{1, ..., \bar{r}\} \cap A(x^*). \end{aligned}$$

Proof. Note that $A(x^*) \neq \varnothing$ because if $A(x^*) = \varnothing$ then $\mu = 0$ and the conditions 1,2,3 of the definition (25.1) are satisfied for the set $\bar{C}$. The rest of the proof is a repetition of the proof of the proposition (25.3). ■

Prop. 25.6 *(Constraint qualification 5). Assume that the following conditions are satisfied.*

1. The functions $h_i(x)$, $i = \bar{m}+1, ..., m$ are linear for some $\bar{m} \le m$.

2. The does not exists a $\lambda = \{\lambda_i\}_{i=1,\dots,m}$ such that

$$-\langle \lambda, \nabla_x h(x^*) \rangle \in N_X(x^*)$$

and not all $\{\lambda_i\}_{i=1,\dots,\bar{m}}$ are zero.

3. Let

$$V_L(x^*) = \{y | \ \langle y, \nabla_x h_i(x^*) \rangle_x = 0, \ i = \bar{m}+1, ..., m\}.$$

Either $V_L(x^*) \cap Interior(N_X(x^*)^*) \neq \varnothing$ or X is convex and $V_L(x^*) \cap ri(N_X(x^*)^*) \neq \varnothing$.

4. There exists a $y \in N_X(x^*)^*$ such that

$$\langle y, \nabla_x h_i(x^*) \rangle_x = 0, \ i = 1, ..., m;$$
$$\langle y, \nabla_x g_j(x^*) \rangle_x < 0, \ j \in A(x^*).$$

Then the vector x^* is pseudonormal.

Proof. We assume that all the conditions of the definition (25.1) hold and reach a contradition.

We introduce the notation

$$\gamma \equiv \langle \lambda, \nabla_x h(x^*) \rangle + \langle \mu, \nabla_x g(x^*) \rangle.$$

According to the condition 4 of this proposition and condition 2 of the definition (25.1), there exists a $y \in N_X(x^*)^*$ such that

$$\langle y, \gamma \rangle_x < 0, \text{ when } A(x^*) \neq \varnothing,$$
$$\langle y, \gamma \rangle = 0, \text{ when } A(x^*) = \varnothing.$$

The condition 1 of the definition (25.1) requires that

$$-\gamma \in N(x^*)$$

thus

$$\langle z, -\gamma \rangle \le 0 \text{ for any } z \in N(x^*)^*.$$

Hence, we already proven the statement for the case $A(x^*) \neq \varnothing$.

It remains to consider the case $A(x^*) = \varnothing$ under the assumption that the conditions 1,2,3 of of the definition (25.1) and the conditions 1,2,3,4 of this proposition are true and arrive to contradiction. By the assumpion $A(x^*) = \varnothing$, we have

$$g_j(x^*) < 0,\ j = 1, ..., r,$$

and by condition 2 of the definition (25.1) we have

$$\mu = 0.$$

The condition 1 of the definition (25.1) implies

$$-\langle \lambda, \nabla_x h(x^*) \rangle = -\sum_{i=1}^{\bar{m}} \lambda_i \nabla_x h_i(x^*) - \sum_{i=\bar{m}+1}^{m} \lambda_i \nabla_x h_i(x^*) \in N_X(x^*).$$

Hence, by the condition 2 of the propositon, all $\{\lambda_i\}_{i=1,\ldots,\bar{m}}$ are zero:

$$\zeta \equiv -\sum_{i=\bar{m}+1}^{m} \lambda_i \nabla_x h_i(x^*) \in N_X(x^*).$$

By the condition 3 there is a y from the interior of $N_X(x^*)^*$ such that

$$\langle y, \nabla_x h_i(x^*) \rangle_x = 0,\ i = \bar{m}+1, ..., m.$$

Hence,

$$\begin{aligned} \langle y, \zeta \rangle_x &= \left\langle y, -\sum_{i=\bar{m}+1}^{m} \lambda_i \nabla_x h_i(x^*) \right\rangle_x \\ &= -\sum_{i=1}^{\bar{m}} \lambda_i \langle y, \nabla_x h_i(x^*) \rangle_x = 0. \end{aligned}$$

Hence, we have found a point $\zeta \in N_X(x^*)$ and an interior point y of $N_X(x^*)^*$ such that

$$\langle y, \zeta \rangle = 0.$$

This is a contradiction. For an interior point of a cone $N_X(x^*)^*$ we must have

$$\langle y, \zeta \rangle < 0.$$

■

26 Lagrangian duality.

We consider the following problem.

Prob. 26.1 *(Primal problem). Find*

$$f^* = \inf_{x \in C} f(x)$$

where

$$C = \{x|\ x \in X,\ h(x) = 0,\ g(x) \leq 0\},$$

$X \subseteq \mathcal{R}^n$, $x \in \mathcal{R}^n$,

$$h(x) = \{h_i(x)\}_{i=1,\dots,m},$$
$$g(x) = \{g_j(x)\}_{j=1,\dots,r},$$

and $f, h_i, g_j : \mathcal{R}^n \to \mathcal{R}$.

We introduce the notation

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle,$$
$$S = \{(h(x), g(x), f(x)) | x \in X\}.$$

26.1 Geometric multipliers.

Def. 26.1.1 *(Geometric multiplier). The pair* (λ^*, μ^*) *is a called a "geometric multiplier" for the problem (26.1) if* $\mu^* \geq 0$ *and*

$$f^* = \inf_{x \in X} L(x, \lambda^*, \mu^*).$$

The following statement directly follows from the definitions (26.1),(26.1.1).

Prop. 26.1.2 *(Visualization lemma). Assume* $-\infty < f^* < +\infty$.

1. The hyperplane in $\mathcal{R}^{m+r+1}$ *with normal* $(\lambda, \mu, 1)$ *that passes through* $(h(x), g(x), f(x))$ *also passes through* $(0, 0, L(x, \lambda, \mu))$.

2. Among all hyperplanes with normal $(\lambda, \mu, 1)$ *that contain the set* S *in the upper halfspace, the highest level of interseption with the axis* $\{(0, 0, w) | w \in \mathcal{R}\}$ *is given by* $\inf_{x \in X} L(x, \lambda, \mu)$.

Prop. 26.1.3 *(Geometric multiplier property). Let (λ^*, μ^*) be a geometric multiplier then x^* is a global minimum of the problem (26.1) if and only if $x^* \in C$ and*

$$L(x^*, \lambda^*, \mu^*) = \min_{x \in X} L(x, \lambda^*, \mu^*),$$
$$\langle \mu^*, g(x^*) \rangle = 0.$$

Proof. Note that $x^* \in C$ implies $h(x^*) = 0$ and $g(x^*) \leq 0$ and the definition (26.1.1) requires $\mu \leq 0$. Hence, Hence, $\langle \lambda^*, h(x) \rangle = 0$ and $\langle \mu^*, g(x^*) \rangle \leq 0$.

Let x^* be a global minimum of the problem (26.1) then

$$f^* = \min_{x \in C} f(x) = f(x^*).$$

By the definition (26.1.1),

$$f^* = \inf_{x \in X} L(x, \lambda^*, \mu^*) = \inf_{x \in X} \{f(x) + \langle \lambda^*, h(x) \rangle + \langle \mu^*, g(x) \rangle\}.$$

Therefore, $\langle \mu^*, g(x^*) \rangle = 0$ and $L(x^*, \lambda^*, \mu^*) = \min_{x \in X} L(x, \lambda^*, \mu^*)$.

The statement is proven similarly in the other direction. ■

Def. 26.1.4 *(Lagrange multiplier). The pair (λ^*, μ^*) is called "Lagrange multiplier of the problem (26.1) associated with the solution x^*" if*

$$0 \in \partial f(x^*) + \langle \lambda^*, \partial h(x^*) \rangle + \langle \mu^*, \partial g \rangle + T_X(x^*)^*$$

and

$$\mu^* \geq 0, \quad \langle \mu^*, g(x^*) \rangle = 0.$$

The following statement is a consequence of the proposition (22.3) and definitions.

Prop. 26.1.5 *Assume that the problem (26.1) has at least one solution x^*.*

1. Let f and $\{g_i\}$ are either convex or smooth , $\{h_i\}$ are smooth, X is closed and $T_X(x^*)$ is convex then every geometric multiplier is a Lagrange multiplier.

2. Let f and $\{g_i\}$ are convex, $\{h_i\}$ are affine and X is closed and convex then the sets of Lagrange and geometric multiplier coincide.

26.2 Dual problem.

Prob. 26.2.1 *(Dual problem). Find*

$$\sup_{\lambda\in\mathcal{R}^m,\mu\in\mathcal{R}^r,\mu\geq 0} q(\lambda,\mu)$$

where

$$q(\lambda,\mu)=\inf_{x\in X} L(x,\lambda,\mu).$$

The dual problem delivers the highest crossing point for the set

$$S=\{(h(x),g(x),f(x))\mid x\in X\}.$$

Note that q is an inf of a collection of affine functions. Hence, it is concave, upper semi-continuous and may be studied with the means of the propositions (14.2),(14.3). In particular, the following statement directly follows from the proposition (14.2), the geometrical interpretation of the (26.1.2) and the definition (26.1.1).

Prop. 26.2.2 *(Duality gap and geometric multipliers). The following alternative takes place.*

1. If $q^*=f^*$ (="there is no duality gap") then the set of geometric multipliers is equal to the set of solutions of the problem (26.2.1).

2. If $q^*<f^*$ (="there is a duality gap") then the set of geometric multipliers is empty.

26.3 Connection of dual problem with minimax theory.

Prop. 26.3.1 *1. The problem (26.1) is equivalent to*

$$\inf_{x\in X}\sup_{\lambda\in\mathcal{R}^m,\mu\in\mathcal{R}^r,\mu\geq 0} L(x,\lambda,\mu).$$

2. The problem (26.2.1) is equivalent to

$$\sup_{\lambda\in\mathcal{R}^m,\mu\in\mathcal{R}^r,\mu\geq 0}\inf_{x\in X} L(x,\lambda,\mu).$$

Proof. Note that

$$\sup_{\lambda\in\mathcal{R}^m,\mu\in\mathcal{R}^r,\mu\geq 0}\left\{f(x)+\langle\lambda,h(x)\rangle+\langle\mu,g(x)\rangle\right\}=\left\{\begin{array}{c} f(x)\,,\text{ if } h(x)=0,\ g(x)\leq 0 \\ \infty\,,\text{ otherwise}\end{array}\right\}.$$

The rest follows from the definitions of the problems (26.1) and (26.2.1).■

Prop. 26.3.2 *(Necessary and sufficient optimality conditions).The vectors (x^*,λ^*,μ^*) form a solution of the problem (26.1) and a geometric multiplier pair if and only if the following four conditions hold.*

$$x^*\in C \tag{1.10}$$

$$\mu^*\geq 0 \tag{1.11}$$

$$L(x^*,\lambda^*,\mu^*)=\min_{x\in X}L(x,\lambda^*,\mu^*) \tag{1.12}$$

$$\langle\mu^*,g(x^*)\rangle=0 \tag{1.13}$$

Proof. If the (x^*,λ^*,μ^*) form a solution of the problem (26.1) and a geometric multiplier pair then the statements (1.10) and (1.11) follow from the definitions and (1.12),(1.13) follow from the proposition (26.1.3).

Conversely, using the conditions of the theorem we obtain

$$f^*\leq f(x^*)=L(x^*,\lambda^*,\mu^*)=\min_{x\in X}L(x,\lambda^*,\mu^*)=q(\lambda^*,\mu^*)\leq q^*.$$

The the equiality

$$f^*=q^*$$

follows from the propositions (26.1.2) and (14.1). ■

27 Conjugate duality.

Prop. 27.1 *For any convex function f*

$$f(x)=\sup_{\{b,\beta|\langle x,b\rangle-\beta<f(x),\forall x\}}\left(\langle x,b\rangle-\beta\right)$$

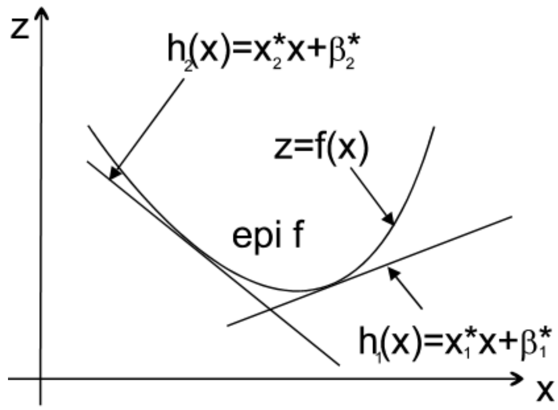

Figure 1.16: *epi* f is the intersection of the upper half planes.

Proof. Each affine function $h(x) = \langle x, b\rangle - \beta$ corresponds to a hyperplane. By the proposition (12.2), for any point below epigraph of f there exists a hyperplane that separates such point from the *epi* f. Hence, for any such point $(x, \mu), x \in dom\ f\ \mu < f(x)$ there exists a pair (b, β) s.t. $\langle x, b\rangle - \beta < 0$ and $\langle y, b\rangle - \beta < f(y)$ for $\forall y \in dom\ f$. ■

Coroll. 27.2 *The set epi f is equal to intersection of the upper half-planes defined by the hyperplanes from proof of the previous statement.*

Let us denote $F = epi\ f$. Let us introduce a set

$$F^* = \{(x^*, \beta^*) | \forall x \in dom\ f,\ \langle x^*, x\rangle - \beta^* \le f(x)\}.$$

Such set is not empty and it is an epigraph of some function because if $(x_0, \beta_0) \in F^*$then $(x_0, \beta) \in F^*$ for all $\beta > \beta_0$. Let us denote such function f^*. By definition

$$\begin{aligned} f^*(x^*) &= \inf\{\beta^* | \forall x \in dom\ f\colon\ \langle x^*, x\rangle - \beta^* \le f(x)\} \\ &= \inf\{\beta^* | \forall x \in dom\ f\colon\ \langle x^*, x\rangle - f(x) \le \beta^*\} \\ &= \sup_{x \in dom\ f} \{\langle x^*, x\rangle - f(x)\}, \end{aligned}$$

$$dom\ f^* = \{x^* | \langle x^*, x\rangle - f(x) < +\infty,\ \forall x \in dom\ f\}.$$

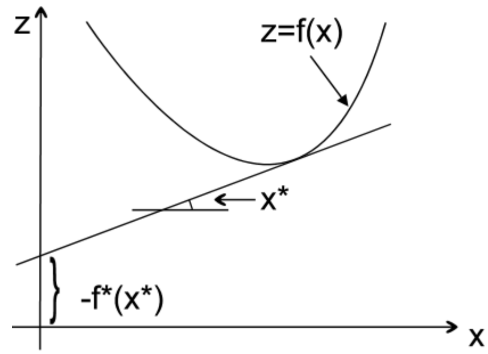

Figure 1.17: Geometrical meaning of $f(x^*)$.

Let us introduce

$$F^{**} = \left\{ (x^{**}, \beta^{**}) \,|\forall x^* \in dom\ f^*,\ \ \langle x^{**}, x^* \rangle - \beta^{**} \le f^*(x^*) \right\},$$

$$\begin{aligned} f^{**}(x^{**}) &= \inf \left\{ \beta^{**} | \forall x^* \in dom\ f^*:\ \langle x^{**}, x^* \rangle - \beta^{**} \le f^*(x^*) \right\} \\ &= \sup_{x^* \in dom\ f^*} \left\{ \langle x^{**}, x^* \rangle - f^*(x^*) \right\}, \end{aligned}$$

$$dom\ f^{**} = \left\{ x^{**} |\, \langle x^{**}, x^* \rangle - f^*(x^*) < +\infty,\ \forall x^* \in dom\ f^* \right\}.$$

Observe that $dom\ f^{**} = dom\ f$. Indeed,

$$\langle x^{**}, x^* \rangle - f^*(x^*) < +\infty,\ \forall x^* \in dom\ f^*$$

means that $\exists c = const$

$$\langle x^{**}, x^* \rangle - f^*(x^*) < c$$

or

$$\langle x^{**}, x^* \rangle - c < f^*(x^*)$$

where the x^* and $\beta = f^*(x^*)$ run through all the hyperplanes that define $epi\ f$. For the same reason the infimum of such c is the $f(x^{**})$:

$$\begin{aligned} f(x^{**}) &= \inf \left\{ c | \forall x^* \in dom\ f^*:\ \langle x^{**}, x^* \rangle - f^*(x^*) \le c \right\} \\ &= \sup_{x^* \in dom\ f^*} \left\{ \langle x^{**}, x^* \rangle - f^*(x^*) \right\}. \end{aligned}$$

Summm. 27.3 *(Conjugate duality theorem). We define the operation of taking a dual function f^* by*

$$f^*(x^*) = \sup_{x \in dom\ f} \{\langle x^*, x\rangle - f(x)\}.$$

Then

$$f^{**} = cl\ f$$

for all proper convex functions.

The closure part: $cl\ f$ comes from the fact that taking affine envelopes includes boundary points of the $epi\ f$ into the final result $epi\ f^{**}$.

27.1 Support function.

Def. 27.1.1 *The indicator function $\delta(\cdot|C)$ of a convex set C is a function of the form*

$$\delta(x|C) = \begin{pmatrix} 0, \textit{if } x \in C \\ +\infty, \textit{ if } x \notin C \end{pmatrix}.$$

The support function $\delta^*(\cdot|C)$ is the conjugate of the indicator function.

According to the definition

$$\begin{aligned} \delta^*(x^*|C) &= \sup_{x \in dom\ \delta(\cdot|C)} \{\langle x^*, x\rangle - \delta(x|C)\} \\ &= \sup_{x \in C} \langle x^*, x\rangle. \end{aligned} \tag{1.14}$$

Note that $\delta^*(x^*|C)$ is a positively homogenous function of x^*. Suppose $f(x)$ is some proper convex positively homogenous function. Consider the conjugate

$$f^*(x^*) = \sup_{x \in dom\ f} \{\langle x^*, x\rangle - f(x)\}.$$

By the positive homogenuity $x \in dom\ f \Rightarrow \lambda x \in dom\ f,\ \forall \lambda > 0$. Consequently, for any $\lambda > 0$,

$$\sup_{x \in dom\ f} \{\langle x^*, x\rangle - f(x)\} = \sup_{x \in dom\ f} \{\langle x^*, \lambda x\rangle - f(\lambda x)\} = \lambda \sup_{x \in dom\ f} \{\langle x^*, x\rangle - f(x)\}.$$

Hence, $f^*(x^*)$ is either 0 or $+\infty$. Introduce the set

$$C = \{x^* | \forall x, \ \ \langle x^*, x\rangle \leq f(x)\}.$$

Such set is a *dom* f^*. Indeed, if $\langle x^*, x\rangle \leq f(x)$ then $\langle x^*, x\rangle - f(x) \leq 0$ and 0 is reached by scaling with λ. Then $f^*(x^*) = 0$. On the other hand if $\langle x^*, x\rangle > f(x)$ then $\langle x^*, x\rangle - f(x) > 0$ and $+\infty$ is reached by scaling.

Summ. 27.1.2 *(Convex homogenous function property). If $f(x)$ is a proper convex positively homegenous function then*

$$\begin{aligned} f^*(x^*) &= \delta(x^* | C), \\ C &= \{x^* | \forall x, \ \ \langle x^*, x\rangle \leq f(x)\} \end{aligned}$$

and

$$\delta^*(x^* | C) = cl \ f.$$

The last part of the summary follows from the summary (27.3). We established one-to-one correspondence between convex sets and proper convex positively homegenous functions.

References

A. S. Lewis & J. M. Borwein: Convex Analysis and Nonlinear Optimization.

R.Tyrrell Rockafellar: Convex Analysis.

Dimitri P. Bertsekas: Convex Analysis and Optimization.

Dimitri P. Bertsekas: Convex Optimization Theory.