SOME REMARKS ON CONVEX SETS

By
TESFA YIGREM

Advisor

Prof.Dr.rer.nat.habil.R.DEUMLICH

A paper Submitted in partial fulfillment as a requirement for the M.sc degree in Mathematics

School of Graduate Studies
Addis Ababa University

June, 2002
Addis Ababa

Preface

This paper is produced as a requirement in partial fulfillment of the master of science degree in Mathematics at Addis Ababa University. It compiles all the presentations I made in the examinors board.

Today the aggregate of subjects called Optimization is a captivating merge of heuristics and rigour of theory and experiment. Truly it can be studied as a branch of pure mathemtics, yet has a wide ranging applications in almost every branch of science and technology. The paper attempts to forward some issues on convex sets which have vital role for optimization in particular and modern mathematics in general. The paper comprises of four main chapters organized in order of dependence of one on the other.

The first chapter introduces basic definitions, theorems and examples on convex sets that seem simple although they are of crucial importance as serving as the basis of many optimization concepts. The second chapter where the theme of the paper commences makes a survey of projections onto a closed convex sets. Optimization is often characterized as the science of determining the 'best' solutions to certain mathematically defined problems which are mostly models of physical reality. It also invloves the study of optimality conditions for problems . In this section too a constrained distance minimization problem is set forth and a corresponding method is offered to solve. Emphesis is then paid to develop necessary and sufficient conditions for optimality. Depending on the closed convex sets in consideration , special optimality criteria are also provided. Furthermore, attempt is made to point out properties of the projection operator.

The third chapter deals about one of the basic properties of convex sets namely separation of convex sets. The important separation theorems are formulated and discussed in detail. Although the Separation Theorem has far-reaching applications, existences of a supporting hyperplane, outer description of a closed convex set, bipolar of a cone and the Farkas' Lemma are some of them where the paper gives due emphesis. The last chapter provides a useful method to approximate convex sets by cones.

I want to express my deepest gratitude to my indefatigable advisor and instructor Pro.Dr.rer.nat.habil.R.Deumlich who cultivated me not only accademically but also in all other social aspects. I strongly apperciate his gallant interest to teach mathematics and computer with decent methods of instruction. His constructive comments, advices and provision of materials like books, projectors and diskettes are extremly invaluable and ever unforgottable. Getting him as my adivisor and instructor really pleased me overwhelimingly.

I am grateful to all my friends for their unreserved moral and material support that encourged and made fruitful my study. Among the ~~money~~ many, Yibeltal Yitayew and Workefrahu Seyoum deserve worth mentioning. I also extend my sincere gratitude to my room mates, Abebe Bezie and Bezunesh Abera, who made my life a pleasure through out my study.

Tesfa Yigrem

Table of Contents

Contents **Page**

1. Introduction

The prime working space of this paper is the pre-hilbert space $\mathbf{R}^n$. The space $\mathbf{R}^n$ can be possibly equipped with the scalar product defined by

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i, \quad x = (x_i),\ y = (y_i),\ i = 1, 2, ..., n.$$

However, effort is made at times to relax the situation in more mathematical settings. The space $\mathbf{R}^n$ is also complete with the norm $\|x\| = \sqrt{\langle x, x\rangle}$ and hence a Hilbert space. The materials presented are ingeneral basic and noteworthy not only for optimization but also have diverse ranging applications reaching into other branches of modern mathematics.

Pertinent definitions, basic theorems, side remarks helping to understand delicate points and to indicate rooms for further investigation in other spaces, solved problems that serve to illustrate definitions and applications of theorems, and other descriptive concepts mainly constitute the body of the paper.

1.1 Convex Sets

In modern Mathematics way or convexity plays in one another fundamental role . The interdisciplinary field, optimization, is to be worth mentioning where it takes a central place. Many optimization problems are either convex or can be transferred into a convex problem. Convex sets have also a widespread applications in mathematical programming which is of course an integral part of optimization theory. Due to its basic importance here too we deal with convex sets at some lengh.

Definition1.1.1: A set $C \subseteq \mathbf{R}^n$ is said to be convex if and only if

$$\alpha x + (1-\alpha)y \in C \text{ for each } x, y \in C \text{ and } \alpha \in [0,1].$$

It means geometrically the line segment

$$[x, y] := \{z \in R^n \mid z = \alpha x + (1-\alpha)y,\ \alpha \in [0, 1]\}$$

is entirely contained in C whenever its end points x and y are in C.

Said otherwise, C is convex if and only if

$$[x, y] \subseteq C \text{ for all } x, y \in C.$$

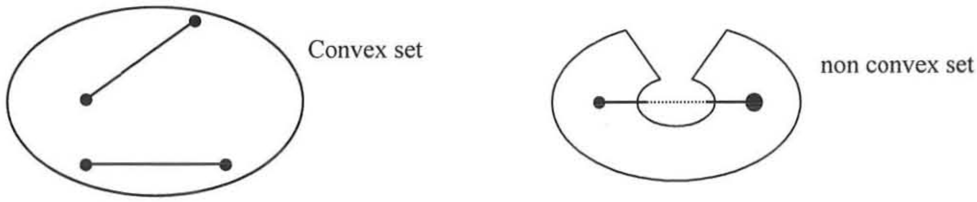


Fig.1.1.1, Convex and nonconvex sets

An immediate consequence of the definition is that any two points of C can be linked by a continuous path. That is C is also path connected.

Example1.1.1 Affine Hyperplanes

Definition1.1.2: Let $(s, r) \in \mathbf{R}^n \times \mathbf{R}$, $s \neq 0$. Then the set

$$H_{s,r} := \{x \in \mathbf{R}^n \mid \langle s, x\rangle = r\}$$

is said to be a hyperplane. If r = 0, then thesubspace $H_{s,0}$ is called a linear or a vector hyperplane.

For x, y $\in H_{s,r}$ and $\alpha \in [0,1]$

$$\langle s, \alpha x + (1-\alpha y)\rangle = \alpha r + (1-\alpha) r = r.$$

$H_{s,r}$ is therefore a convex set. Fix s and let r describes **R**. Then the hyperplanes $H_{s,r}$ are translations of the same linear or vector hyperplane $H_{s,0}$. The hyperplane $H_{s,r}$ in $\mathbf{R}^n$ is always n -1 dimensional and thus of codimension (the dimension of its coplementary space) 1. $H_{s,r}$ can be also considered as the solution set of the linear equation

$$\langle s, x\rangle = r, \; x \in \mathbf{R}^n.$$

Hyperplanes play paramount role in separation of convex sets: point of emphesis in the sequal.

Example1.1.2 Affine Manifolds

Definition1.1.3: Let $V \subseteq \mathbf{R}^n$. Then V is said to be an affine manifold if and only if

$$\alpha x + (1-\alpha)y \in V \text{ for each } x, y \in V \text{ and } \alpha \in \mathbf{R}.$$

Clearly affine manifolds are convex sets. Indeed convex sets are special case of affine manifolds.

For $v \in V$, consider the set

$$M := V-\{v\}.$$

Let $x \in M$ and $\alpha \in \mathbf{R}$. Then

$$\alpha x = \alpha (v_1 - v) = (\alpha v_1+(1-\alpha)v) - v \in V \text{ where } v_1 \in V.$$

Furthermore, for $x, y \in M$

$$x + y = (v_1 - v) +(v_2 - v) = (v_1 + v_2) - v \in M \text{ where } v_1,v_2 \in V.$$

M is thus a subspace of $\mathbf{R}^n$ which is independent of the particular choice of v. We can therefore equivalently define affine manifolds as translation of a subspace.

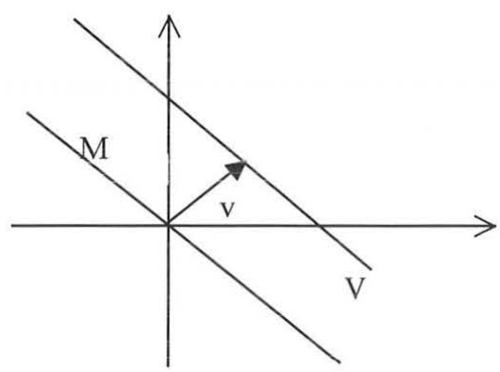


Fig.1.1.2, Translating a subspace

Example 1.1.3 Halfspaces of $\mathbf{R}^n$

Definition1.1.4: Let $(s, r) \in \mathbf{R}^n \times \mathbf{R}$, $s \neq 0$. Then the sets

(i) $\{x \in \mathbf{R}^n \mid \langle s, x\rangle \leq r\}$ and

$\{y \in \mathbf{R}^n \mid \langle s, y\rangle \geq r\}$

are called the closed halfspaces.

(ii) $\{x \in \mathbf{R}^n \mid \langle s, x\rangle < r\}$ and

$\{y \in \mathbf{R}^n \mid \langle s, y\rangle > r\}$

are called the open halfspaces.

Obviously, the halfspaces of $\mathbf{R}^n$ are convex sets. An open halfspace is naturally an open set being the interior of the corresponding closed halfspace and a closed halfspace is a closed set being the closure of the corresponding open halfspaces. Clearly the boundaries of the halfspaecs are hyperplanes.

1.1 Convexity Preserving Operations

Convexity is stable under some mathematical operations. In what follows we give some of the pertinent once.

Theorem1.2.1: Let J be an arbitrary index set and C, $C_j \subseteq \mathbf{R}^n$, i∈J be convex sets. Furthermoere, let the convex sets $A_i \subseteq \mathbf{R}^{n_i}$ for i = 1, 2, ..., m. Then

(i) $\bigcap_{j\in J} C_j$,

(ii) $\alpha A_1 + \beta A_2 := \{x \mid x = \alpha x_1 + \beta x_2, x_1 \in A_1, x_2 \in A_2, \alpha, \beta \in \mathbf{R}\}$,

(iii) $A_1 \times A_2 \times \ldots \times A_m$,

(iv) $T(C) = \{z \in \mathbf{R}^{n_1} \mid z = T(x), x \in C\}$ for a linear mapping T: $\mathbf{R}^n \to \mathbf{R}^{n_1}$ are convex sets.

Proof: (i) Let x, y $\in \bigcap_{j\in J} C_j$. Then x, y $\in C_j$ for each j ∈ J.

By the convexity of C_j , for each j∈ J we get

$$\lambda x + (1-\lambda)y \in C_j \text{ for all } j \in J \text{ and for all } \lambda \in [0, 1].$$

Thus

$$\lambda x + (1-\lambda)y \in \bigcap_{j\in J} C_j .$$

And hence

$\bigcap_{j\in J} C_j$ is always a convex set.

In contrast to this operation, the union of convex sets may not be convex.

(ii) Let x, y $\in \alpha A_1 + \beta A_2$. Then

$$x = \alpha x_1 + \beta x_2,\ y = \alpha y_1 + \beta y_2, \text{ for some } y_1, x_1 \in A_1 \text{ and } y_2, x_2 \in A_2.$$

From this

$$\begin{aligned}\lambda x + (1-\lambda)y &= \lambda(\alpha x_1 + \beta x_2) + (1-\lambda)(\alpha y_1 + \beta y_2),\\ &= \lambda(\alpha x_1 + \beta x_2) + (1-\lambda)(\alpha y_1 + \beta y_2),\\ &= \alpha(\lambda x_1 + (1-\lambda)y_1) + \beta(\lambda x_2 + (1-\lambda) y_2),\\ &= \alpha z_1 + \beta z_2\end{aligned}$$

where $z_1 := \lambda x_1 + (1-\lambda)y_1 \in A_1$ and $z_2 := \lambda x_2 + (1-\lambda) y_2 \in A_2$ for all $\lambda \in [0,1]$.

This implies that

$$\lambda x + (1-\lambda)y = \alpha z_1 + \beta z_2 \in \alpha A_1 + \beta A_2.$$

(iii) Let $x, y \in A_1 \times A_2 \times \ldots \times A_m$. Then

$$x = (x_1, x_2, \ldots, x_m) \text{ and } y = (y_1, y_2, \ldots, y_m) \text{ where } x_i, y_i \in A_i \text{ for } i = 1, 2, \ldots, m.$$

For $\lambda \in [0, 1]$ we have that

$$\lambda x + (1-\lambda)y = (\lambda x_1 + (1-\lambda)y_1, \lambda x_2 + (1-\lambda)y_2, \ldots, \lambda x_m + (1-\lambda)y_m)$$
$$= (z_1, z_2, \ldots, z_m)$$

where $z_i := \lambda x_i + (1-\lambda)y_i \in A_i$ for each $i = 1, 2, \ldots, m$.

Hence

$$\lambda x + (1-\lambda)y \in A_1 \times A_2 \times \ldots \times A_m.$$

(iV) Let $z_1, z_2 \in T(C)$. Then

$$z_1 = T(x),\ z_2 = T(y), \text{ for some } x, y \in C.$$

For $\lambda \in [0, 1]$, we thus have

$$\lambda z_1 + (1-\lambda)z_2 = \lambda T(x) + (1-\lambda)T(y) = T(\lambda x + (1-\lambda)y) = T(z) \in T\ (C)$$

where $Z := \lambda x + (1-\lambda)y \in C$. As a result the image of convex sets under a linear mapping is convex. □

Remark1: Theorem 1.2.1 is also valid in any vector space X with the same proof.

Example 1.1.4: Let $(s_1, r_1), (s_2, r_2), \ldots, (s_m, r_m)$ be m given elements of $\mathbf{R}^n \times \mathbf{R}$. Then Consider the set

$$A := \{x \in \mathbf{R}^n \mid \langle s_i, x\rangle \le r_i \ \ i = 1, 2, \ldots, m\}.$$

A is then a convex set which can be confirmed by considering it as the intersection of m convex halfspaces.

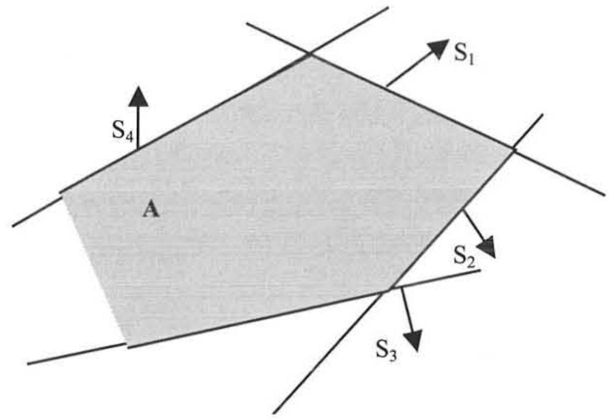


Fig.1.2.1, Intersection of halfspaces

Afruitful observation is that the above construction can be also applied to the examples in section(1.1). For instance a hyperplane is the intersection of two (closed) halfspaces and an affine manifold is inturn the intersection of finitely many affine hyperplanes. Piecing these and other examples of convex sets we notice that they all can be considered as the interesection of sufficiently many closed halfspaces. Accordingly, the operation making intersections and the closed halfspaces are basic objects used frequently in the space of convex analysis. We conclude this section with two toplogical operations preserving convexity.

Theorem1.2.2: Let $C \subseteq \mathbf{R}^n$ be a convex set. Then its interior (intC) and closure (clC or $\overline{C}$) are convex sets .

Proof: Let x, y $\in$ intC and B(0,δ) be a neighbourhood of the origin 0 such that x +B(0, δ) is in C. For $\lambda \in (0, 1)$ we have that

$$\lambda B(0, \delta) \text{ is also a neighbourhood of } 0.$$

By the convexity of C it follows that

$$\lambda x + (1-\lambda)y + \lambda B(0,\delta) = \lambda (x + B(0,\delta)) + (1-\lambda)y \subseteq C .$$

As a result

$$\lambda x + (1-\lambda)y \in \text{intC for each } \lambda \in (0, 1).$$

Regarding to the convexity of the closure, consider x, y $\in$ clC and the sequences (x_k), (y_k) in C such that

$$\lim_{k\to\infty} x_k = x, \quad \lim_{k\to\infty} y_k = y .$$

Since C is convex we have that

$$\lambda x_k + (1-\lambda)y_k \in C \text{ for all } \lambda \in (0,1) \text{ and } k \in \mathbf{N}.$$

Thus

$$\lambda x + (1-\lambda)y = \lambda \lim_{k\to\infty} x_k + (1-\lambda) \lim_{k\to\infty} y_k = \lim_{k\to\infty} (\lambda x_k + (1-\lambda)y_k) \in \text{cl}C .$$

All convex sets have therefore convex interior and convex closure. □

1.3 Convex Combinations and Convex hull

The operations get implemented in section (1.2) took convex sets and made new convex sets out of them. The present is however focused on operation which takes nonconvex set and makes a new convex set with it.

Definition1.3.1: Let $x_1, x_2, ..., x_k \in \mathbf{R}^n$, $\lambda_1, \lambda_2, ..., \lambda_k \in \mathbf{R}$. Then

$$z := \sum_{i=1}^{k} \lambda_i x_i, \text{ where } \sum_{i=1}^{k} \lambda_i = 1$$

is said to be an affine combination of the points $x_1, x_2, ..., x_k$.

From its definition an affine combination is a particular linear combination. An affine manifold contains all its affine combinations. The intersection of affine manifolds is obviously an affine manifold. For a given set S we can therefore associate the intersection of all affine manifolds containing it.

Definition 1.3.2: Let $S \subseteq \mathbf{R}^n$. Then the set

$$\text{aff}S := \bigcap\{K \mid S \subseteq K \subseteq R^n, \text{ K affine manifold}\}$$

is said to be the **affine hull** of S.

Clearly, affS is the affine manifold generated by S or the smallest (in the sense of inclusion) affine manifold containing S. It can be constructed directly from the set S by collecting all the affine combinations of elements of S. This can be seen easily in the way to follow. Start from a point $x_0 \in S$ and take the linear hull of the set S - $\{x_0\}$ (lin(S-x_0)) and come back by adding x_0 to the result x_0 + lin(S-x_0) which is just affS. For a convex set

S with $\text{int}S \neq \phi$, affS expands to the whole space $\mathbf{R}^n$. By definition it follows that the operation "taking a hull" is monotone increasing in the sense that if $S_1 \subseteq S_2$, then

$$\text{aff}S_1 \subseteq \text{aff}S_2.$$

Definition1.3.2: Let $x_1, x_2, \ldots, x_k \in \mathbf{R}^n$, $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbf{R}$. Then

$$z := \sum_{i=1}^{k} \lambda_i x_i, \text{ where } \sum_{i=1}^{k} \lambda_i = 1, \lambda_i \geq 0, i = 1, 2, 3, \ldots, k$$

is called a convex combination of the points $x_1, x_2, \ldots, x_k$.

A convex combination is again a particular affine combination. Since convexity is stable under intersection, for a given set S we can also intersect all convex sets containing it.

Definition1.3.4: Let $S \subseteq \mathbf{R}^n$. Then the set

$$\text{conv}S: == \bigcap\{K \mid S \subseteq K \subseteq R^n, \text{K convex}\}$$

is said to be the **convex hull** of S.

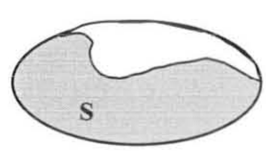


Fig.1.3.1, Convex hull of

The convex hull of S is the smallest (in the sense of inclusion) convex set containing S and is also equal to the set of all convex combination of finite number of points in S.

2. Projection onto a closed Convex set

Apart from the intrinsic interest of the subject, the study of optimization techniques is imperative due to its very wide field of applications. Optimization problems apear almost daily in operational research, economics, statestics, geometry, physics, control theory, chemical engineering, in many other subjects and social affairs. Indeed it has been claimed that every one is optimizing (choosing the best out of the possible) some thing all the time. In the section set forth an elegant techinque is displayed in order to solve a constrained minimization problem.

Let $C \subseteq \mathbf{R}^n$ be a non-empty closed set. Fix a point $x \in \mathbf{R}^n$ and consider the constrained optimization problem

$$\min \left\{ \frac{1}{2}\|x - y\|^2 \mid y \in C \right\}. \tag{2.1}$$

Here we are interested to search out those points of C that are nearest to x for the Euclidean distance. In doing so, let $\beta := \inf \left\{ \frac{1}{2}\|x - y\|^2 \mid y \in C \right\}$. Then by definition of the infimum there is a sequence (y_n) in C such that

$$\frac{1}{2}\|x - y\|^2 \xrightarrow{n \to \infty} \beta \cdot$$

Consequently, for a positive number ε, there is an index n_0 such that

$$\frac{1}{2}\|x - y\|^2 \leq \beta + \varepsilon \text{ for all } n \geq n_0.$$

or

$$\|x - y\| \leq (2\beta + 2\varepsilon)^{\frac{1}{2}} =: \alpha \text{ for all } n \geq n_0.$$

Thus

$$\|y_n\| = \|y_n - x + x\| \leq \|y_n - x\| + \|x\| \leq \alpha + \|x\| \text{ for all } n \geq n_0.$$

Therefore the sequence (y_n) is bounded in the finite dimensional space $C \subseteq \mathbf{R}^n$. Hence by the Bolzano-Weierstrass Theorem there is a subsequence (y_{n_j}) such that

$$(y_{n_j}) \xrightarrow{j \to \infty} y_x \in \mathbf{R}^n.$$

But since $C \subseteq \mathbf{R}^n$ is closed we get

$y_x \in C$.

Furthermore, by the continuity of the norm we have that

$$\beta = \lim_{j\to\infty} \frac{1}{2}\|y_{n_j} - x\|^2 = \frac{1}{2}\|y_x - x\|^2 .$$

Consequently by the definition of β,

$$\frac{1}{2}\|y_x - x\|^2 \leq \frac{1}{2}\|y - x\| \quad \text{for each } y \in C.$$

Therefore there always exists a point $y_x \in C$ nearest to x and

$$\frac{1}{2}\|y_x - x\|^2 = \min_{y\in C} \frac{1}{2}\|y - x\|^2 . \quad \square \tag{2.2}$$

Remark1: For a normed space X, the existence result above is valid for a complete space C or when $C \subseteq V$ where V is a finite dimensional subspace of X.

The materials presented so far only addresses the solvabilty of (2.1). The usual unescapable question inseparable with existence is that of uniqueness. But under only the aforementioned hypotheses, the solution need not be unique. Uniqueness crucially needs additional supposition. The next lemma gives the answer what has to be added to formulate unique solvability.

Lemma 2.1.1: Let $C \subseteq \mathbf{R}^n$ be a non-empty, closed and convex set. Then for each $x \in \mathbf{R}^n$,

$$\min_{y\in C} \frac{1}{2}\|y - x\|^2 \tag{2.3}$$

is unique solvable.

Proof: Let $f(y) := \frac{1}{2}\|y - x\|^2$, $y \in C$. Let y_1 and y_2 be two different solutions of (2.3). That is

$$f(y_1) = f(y_2) \leq f(y) \text{ for all } y \in C.$$

As C is convex we get that

$$\frac{y_1 + y_2}{2} \in C.$$

Applying the parallelogram equation we have

$$f\left(\frac{y_1 + y_2}{2}\right) = \frac{1}{2}\left\|\frac{y_1 + y_2}{2} - x\right\|^2 = \frac{1}{2}\left\|\frac{y_1 - x}{2}\right\|^2 + \frac{1}{2}\left\|\frac{y_2 - x}{2}\right\|^2 ,$$

$$= \frac{1}{4}\|y_1 - x\|^2 + \frac{1}{4}\|y_2 - x\|^2 - \frac{1}{8}\|y_1 - y_2\|^2,$$

$$= \frac{1}{2}f(y_1) + \frac{1}{2}f(y_2) - \frac{1}{8}\|y_1 - y_2\|^2,$$

$$= f(y_1) - \frac{1}{8}\|y_1 - y_2\|^2 < f(y_1).$$

But this is a contradiction. □

Remark2: The uniqueness in Lemma 2.1.1 is also valid for a non-empty, closed and convex subset C of a pre- Hilbert space X with the same proof.

Definition 2.1.1: Let $C \subseteq \mathbf{R}^n$ be a non-empty, closed and convex set. Then the mapping

P_C: $\mathbf{R}^n \longrightarrow C$ where

$$P_C(x) = y_x \text{ and } \frac{1}{2}\|y_x - x\|^2 = \min \frac{1}{2}\|y - x\|^2, \; y \in C$$

is called a **Projection.**

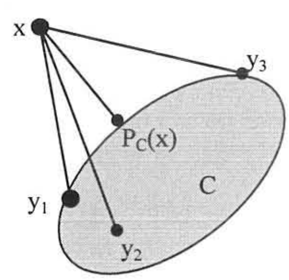


Fig.2.1.1, Projection onto a closed convex set

Definition.1.2. Let $C \subseteq \mathbf{R}^n$ and $x \in \mathbf{R}^n$. Then $y_0 \in C$ is said to be the best approximation of x with regard to C if and only if

$$\|y_0 - x\| = \min_{y \in C} \|y - x\|.$$

Let $f(y) := \frac{1}{2}\|y - x\|^2$ and $g(y) := \|y - x\|$, $y \in C$. Then since f is non-negative and monotone increasing, f and g have the same minimum point in C. Accordingly, $P_C(x)$ is the best approximation of x with regard to C. The projection operator has remarkable application in separation of convex sets; a point of discussion in Chapter 3.1.

The main difficulties in optimization is to arrive at the best possible solution in any given set of problems. Of course, many situations arise where the" best" is unattainable for one reason or another. Actually what is " best" to some body may be "worest "to some one else. In the restricted situation here however the best acheivable and indeed what follows is a necessary and sufficient condition when a point to be the projection of another point.

Theorem 2.1.2: (The Characterization Theorem)

Let $C \subseteq \mathbf{R}^n$ be a non-empty, closed and convex set. A point $y_x \in C$ is the projection $P_C(x)$ if and only if

$$\langle x - y_x, y - y_x \rangle \leq 0 \text{ for all } y \in C. \tag{2.4}$$

Proof: Let $y_x = P_C(x)$ and $f(y) := \frac{1}{2}\|y - x\|^2$. Then by the convexity of C

$$y_x + \lambda(y - y_x) \in C \text{ for all } \lambda \in (0, 1) \text{ and } y \in C.$$

Now, $f(y_x) = \frac{1}{2}\|y_x - x\|^2 \leq f(y_x + \lambda(y - y_x)) = \frac{1}{2}\|y_x + \lambda(y - y_x)\|^2$,

$$= \frac{1}{2}[\ \|y_x - x\|^2 + \lambda^2\|y - y_x\|^2 + 2\lambda\langle y - y_x, y_x - x\rangle\ ],$$

or

$$0 \leq \frac{\lambda^2}{2}\|y - y_x\|^2 + \lambda\langle y - y_x, y_x - x\rangle \cdot \tag{2.5}$$

Dividing (2.5) by $\lambda > 0$ results

$$0 \leq \lambda\|y - y_x\|^2 + \langle y - y_x, y_x - x\rangle .$$

Letting $\lambda \to 0$ and using properties of the scalar product we have that

$$\langle x - y_x, y - y_x \rangle \leq 0 \qquad \text{for all } y \in C.$$

Conversely, let $y_x \in C$ and $\langle x - y_x, y - y_x \rangle \leq 0$ for all $y \in C$.

Case1: If $y_x = x$, then trivially $P_C(x) = y_x$.

Case2 : If $y_x \neq x$, then

$$0 \geq \langle x - y_x, y - y_x \rangle = \langle x - y_x, y - x + x - y_x \rangle,$$
$$= \|x - y_x\|^2 + \langle x - y_x, y - x \rangle.$$

Applying the Cauchy-Schewarz inequality gives

$$0 \geq \|x - y_x\|^2 - \|x - y\|\|x - y_x\|. \tag{2.6}$$

dividing (2.6) by $\|x - y_x\| > 0$ yields,

$$\|x - y\| \geq \|x - y_x\| \text{ for all } y \in C.$$

Which means that $P_C(x) = y_x$. □

The geometrical meaning of the Charactrization Theorem is easy to apperciate. From (2.4) we have

$$\langle x - y_x, y - y_x \rangle = \|x - y_x\|\|y - y_x\| \cos\theta \leq 0 \quad \text{for all } y \in C.$$

This implies that

$$\theta \in [\frac{\pi}{2}, \pi].$$

In other words the angle between $x - y_x$ and $y - y_x$ is obtus for all $y \in C$.

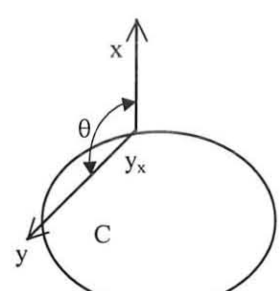


Fig.2.1.2, The angle characterization of a projection

Remark 3: Obvously, the Characterziation Theorem can be also extended to any pre-Hilbert space with the same proof.

By means of the scalar we can define the notion of orthogonality as below.

Definition 2.1.3: Let $C \subseteq \mathbf{R}^n$. Then

(a) $x, y \in \mathbf{R}^n$ are said to be orthogonal ($x \perp y$) if and only if

$$\langle x, y \rangle = 0$$

(b) $C^{\perp} := \{x \in \mathbf{R}^n \mid x \perp y, \text{ for each } y \in C\}$

is called an orthogonal space of C.

Corollary 2.1.3: Let C be a subspace of $\mathbf{R}^n$ and $y_x \in C$. Then $P_C(x) = y_x$ if and only if

$$\langle x - y_x, y\rangle = 0 \text{ i.e, } x - y_x \in C^{\perp} \text{ for all } y \in C.$$

Proof: Since C is a subspace of $\mathbf{R}^n$, $y - y_x \in C$ for each $y \in C$.

Employing the Characterization Theorem we have that

$$\langle x - y_x, y_x - y - y_x\rangle = \langle x - y_x, -y\rangle \leq 0 \text{ for all } y \in C. \tag{2.7}$$

or

$$\langle x - y_x, y\rangle \geq 0 \text{ for each } y \in C.$$

And again (2.7) implies that

$$\langle x - y_x, y\rangle \leq 0 \quad \text{for all } y \in C.$$

Hence

$$\langle x - y_x, y\rangle = 0 \text{ for all } y \in C.$$

Conversely, let $x - y_x \perp y$ for each $y \in C$. Since C is a subspace we get

$$y = y_x + z, \ z \in C.$$

To this effect,

$$\frac{1}{2}\|x - y\|^2 = \frac{1}{2}\|x - y_x - z\|^2 = \frac{1}{2}\langle x - y_x - z, x - y_x - z\rangle \cdot$$

Thus

$$P_C(x) = y_x. \qquad \square$$

Passing from Corollary 2.1.3 to (2.4) shows that convex analysis is the realm of inequalities in contrast with linear analysis.

Example: Let $X = l_R^2 = \{ (y_n) \mid \sum_{n=1}^{\infty} (y_n)^2 < \infty, y_n \in \mathbf{R}, n \in N \}$. Then X is a pre-Hilbert space equipped with the scalar product defined by

$$\langle z, y\rangle = \sum_{i=1}^{\infty} y_i z_i, \ y = (y_i), \ z = (z_i), \ i \in N.$$

Let $x = \begin{cases} n, & \textit{if } n \leq 4 \\ 0, & \textit{otherwise} \end{cases}$, $n \in N$ and $C := \{ y \in X \mid y = \alpha x_0, \alpha \in \mathbf{R}\}$ where

$$x_0 = (z_n) = \begin{cases} -(n+1), & \text{if } n = 2,3,4 \\ 0, & \text{otherwise} \end{cases} \quad n \in \mathbf{N}.$$

Obviously, $x \in l_R^2$ and C is a subspace of X. Let $y_x \in C$ be the projection $P_C(x)$.

Then

$$y_x = \beta x_0 \text{ for some } \beta \in \mathbf{R}$$

By corollary 2.1.3 we have

$$\langle x - \beta x_0,\ y\rangle = 0 \ \text{ for all } y \in C.$$

But $x - \beta x_0 = (1, 2+3\beta, 3+4\beta, 4+5\beta, 0, 0, 0, \dots)$.

Thus

$$\langle x - \beta x_o, y\rangle = -3\alpha(2 + 3\beta) - 4\alpha(3 + 4\beta) - 5\alpha(4 + 5\beta) = 0 \text{ for all } \alpha \in \mathbf{R}. \qquad (2.8)$$

Here (2.8) holds true if and only if

$$\beta = \frac{-19}{25} \text{ and hence } y_x = \beta x_0 = \frac{-19}{25} x_0 = P_C(x).$$

Definition2.1.4. 1) Let $\mathbf{R}^n = U \oplus U^{\perp}$. Then the projection

$P_U : \mathbf{R}^n \to U$ along $U^{\perp}$ is said to be the orthogonal projection of $\mathbf{R}^n$ onto U.

2) Let A: $\mathbf{R}^n \to \mathbf{R}^n$ be a linear mapping.Then A is said to be

a) symmetrical if and only if $\langle Ax, y\rangle = \langle x, Ay\rangle$ for all x, y $\in \mathbf{R}^n$.

b) bounded if and only if there is $\lambda \geq 0$ such that

$$\|A(x)\| \leq \lambda \|x\| \text{ for all } x \in \mathbf{R}^n.$$

c) nonexpansive if and only if $\|A(x)\| \leq \|x\|$ for all $x \in \mathbf{R}^n$.

3) A mappingA: $\mathbf{R}^n \to \mathbf{R}^n$ is said to be monotone increasing if and only if

$$\langle x - y, A(x) - A(y)\rangle \geq 0 \text{ for all } x, y \in \mathbf{R}^n$$

There is a fundamental connection between the prperty of A to be bounded and its continuity as given below.

Theorem 2.1.4: Let A: $\mathbf{R}^n \to \mathbf{R}^n$ be a linear mapping.Then

A is continuous if and only if A is bounded.

Proof: Let A be bounded. Then there is $\lambda \geq 0$ such that

$$\|A(x)\| \leq \lambda \|x\| \text{ for all } x \in \mathbf{R}^n.$$

Let $(x_n) \xrightarrow{n\to\infty} x$. Then we have that

$$\|A(x_n) - A(x)\| = \|A(x_n - x)\| \le \lambda \|x_n - x\| \xrightarrow{n\to\infty} 0.$$

It follows that

$$A(x_n) \xrightarrow{n\to\infty} A(x). \text{ That is A is continuous.}$$

Conversely, we prove the contrapositive. Suppose that A is not bounded. Then for each $n \in N$ there is an x_n such that

$$\frac{\|A(x_n)\|}{\|x_x\|} > n \text{ or } \|A(x_n)\| > n\|x_n\|.$$

Let $y_n := \frac{x_x}{\|x_n\|}$. Then we have $\|y_n\|=1$ and $\|A(y_n)\| = \left\|A(\frac{x_x}{\|x_n\|})\right\| = \frac{\|A(x_n)\|}{\|x_n\|} > n.$

That is

$$\|A(y_n)\| \xrightarrow{n\to\infty} \infty.$$

Let $z_n := \frac{y_x}{\|A(y_n)\|}$. Then $\|y_n\| = \frac{1}{\|A(y_n)\|} \xrightarrow{n\to\infty} 0.$

Now we assume that A is continuous. Then from $z_n \xrightarrow{n\to\infty} 0$ we get

$$A(z_n) \xrightarrow{n\to\infty} A(0) = 0.$$

That means

$$\|A(z_n)\| \xrightarrow{n\to\infty} 0.$$

But on the other hand we get

$$\|A(z_n)\| = \left\|A(\frac{y_x}{\|A(y_n)\|})\right\| = \frac{\|A(y_n)\|}{\|A(y_n)\|} = 1.$$

But this is a contradiction. So we proved that if A is not bounded, then A is not continuous. Hence if A is continuous, then A is bounded. □

Theorem 2.1.5: Let $C \subseteq \mathbf{R}^n$ be a non-empty, closed and convex set. Then

(i) The fixed points of P_C is C itself. That is

$$\{x \in \mathbf{R}^n \mid P_C(x) = x\} = C.$$

Consequently we get that

$$P_C^2 = P_C.$$

(ii) $\|P_C(x_1) - P_C(x_2)\|^2 \leq \langle P_C(x_1) - P_C(x_2),\ x_1 - x_2\rangle$ for all $x_1, x_2 \in \mathbf{R}^n$ and thus P_C is monotone increasing.

(iii) If C is a subspace of $\mathbf{R}^n$, then

a) P_C is a linear mapping and conversely.

b) P_C is an orthogonal projection.

c) P_C is continuous and symmetrical.

d) P_C is nonexpansive.

Proof: (i) follows simply by definition of P_C.

(ii) Applying the Characterzation Theorem we get

$$\langle x_1 - P_C(x_1),\ P_C(x_2) - P_C(x_1) \rangle \leq 0 \text{ and} \tag{2.9}$$

$$\langle x_2 - P_C(x_2),\ P_C(x_1) - P_C(x_2) \rangle \leq 0. \tag{2.10}$$

Adding (2.9) and (2.10) yields

$$\langle x_2 - x_1 + P_C(x_1) - P_C(x_2),\ P_C(x_1) - P_C(x_2) \rangle \leq 0$$

or

$$\|P_C(x_1) - P_C(x_2)\|^2 + \langle x_2 - x_1, P_C(x_1) - P_C(x_2)\rangle \leq 0.$$

This implies

$$\|P_C(x_1) - P_C(x_2)\|^2 \leq \langle x_1 - x_2,\ P_C(x_1) - P_C(x_2)\rangle. \tag{2.11}$$

From (2.11) we get

$$0 \leq \|P_C(x_1) - P_C(x_2)\|^2 \leq \langle x_1 - x_2,\ P_C(x_1) - P_C(x_2)\rangle \text{ for all } x_1, x_2 \in \mathbf{R}^n$$

That is

$$\langle x_1 - x_2,\ P_C(x_1) - P_C(x_2)\rangle \geq 0 \text{ for all } x_1, x_2 \in \mathbf{R}^n.$$

P_C is thus monotone increasing.

(iii) (a) follows trivially from Corollary 2.1.3 and $P_C(x) \in C$ for each $x \in \mathbf{R}^n$.

(b) $\ker P_C = \{x \in \mathbf{R}^n \mid P_C(x) = 0\} = \{x \in \mathbf{R}^n \mid \langle x - 0,\ y\rangle = 0 \text{ for } y \in C\} = C^{\perp}$

Obviously, $\operatorname{im} P_C = C$ and $C \oplus C^{\perp} \subseteq \mathbf{R}^n$.

Consider $x \in \mathbf{R}^n$, then $P_C(x) \in C$ and $x - P_C(x) \in C^{\perp}$.

So we have that

$$x = y_0 + P_C(x) \text{ for some } y_0 \in C^{\perp}.$$

Hence we get $\mathbf{R}^n \subseteq C \oplus C^{\perp}$.

Let $x_0 \in C \cap C^{\perp}$. Then

$$\langle x_0, x_0 \rangle = 0 \text{ i.e. } x_0 = 0.$$

Therefore $\mathbf{R}^n = C \oplus C^{\perp}$. Indeed this decomposition is unique and each $x \in \mathbf{R}^n$ has the unique representation $x = P_C(x) + P_C^{\perp}(x)$. To this effect the subspace of $\mathbf{R}^n$ can be characterized by the existence and the uniqueness of orthogonal projection.

(c) From (a) and (b) we have that P_C is an orthogonal linear operator.

Let $x \in \mathbf{R}^n$. Then

$$x = u + v,\ u \in \ker P_C = C^{\perp} \text{ and } v \in C.$$

Now, $P_C(x) = P_C(u + v) = P_C(u) + P_C(v) = 0 + P_C(v) = v$.

$$\|P_C(x)\|^2 = \|P_C(v)\|^2 = \|v\|^2 \le \|u\|^2 + \|v\|^2 = \langle u, u\rangle + \langle v, v\rangle = \langle u + v, u + v\rangle = \|x\|^2.$$

or

$$\|P_C(x)\| \le \|x\| \text{ for each } x \in \mathbf{R}^n. \tag{2.12}$$

This implies that

P_C is bounded .

Then by Theorem 2.1.4 P_C is continuous.

$$\begin{aligned}\langle P_C(x), x\rangle &= \langle P_C(x),\ x_1 - P_C(x_1) + P_C(x_1)\rangle, \\ &= \langle P_C(x),\ P_C(x_1)\rangle \text{ as } x_1 - P_C(x_1) \in C^{\perp}, \\ &= \langle P_C(x) - x + x,\ P_C(x_1)\rangle = \langle x,\ P_C(x_1)\rangle \text{ as } -x + P_C(x) \in C^{\perp} \text{ for}\end{aligned}$$

all $x_1,\ x_2 \in \mathbf{R}^n$.

This again implies that

P_C is a symmetrical operator.

d) Trivially follows from (2.12). □

2.2 Projection onto a closed convex cone

Definition 2.2.1: Let $K \subseteq \mathbf{R}^n$. Then K is called a cone if and only if

$$\alpha K \subseteq K \text{ for all } \alpha \in \mathbf{R}_+ \backslash \{0\}. \tag{2.13}$$

To illustrate the notion of a cone we consider the set

$$K: = \{x \in \mathbf{R}^2 | -2x_1 + x_2 \leq 0, x_1 - x_2 \geq 0 \}.$$

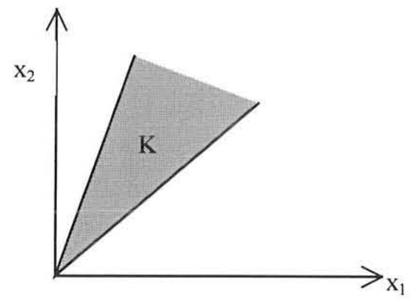


Fig.2.2.1, an example of a cone

Although a cone has a vertex in the usual representation of geometrical objects, a subspace is an example that fails to possess. The vertex in the above example is at the origin.

Obviously, (2.13) can be written to the form

$$x \in K \text{ implies } \alpha x \in K \text{ for all } \alpha \in \mathbf{R}_+ \backslash \{0\}.$$

We have that

$$\{\alpha x \mid \alpha > 0\} \subseteq K \text{ for all } x \in K.$$

That is the ray generated by the vector $x \in K$ is always in K. By definition a cone need not contain the element 0.

Examples: Let A be an (m,n)- matrix on **R**. We define $x \in \mathbf{R}^n$, $x \geq 0$ if and only if $x_i \geq 0$ for all $i \in \{1,2,...,n\}$. Then

(1) subspaces of $\mathbf{R}^n$;

(2) $\Omega_+ : = \{x \in \mathbf{R}^n \mid x \geq 0 \}$,

(3) $\{ x \in \mathbf{R}^n \mid Ax \geq 0\}$,

(4) $\{ y \in \mathbf{R}^n \mid y = Ax , x \geq 0, x \in \mathbf{R}^n\}$

are cones.

Convexity of a given set is simple to check if the set is already known to be a cone. The next theorem provides a necessary and a sufficient condition for a cone to be convex.

Theorem 2.2.1: A cone $K \subseteq \mathbf{R}^n$ is convex if and only if

$$x, y \in K \text{ implies } x + y \in K.$$

Poof: Let K be convex and x, y $\in$K. Then

$$\frac{x+y}{2} \in K.$$

That is

$$x + y \in 2K \subseteq K.$$

Conversely, let x, y $\in$ K. Then αx, βy $\in$ K for all α, β $\in \mathbf{R}_+ \backslash \{0\}$.

Consequently,

$$\alpha x + (1 - \alpha)y \in K \text{ for all } \alpha \in (0, 1).$$

Convex cones being among the simplest convex sets play paramount role. They are basic in convex analysis as subspaces do in linear analysis. A convex cone can be characterized as an intersetion of a subspace with a linear hyperplane.

Definition2.2.2: Let $x_1, x_2, ..., x_k \in \mathbf{R}^n$, $\lambda_1, \lambda_2, ..., \lambda_k \in \mathbf{R}_+$. Then

$$z := \sum_{i=1}^{k} \lambda_i x_i$$

is said to be the conical combination of the points $x_1, x_2, ..., x_k$.

The set coneS of all conical combinations from a given non-empty set S $\subseteq \mathbf{R}^n$ is called the **conical hull** of S.

If $S = \{x_1, x_2, ..., x_k\} \subseteq \mathbf{R}^n$, then

$$\text{coneS} = \{\sum_{i=1}^{k} \alpha_i x_i \ , \alpha_i \in R_+ \ , i = 1,2,3,..., k\}.$$

Every conical hull trivially contains the zero element. In order to form coneS one can interesct all convex cones containing S and append 0 to the result.

Definition 2.2.3: Let K be a convex cone. Then the polar cone K^0 of K (called negative polar cone by some authors) is defined by

$$K^0: = \{ y \in \mathbf{R}^n \mid \langle y, x \rangle \le 0 \text{ for all } x \in K\}.$$

For y$\in K^0$, we have that

$\langle y, x \rangle = \|y\|\|x\|\cos\theta \le 0$ for all x $\in$ K. This implies that the angle between x and y is obtuse for each x $\in$ K.

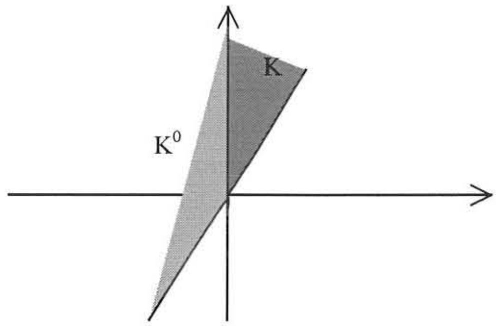


Fig.2.2.2, A cone and its ploar cone

To illustrate the notion of a polar cone, consider

$$K: =\{ \sum_{i=1}^{k} \alpha_i x_i \ ,\alpha_i \in R_+ \ ,i = 1,2,3,..., k\}, \text{ the conical hull of k points } x_1, x_2, ..., x_k \text{ in } \mathbf{R}^n,$$

then

$$K^0 = \{y \in \mathbf{R}^n \mid \langle y, \ x_i \rangle \ \leq \ 0 \ \text{ for all } i = 1, 2, ..., k\}.$$

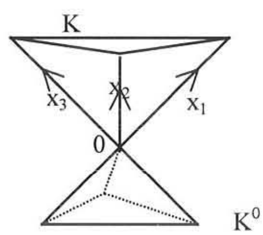


Fig.2.2.3, Example of a polar cone

By definition the polar cone depends on the scalar product $\langle ., \ .\rangle$. Change of $\langle ., \ .\rangle$ changes K^0 even K remaining the same. As it will be seen later on the polar of K^0 is the closure of K. Evidently, if K is closed, then the polar of K^0 becomes K itself.

Theorem 2.2.2: Let K, K_1, K_2, $K_j \subseteq \mathbf{R}^n$, $j \in J$ be convex cones where J is an index set. Then

a) K^0 is a closed convex cone.

b) $K_1 \subseteq K_2$ implies $K_2^0 \subseteq K_1^0$.

c) $K \cap K_0 = \{0\}$.

d) $K^0 = K^\perp$ if K is a subspace of $\mathbf{R}^n$.

e) $(\alpha K)^0 = -K^0$ for $\alpha \in \mathbf{R} \backslash \mathbf{R}_+$.

f) $(\bigcup_{j \in J} K_j)^0 = \bigcap_{j \in J} K_j^0$.

Proof: a) Let (y_n) be a sequnce in K^0 such that $y_n \xrightarrow{n\to\infty} y \in \mathbf{R}^n$. Then

$$\langle y_n, x\rangle \leq 0 \text{ for all } x \in K.$$

Due to the continuity of the scalar prodauct we get

$$0 \geq \lim_{n\to\infty}\langle y_n, x\rangle = \langle \lim_{n\to\infty}, x\rangle = \langle y, x\rangle \text{ for all } x \in K.$$

This implies

$$y \in K^0.$$

b) Let $y \in K_2^0$. Then $\langle y, x\rangle \leq 0$ for all $x \in K_2$. But as $K_1 \subseteq K_2$, $\langle y\ , x\rangle \leq 0$ for all $x \in K_1$.

Thus we have that

$$y \in K_1^0.$$

Polarity thus establishes a correspondence in the set of closedconvex cones which is order-reverting.

c) Let $y \in K \cap K_0$. Then $\langle y\ , x\rangle \leq 0$ for all $x \in K$. In particular for $x = y$, we get

$$\langle y\ , y\rangle = 0 \text{ .i.e. } y = 0.$$

Hence 0 is the only possible element in $K \cap K_0$.

d) Clearly, $K^{\perp} \subseteq K^0$. Let $y \in K^0$, $y \neq 0$. Then

$$\langle y, x\rangle \leq 0 \text{ for all } x \in K. \tag{2.14}$$

But since K is a subspace, $x \in K$ implies $\lambda x \in K$ for all $\lambda \in \mathbf{R}$.

Thus (2.14) becomes

$$\lambda \langle y, x\rangle \leq 0 \text{ for all } \lambda \in \mathbf{R}. \tag{2.15}$$

If $\langle y\ , x\rangle \neq 0$, then by approperiate choice of $\lambda \in \mathbf{R}$, the left hand side of (2.15) can be made arbitrary large. Hence (2.15) holds true only if $\langle y, x\rangle = 0$.

That is

$$y \in K.$$

As a result, if K is a subspace, then K^0 becomes its orthogonal and thus polarity generalizes orthogonality.

e) Let $\alpha \in \mathbf{R}\backslash\mathbf{R}_+$. Then

$$(\alpha K)^0 = \{y \in \mathbf{R}^n \mid \langle y, \alpha x\rangle = \alpha\langle y, x\rangle \leq 0 \text{ for all } x \in K\},$$

$$= \{y \in \mathbf{R}^n \mid \langle y, x\rangle \geq 0 \text{ for all } x \in K\},$$

$$= \{y \in \mathbf{R}^n \mid -\langle y, x\rangle \le 0 \text{ for all } x \in K\},$$

$$= -K^0.$$

f) $$(\bigcup_{j\in J} K_j)^0 = \{y \in \mathbf{R}^n \mid \langle y, x\rangle \le 0 \text{ for all } x \in \bigcup_{j\in J} K_j\},$$

$$= \{y \in \mathbf{R}^n \mid \langle y, x\rangle \le 0 \text{ for all } x \in K_j \text{ and for each } j \in J\},$$

$$= \bigcap_{j\in J} K_j^0 . \qquad \square$$

Like in subspaces the Characterzation Theorem takes the next special form in conical situations.

Theorem 2.2.3: Let $K \subseteq \mathbf{R}^n$ be a closed convex cone. Then $y_x = P_K(x)$ if and only if

$$y_x \in K,\ x - y_x \in K^0 \text{ and } \langle x - y_x,\ y_x\rangle = 0. \tag{2.16}$$

Proof: Let $y_x = P_K(x)$. Then by the Characterization Theorem we have

$$\langle x - y_x,\ y - y_x\rangle \le 0 \qquad \text{for all } y \in K.$$

Put $y = \alpha\, y_x$ for arbitrary $\alpha \in \mathbf{R}, \alpha > 0$.

Correspondigly we get

$$\langle x - y_x, \alpha y_x - y_x\rangle = (\alpha - 1)\langle x - y_x,\ y_x\rangle \le 0 \qquad \text{for all } \alpha > 0.$$

But $\alpha - 1$ can assume either of the signs which results

$$\langle x - y_x,\ y_x\rangle = 0.$$

Moreover, $\langle x - y_x,\ y - y_x\rangle = \langle x - y_x,\ y\rangle + \langle x - y_x,\ -y_x\rangle \le 0,$

or $\quad \langle x - y_x,\ y - y_x\rangle = \langle x - y_x,\ y\rangle \le 0 \quad$ for all $y \in K$.

Thus

$$x - y_x \in K^0. \tag{2.17}$$

Conversely, assume y_x satisfies (2.16) and for $y \in K$, consider

$$f(y) := \frac{1}{2}\|x - y\|^2 = \frac{1}{2}\|x - y_x + y_x - y\|^2 \ge f(y_x) + \langle x - y_x,\ y_x - y\rangle \quad \text{for each } y \in K.$$

But (2.16) and (2.17) implies

$$\langle x - y_x,\ y_x - y\rangle = -\langle x - y_x,\ y\rangle \ge 0 \text{ for each } y \in K.$$

Hence

$$f(y) \ge f(y_x) + \langle x - y_x,\ y_x - y\rangle \ge f(y_x) \text{ for each } y \in K.$$

This implies

$$y_x = P_K(x). \qquad \square$$

Example: Let x = (1, 0) and $K := \{ \alpha_1 x_1 + \alpha_2 x_2 \mid \alpha_i \in R_+, i = 1, 2\}$ where

$$x_1 = (1, 2) \text{ and } x_2 = (-2, 1).$$

Let $y_x = \alpha_1 x_1 + \alpha_2 x_2 = (\alpha_1 - 2\alpha_2, 2\alpha_1 + \alpha_2)$, $\alpha_i \geq 0$, i = 1, 2 be the projection $P_K(x)$.

From $x - y_x = (1 - \alpha_1 + 2\alpha_2, 2\alpha_1 + \alpha_2) \in K^0 = \{y \in \mathbf{R}^2 \mid \langle y, x_i \rangle \leq 0, i = 1, 2\}$, we have

$$5\alpha_1 \geq 1 \text{ and } 4\alpha_1 + 5\alpha_2 \geq -2, \alpha_i \geq 0, i = 1,2 \tag{2.18}$$

Furthermore, from $\langle x - y_x, y_x \rangle = 0$ we get

$$(1 - \alpha_1 + 2\alpha_2)(\alpha_1 - 2\alpha_2) - (2\alpha_1 + \alpha_2)^2 = 0. \tag{2.19}$$

Solving the system of equation (2.18) and inequalities (2.17) togther results

$$\alpha_1 = \frac{1}{5} \text{ and } \alpha_2 = 0.$$

Therefore

$$y_x = \alpha_1 x_1 = (\tfrac{1}{5}, \tfrac{2}{5}) = P_K(x).$$

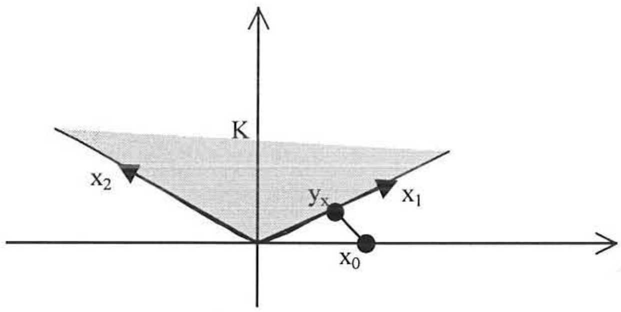


Fig.2.2.4

The following properties are some immediate consequences of Theorem 2.2.3.

For all $x \in \mathbf{R}^n$,

(i) $P_K(x) = 0$ if and only if $x \in K^0$.

(ii) $P_K(\alpha x) = \alpha P_K(x)$ for all $\alpha \in \mathbf{R}$, $\alpha \geq 0$.

(iii) $P_K(-x) = - P_{-K}(x)$.

(iv) $P_K(x) + P_{K0}(x) = x$.

The first three properties some how generalize the linearity of the projection onto a subspace C while the last one plays the role of the representation $x = P_C(x) + P_{C^\perp}(x)$.

Theorem 2.2.4: (J.J Moreau)

Let K be a closed convex cone. For $x, x_1, x_2 \in \mathbf{R}^n$, the properties below are equivalent.

(a) $x = x_1 + x_2$ with $x_1 \in K$, $x_2 \in K^0$ and $\langle x_1, x_2 \rangle = 0$.

(b) $x_1 = P_K(x)$ and $x_2 = P_{K0}(x)$.

Proof: (a) $\Rightarrow$ (b):

Let $x = x_1 + x_2$, $x_1 \in K$, $x_2 \in K^0$ and $\langle x_1, x_2 \rangle = 0$.

$x = x_1 + x_2$ implies $x - x_1 = x_2 \in K^0$ and $x_1 \in K$. Then consider

$$\langle x - x_{1,}\ x_1 \rangle = \langle x_1 + x_2 - x_1,\ x_1 \rangle = \langle x_2,\ x_1 \rangle = 0 .$$

Hence we have that

$$x_1 = P_K(x).$$

Seemingly,

$x = x_1 + x_2$ implies $x - x_2 = x_1 \in K^{00} = K$ and $x_2 \in K^0$.

Then consider

$$\langle x - x_2,\ x_2 \rangle = \langle x_1 + x_2 - x_2,\ x_2 \rangle = \langle x_1,\ x_2 \rangle = 0 .$$

Hence

$$x_2 = P_{K0}(x).$$

(b) $\Rightarrow$ (a):

Let $x_1 = P_K(x)$ and $x_2 = P_{K0}(x)$. Then by property (iv) above $P_K(x) + P_{K0}(x) = x = x_1 + x_2$. On the other hand P_K and P_{K0} are projections **onto** K and K^0 respectively which assures that

$$x_1 \in K \text{ and } x_2 \in K^0.$$

Finally, applying Theorem 2.2.3 we get

$$\langle x - x_{1,}\ x_1 \rangle = 0 \text{ and}$$

$$\langle x - x_2,\ x_2 \rangle = 0.$$

Adding these two equations gives

$$0 = \langle x - x_1, x_1\rangle + \langle x - x_2, x_2\rangle = 2\langle x_1, x_2\rangle .$$

Therefore

$$\langle x_1, x_2\rangle = 0.$$

Unlike the decomposition in subspaces, the representation $x = x_1 + x_2$, $x_1 \in K$, $x_2 \in K^0$ here is not unique.

3. Separation and Its Applications

3.1 Separation of Convex Sets

Among the far-reaching properties of convex sets,the possibility to separate two convex sets by a hyperplane plays vital role not only in optimization but also in different fields of mathematics. The duality theory in optimization is to be worth mentioning whre the application of separation takes a central place. In the following a brief introduction to separation with some of its applications are considered.

Definition 3.1.1: Let S_1 and S_2 be two non empty convex sets in $\mathbf{R}^n$. Then S_1 and S_2 are said to be :

(a) separable if and only if there is $a \in \mathbf{R}^n$, $a \neq 0$ and $\alpha \in \mathbf{R}$ such that

$$\langle a,x\rangle \leq \alpha \leq \langle a,y\rangle \text{ for all } x \in S_1 \text{ and } y \in S_2. \tag{3.1}$$

(b) strictly separable if and only if there is $a \in \mathbf{R}^n$, $a \neq 0$ such that

$$\sup\{\langle a,x\rangle \mid x \in S_1\} < \inf\{\langle a,y\rangle \mid y \in S_2\}. \tag{3.2}$$

Separation of convex sets has the following simple geometrical interpretation. From $\langle a,x\rangle \leq \alpha \leq \langle a,y\rangle$, it follows that

$$\langle a,x\rangle - \alpha \leq 0 \leq \langle a,y\rangle - \alpha \text{ for all } x \in S_1 \text{ and } y \in S_2.$$

Thus

$S_1 \subseteq H^- := \{x \in R^n \mid \langle a, x\rangle - \alpha \leq 0\}$ and

$S_2 \subseteq H^+ := \{x \in R^n \mid \langle a, x\rangle - \alpha \geq 0\}$. That means S_1 and S_2 are subsets of the two opposite halfspaces H^- and H^+ determined by the hyperplane

$$H := \{x \in \mathbf{R}^n \mid \langle a,x\rangle = \alpha\}.$$

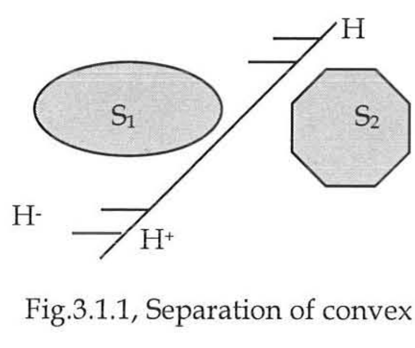


Fig.3.1.1, Separation of convex sets

Therefore if S_1 and S_2 are two disjoint convex sets, then a convex set namely an affine hyperplane can be squeezed in between. This extremly important property follows directly from that of the projection operator onto a closed convex set. Neverthless, when either S_1 or S_2 fails to be convex a separating hyperplane in between may not exist. Separation theorems are formulated in most cases with regard to convex sets.

Theorem 3.1.1: (The Separation Theorem)
Let $C \subseteq \mathbf{R}^n$ be a non-empty, closed and convex set with $x \in \mathbf{R}^n \backslash C$. Then there exists a point $s \in \mathbf{R}^n$ such that

$$\langle s, x\rangle \quad > \sup\{\langle s, y\rangle \mid y \in C\}. \tag{3.3}$$

Proof: Applying the projection operator (see chapter 2.1) simplifies the proof.

Put $s := x - P_C(x)$. Since $x \notin C$, $x \neq P_C(x)$ implies $s \neq 0$. Then by the Characterization Theorem we get that

$$0 \geq \langle x - P_C(x), y - P_C(x)\rangle = \langle s, y - x + s\rangle = \langle s, y\rangle - \langle s, x\rangle + \|s\|^2 \text{ for each } y \in C.$$

or

$$\langle s, x\rangle \geq \|s\|^2 + \langle s, y\rangle \text{ for all } y \in C.$$

Taking the suprema gives

$$\langle s, x\rangle \geq \|s\|^2 + \sup\{\langle s, y\rangle \mid y \in C\}.$$

As $\|s\|^2 > 0$, we have that

$$\langle s, x\rangle > \sup\{\langle s, y\rangle \mid y \in C\}. \qquad \square$$

The Separation Theorem has an obvious geometrical meaning. Because $s \neq 0$, it defines hyperplanes $H_{s,r}$ which are translations of the same linear or vector hyperplane $H_{s,0}$ when r describes $\mathbf{R}$. In particular setting

$$r := \frac{1}{2}\left(\langle s, x\rangle + \sup\langle y, s\rangle, y \in C\right) \text{ yields },$$

$$\langle s, x\rangle + \sup\langle y, s\rangle - 2r = 0, \quad y \in C. \tag{3.4}$$

From (3.3), $\langle s,x\rangle = k + \sup\langle y,s\rangle$, y $\in$C for some k $\in\mathbf{R}$, k > 0.

Consequently (3.4) becomes

$$k + 2\sup\langle y,s\rangle = 2r,\ y \in C.$$

This again implies

$$k + 2\langle y,s\rangle \le 2r \text{ for each } y \in C.$$

As k>0 we thus have

$$\langle y,s\rangle - r < 0 \text{ for all } y \in C. \tag{3.5}$$

On the other hand $\sup\langle y,s\rangle = \langle s,x\rangle - k,\ y \in C.$

With this representation (3.4) again becomes

$$2\langle s,x\rangle - k = 2r.$$

It follows that

$$\langle s,x\rangle - r > 0. \tag{3.6}$$

Combing (3.5) and (3.6) reflects that there is a hyperplane

$$H_{s,r} := \{z \in \mathbf{R}^n \mid \langle z,s\rangle = r\}$$

separating the two convex sets C and {x}where $C \subseteq H^-_{s,r}$ and $\{x\} \subset H^+_{s,r}$.

Example: Let $C = \{y \in \mathbf{R}^n \mid y = \alpha x_0,\ \alpha \in \mathbf{R}\}$ where

$x_0 = (1, 2, 3, 0, 0, 0, \ldots, 0)$ and $x = (0,1, 2, 0, 0, 0, \ldots,0)$.

Obviously, $x \notin C$ and C is a non-empty, closed and convex subset of $\mathbf{R}^n$. Hence by the Separation Theorem there exists a point $s \in \mathbf{R}^n$ such that

$$\langle s,x\rangle > \sup\{\langle s,y\rangle \mid y \in C\}.$$

If we set $P_C(x) = y_x = (\beta, 2\beta, 3\beta, 0, 0, \ldots, 0)$ for some $\beta \in \mathbf{R}$, then

$$\langle x - y_x, y\rangle = 0 \text{ for all } y \in C.$$

This implies

$$-\alpha\beta + 2\alpha(1-2\beta) + 3\alpha(2-3\beta) = 0 \text{ for all } \alpha \in \mathbf{R}.$$

or

$$\beta = \frac{4}{7} \text{ and } y_x = \frac{4}{7}x_0.$$

Now, x - y_x = $(-\frac{4}{7}, -\frac{1}{7}, \frac{2}{7}, 0, 0, \ldots, 0)$ and $\langle x - y_x, x\rangle = \frac{-1}{7} + \frac{4}{7} = \frac{3}{7}$.

Setting s: = x - y_x yields

$$\langle s, x\rangle = \frac{3}{7} \text{ and } \langle s, y\rangle = \frac{-4}{7}\alpha + \frac{-2}{7}\alpha + \frac{6}{7}\alpha = 0 \text{ for all } y \in C.$$

so

$$\langle s, x\rangle = \frac{3}{7} > \sup\{\ \langle s, y\rangle \mid y \in C\} = 0.$$

Futhermore, the hyperplane $H_{s,r} = \{z \in \mathbf{R}^n \mid \langle s, z\rangle = r\}$ where $r = \frac{3}{14}$ separates

{x} and set C.

The Separation Theorem can be extended to the forthcoming corollary.

Corollary 3.1.2: (Strict Separation of Convex Sets)

Let $C_i \subseteq \mathbf{R}^n$, i = 1, 2 be two disjoint non-empty, closed and convex sets. If C_2 is bounded, then there exists $s \in \mathbf{R}^n$ such that

$$\sup_{y_1 \in C_1} \langle s, y_1\rangle < \min_{y_2 \in C_2} \langle s, y_2\rangle. \tag{3.7}$$

Proof: Clearly Since C_1 and C_2 are convex sets , C_1 - C_2 is also a convex set (see chapter 1.2 (ii)). Besides because C_2 is a closed and bounded set in the finite dimensional space $\mathbf{R}^n$ and hence compact, C_1 - C_2 again becomes a closed set. To say that C_1 and C_2 are disjoint is exactly to mean that $0 \notin C_1 - C_2$. Thus the Separation Theorem guarantees an $s \in \mathbf{R}^n$ such that

$$0 = \langle s, 0\rangle > \sup\ \{\ \langle s, y\rangle \mid y \in C_1 - C_2\}.$$

This means

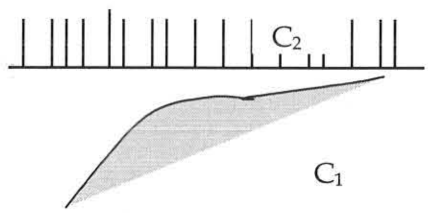
$$0 > \sup\ \{\langle s, y_1\rangle \mid y_1 \in C_1\} + \sup\{\langle s, -y_2\rangle \mid y_2 \in C_2\},$$

$$= \sup\ \{\langle s, y_1\rangle \mid y_1 \in C_1\} - \inf\{\langle s, y_2\rangle \mid y_2 \in C_2\}, \tag{3.8}$$

Now, C_2 being bounded, the infimum in (3.8) is a minimum and of course finite which can be moved to the left hand side.i.e.

$$\min\{\langle s, y_2\rangle \mid y_2 \in C_2\} > \sup\{\langle s, y_1\rangle \mid y_1 \in C_2\}. \quad \square$$

For strict separation compactness of either of the sets deserve a prominent role. When both C_1 and C_2 are unbounded corollary 3.1.2 may fail as the figure below illustrates.



$$\sup_{y_1 \in C_1} \langle s, y_1 \rangle \leq \inf_{y_2 \in C_2} \langle s, y_2 \rangle$$

Fig.3.1.2

3.2 Applications of Separations

The separation properties introduced heretofore have various applications of which the present section is claimed to study some of them.

(I) Existence of Supporting Hyperplanes

Definition 3.2.1:The set $H_{s,r}$: $= \{x \in \mathbf{R}^n \,|\, \langle x, s \rangle = r\}$ is said to be a supporting hyperplane of set C

(a) if $\langle s, y \rangle \leq r$ for all y $\in$C, i.e. C $\subseteq H^-_{s,r}$. (3.9)

(b) at the point x $\in$ C if C $\subseteq H^-_{s,r}$ and x $\in$ $H_{s,r}$. (3.10)

For illustration we consider the following

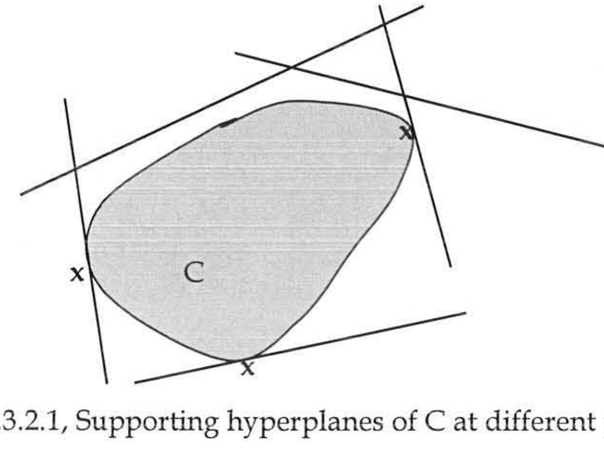


Fig.3.2.1, Supporting hyperplanes of C at different points

Notice that a supporting hyperplane of C may or may not touch set C. If $H_{s,r}$ supports set C, then $H_{-s,-r}$ also supports it. Naturally the inequality sign in (3.9) can be reversed. Consequently $H_{s,r}$ supports set C if C is entirely contained in one of the two closed halfspaces $H^+_{s,r}$ or $H^-_{s,r}$. Accordingly, when C has a supporting hyperplane at $x \in C$, then x is a boundary point of C. The whole space is an instance of a set that possesses no supporting hyperplane.

We recall from toplogy that a subspace A of $\mathbf{R}^n$ can be equipped with a toplogy relative to A by defining the " open balls" $B(x, \delta) \cap A$ for $x \in A$. Then A becomes a toplogical space by its own. In convex analysis the toplogy of $\mathbf{R}^n$ is of modert interest. However, the toplogies relative to affine manifolds turn out to be much resourceful.

Definition 3.2.2: The relative interior riC of a convex set $C \subseteq \mathbf{R}^n$ is the interior of C for the toplogy relative to the affine hull of C.

By definition $x \in$ riC if and only if $x \in$ affC and there exists $\delta > 0$ such that

$$\text{aff}C \cap B(x, \delta) \subseteq C.$$

Obviously, riC $\subseteq$ C and the cluster points of set C are in the closed set affC. As a result the relative closure of C is exactly clC. To the contrary the the boundary is affected and indeed the relative boundary of C symbolically becomes

$$\text{rbd}C := \text{cl}\, C \setminus \text{ri}C.$$

Theorem 3.2.1: Let $C \subseteq \mathbf{R}^n$, $C \neq \mathbf{R}^n$ be a convex set. Then C has a supporting hyperplane.

Proof: Since $C \neq \mathbf{R}^n$, clC $\neq \mathbf{R}^n$. Indeed otherwise we have that

$$C \supseteq \text{ri}C = \text{ricl}C = \text{ri}\,\mathbf{R}^n = \mathbf{R}^n.$$ But this is a contradiction.

Now, let $x \in \mathbf{R}^n \setminus$cl C. Then by the Separation Theorem there is a hyperplane $H_{s,r}$ which separates clC and {x}where

$$\text{cl}\, C \subseteq H^-_{s,r} \text{ and } \{x\} \subseteq {}^+_{s,r} H.$$

Evidently this hyperplane is our asserted support of set C. □

From the above theorem we have that every proper convex subset C of $\mathbf{R}^n$ deos have a supporting hyperplane. A slightly more and useful formulation is still the following lemma.

Lemma 3.2.2: Let $C \subseteq \mathbf{R}^n$, $C \neq \mathbf{R}^n$ be a non-empty convex set. Then there exists a supporting hyperplane at each boundary point x of C.

Proof: Let x be a boundary point of C. Since C, clC and their complements have the same boundary, a sequence (x_k) can be found where

$$(x_k) \in \mathbf{R}^n \backslash \mathrm{cl}\, C \text{ and } x_k \xrightarrow{k \to \infty} x, \ k \in N.$$

As cl C is a closed convex set, for each $k \in N$, by the Separation Theorem we have some $s_k \in \mathbf{R}^n$ such that

$$\langle s_k, x_k \rangle > \sup\{\langle s_k, y\rangle \mid y \in cl\, C\}.$$

This implies

$$\langle s_k, x_k \rangle - \langle s_k, y\rangle = \langle s_k, x_k - y\rangle > 0 \text{ for all } y \in \mathrm{cl}C.$$

Perhaps we have a sequence (H_{s_k, r_k}) of hyperplanes where H_{s_k, r_k} strictly separates C and $\{x_k\}$with

$$C \subseteq H^-_{s_k, r_k} \text{ and } \{x_k\} \subseteq H^+_{s_k, r_k}.$$

WLoG, we can assume $\|s_k\| = 1$. Then the distance of H_{s_k, r_k} to the origin 0 is given by

$$d_K = \left| \frac{\langle s_k, 0\rangle - r_k}{\|s_k\|} \right| = |-r_k| = |r_k|.$$

Obviusly, $|r_k| \leq \max\{\|x_1\|, \|x_2\|, \ldots, \|x\|\} =: m < \infty.$

Thus m has to be a real number since the sequence is convergent and therefore bounded. Then we consider the sequence (s_K, r_K) which lies in the set

$$M := \{(a, \alpha\} \mid \|a\| = 1, |\alpha| \leq m\}.$$

Clearly, M is compact and hence each sequence in M has a convergent subsequence.

This implies that

$$(s_{k_j}, r_{k_j}) \xrightarrow{j\to\infty} (s, r_0) \in M \tag{3.11}$$

Again (3.11) inturn implies

$$s_{k_j} \xrightarrow{j\to\infty} s.$$

Finally by the continuity of the scalar product we have

$$\lim_{j\to\infty} \langle s_{k_j}, x_{k_j} \rangle = \langle s, x - y\rangle \geq 0 \qquad \text{for each } y \in C.$$

Put $r := \langle s, x\rangle$. Then

$$r \geq \langle s, y\rangle \qquad \text{for all } y \in C.$$

Hence the set

$$H_{s,r} := \{z \in R^n \mid \langle z, y\rangle = r = \langle s, y\rangle\}$$

is the desired hyperplane that supports C at x. □

(II) Outer Description of Closed Convex Set

In convex analysis intersecting convex sets is an operation of utmost importance. Closing a convex set ended up in intersecting the closed convex sets containing it. We recall (chapter1.2) that convexity allowed the intersection to be restricted to a simple class of closed convex sets namely the closed half spaces. Fortunate enough, Theorem 3.2.1 ensures that a non-empty proper subset C of $\mathbf{R}^n$ has at least one supporting hyperplane. So if we let an index set:

$$\sum C := (s, r) \in R^n \times R \mid C \subseteq H^-_{s,r} \}$$

where

$$H^-_{s,r} := \{y \in R^n \mid \langle s, y\rangle \leq r\}, \text{ then } \sum C \text{ is non -empty.}$$

As illustrated below we can therefore intersect all half spaces indexed in $\sum C$.

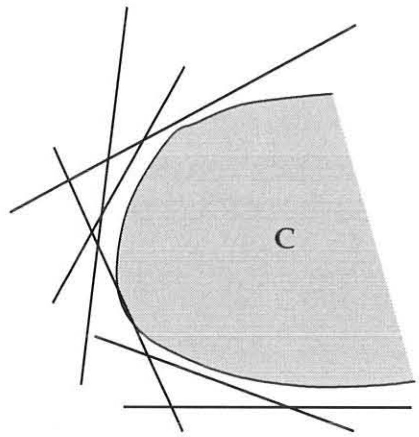


Fig.3.2.2, Outer description of a closed convex set.

Put $C^* := \bigcap_{(s,r)\in \sum C} H_{s,r}^- = \{ z \in R^n \mid \langle s, y\rangle \le r$ whenever $\langle s, y\rangle \le r, \forall y \in C \}$.

Actually $\sum C$ is redundant since it contains too many r's. The definition of C^* could rather be slightly simplified. For s $\mathbf{R}^n$, consider the number

$$r_s := \inf \{ r \in \mathbf{R} \mid (s, r) \in \sum C \}.$$

Letting s vary (s, r_s) describes $\sum C$. With this new description C^* becomes

$$C^* = \{ z \in R^n \mid \langle s, z\rangle \le \sup_{y\in C} \langle s, y\rangle \}.$$

Obviously, $C \subseteq C^*$ and C^* is a closed convex set.

Theorem 3.2.3: Let $C \subseteq \mathbf{R}^n$, $C \neq \mathbf{R}^n$ be a non-empty convex set. Then the set C^* defined above is the closure of C.

Proof: We know that the closure of C, clC, is the intersection of all closed sets containing C. Thus by construction

$$C^* \supseteq \mathrm{cl}C.$$

To verify the other inclusion one can equivalently prove the contra positive. That is if x does not belong to the closure of C, then it is not an element of C^*. To do so assume that

$$x \notin \mathrm{cl}C.$$

Then by Separation Theorem there exists $s_0 \in \mathbf{R}^n$, $s_0 \neq 0$, such that

$$\langle s_0 , y\rangle > \sup_{y\in C} \langle s_0 , y\rangle =: r_0 .$$

It follows that

$$\langle s_0, y\rangle \le r_0 \text{ for all } y \in C.$$

So we get

$$C \subseteq H^-_{s_0,r_0} \text{ and hence } (s_0, r_0) \in \sum C .$$

But $x \notin H^-_{s_0,r_0}$ which implies

$$x \notin C^* .$$

To this effect

$$C^* \subseteq \mathrm{cl}C. \qquad \square$$

A fruitful message of Theorem 3.2.3. is the forthcoming corollary whose proof trivially follows.

Corollary 3.2.4: Let $C \subseteq \mathbf{R}^n$, $C \neq \mathbf{R}^n$ be a non-empty, closed and convex set. Then C is the intersection of all closed half spaces containing it.

A closed convex set can therefore be defined as the intersection of all closed half spaces containing it. A useful special case is indeed the next definition.

Definition 3.2.4: A closed convex polyhedron is the intersection of finitely many half spaces.

Let $(s_1, r_1), (s_2, r_2), ...,(s_m, r_m) \in \mathbf{R}^n \times \mathbf{R}$ with $s_j \neq 0, j = 1, 2, ..., m$. Then

$$P := \{ x \in \mathbf{R}^n \mid \langle s_j, x \rangle \le r_j \text{ for } j = 1, 2, ..., m \}$$

is a closed convex polyhedron. P can be also represented using matrix notation as

$$P = \{ x\in \mathbf{R}^n \mid Ax \le b \}$$

whereA is the matrix with rows s_j and $b \in \mathbf{R}^n$ with coordinates r_j.

In the case where b = 0 we get a closed convex polyhedral cone.

(III) Bipolar of a convex cone

Definition 3.2.5: Let $C \subseteq \mathbf{R}^n$ be a non-empty set. Then the function

$$\sigma_C : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\} \text{ defined by}$$

$$\sigma_C(x) := \sup \{\langle x, y\rangle \mid y \in C\} \tag{3.12}$$

is called the **support function** of C.

The motivation to formulate the aforementioned definition basically arises from the right hand side of (3.3).

Obviously, $\langle x, y\rangle \leq \sigma_C(x)$ for all $x \in \mathbf{R}^n$ and $y \in C$. This characterizes the elements of set C. For a given set C, the support function is strongly attached to the scalar product $\langle .,. \rangle$. Change of the scalar product changes σ_C even C remaining the same. The supremum in (3.12) could be finite or infinite that might be achieved on C or not. In this context, C can be possibly interpreted as an index set where $\sigma_C(.)$ is the supremum of the collection $\langle s, . \rangle$ over C. When C is bounded, then its support function is finite everywhere. Furthermore, it is simple to notice that σ_C is also the support function of the closure of C and its closed convex hull. It is therefore logical to consider only support function of a non-empty closed convex set. The support functions are sublinear and $\sigma_C(0) = 0$. However, $\sigma_C(x) = 0$ need not imply that $x = 0$.

The definition of a polar cone was given in section (2.2) where attempt has made to point out some interesting properties. Here we want to consider one more similarity with the notion of orthogonality in linear analysis.

Definition 3.2.6: Let $K \subseteq \mathbf{R}^n$ be a convex cone. Then the bipolar K^{00} of K is the polar of K^0.

The theorem set forth gives rise to the relation between K and its bipolar cone K^{00}.

Theorem 3.2.5: Let $K \subseteq \mathbf{R}^n$ be a convex cone. Then $K^{00} = \text{cl } C$.

Proof: Employing definition 3.2.5 makes the proof easier. Due to its conical character, clK has a very special support function.

$$\sigma_{clK}(s) = \sup\{\langle y, s\rangle \mid y \in clK\},$$
$$= \sup\{\alpha\langle y, s\rangle \mid y \in clK, \alpha \in R, \alpha > 0\},$$
$$= \begin{cases} 0, & \text{if } \langle s, y\rangle \leq 0 \text{ for all } y \in clK, \\ \infty, & \text{otherwise.} \end{cases}$$
$$= \begin{cases} 0, & \text{for all } s \in K^0, \\ \infty, & \text{otherwise.} \end{cases}$$

In other words σ_{clK} is 0 on K^0 and infinite elsewhere.

Evidently,

$$x \in \text{cl K if and only if } \langle s, x\rangle \leq \sigma_{clK}(x). \tag{3.13}$$

Thus the statement in (3.13) becomes

$x \in$ clK if and only if

$\langle s, x\rangle \leq 0$ for all $s \in K^0$ and arbitrary $\langle s, x\rangle$ for $s \notin K^0$.

This is clear enough to recall the definition of K^{00} and to notice the required conclusion K^{00} = clC. □

In the case where K is already closed we have trivially $K^{00} = K$. Let $H_{s,r}$ be a hyperplane supporting cone K at its boundary point x. Then

$$\langle k, s\rangle \leq r = \langle s, x\rangle \qquad \text{for all } k \in K.$$

or

$$\alpha\langle k, s\rangle \leq \langle s, x\rangle \text{ for all } k \in K \text{ and } \alpha \in \mathbf{R}, \alpha > 0. \tag{3.14}$$

Now (3.14) holds true only if

$$\langle k, s\rangle \leq 0 \text{ for all } k \in K \text{ and } \langle s, x\rangle = r \geq 0.$$

This assures that $H_{s,r}$ also supports cone K at 0. Therefore when dealing with support for a cone, it is quiet sufficient to deem only hyperplanes which are subspaces of $\mathbf{R}^n$.

(IV) The Lemma of Minkowski- Farkas

This section is intended to demonstrate another application of the separation property known as the Farkas' Lemma that deals with systems of linear inequalities. We recall that the real linear system $\sum_{i=1}^{m} a_{ij} y_j = b_j$, j = 1,2, ...,n

has a solution

$(y_1, y_2, ..., y_m)$ if and only if each solution $(x_1, x_2, ..., x_n)$ of the linear homogeneous system $\sum_{j=1}^{n} a_i x_j = 0$,i =1, 2, 3, .., m is a solution of the linear equation $\sum_{j=1}^{n} b_j x_j = 0$.

Definitely, if $\sum_{i=1}^{m} a_{ij} y_j = b_j$, j = 1, 2, ...,n. then

$$\sum_{j=1}^{n} b_j x_j = \sum_{j=1}^{n} (b_j x_j \sum_{i=1}^{m} a_{ij} y_i \quad) x_j = \sum_{i=1}^{m} y_i \left(\sum_{j=1}^{n} a_{ij} x_j \right)$$

Which follows that

$$\sum_{j=1}^{n} b_j x_j = 0 \text{ when } \sum_{j=1}^{n} a_i x_j = 0, \text{ i =1, 2, 3, ..., m.}$$

To sum up in a more handy expression, let A be an (n, m)- matrix and $b \in \mathbf{R}^n$. Then the system Ax = b is consistent (has at least one solution in $\mathbf{R}^m$) exactly when $b \in (\ker A^T)^{\perp}$. Representing the columns of A by $s_1, s_2, ..., s_m$ and applying the usual Euclidean notation the above equivalent properties can be formulated as:

$\langle b, x \rangle = 0$ whenever $\langle s_j, x \rangle = 0$, j = 1,2,.., m if and only if

$b \in$ linS where linS is the subspace generated by S: = $\{s_1, s_2, ..., s_m\}$.

The Farkas Lemma is actually nothing but an extension of the aforementioned classical result from linear algebra. Getting into the space of convex analysis where we replace linear hulls by conical hulls, we see that

$$\sum_{j=1}^{n} b_j x_j \geq 0 \text{ whenever } \sum_{i=1}^{m} a_{ij} x_j \geq 0 \ ,$$

for the scalars $y_j \geq 0$ for each $j = 1, 2, 3, ..., m$.

We shall first state the Farkas Lemma with out proof, as it will be an immediate consequence of Theorem 3.2.9

Lemma3.2.6: (Farkas' I)

Let $b, s_1, s_2, ..., s_m$ be given in $\mathbf{R}^n$. Put

$$A := \{ x \in \mathbf{R}^n \mid \langle s_j, x \rangle \leq 0, j = 1, 2, ..., m\}. \tag{3.15}$$

$$B := \{ x \in \mathbf{R}^n \mid \langle b, x \rangle \leq 0 \}. \tag{3.16}$$

Then

$$A \subseteq B \text{ if and only if } b \in \text{Cone}\{ s_1, s_2, ..., s_m \}.$$

Another equivalent way of saying $b \in \text{Cone}\{s_1, s_2, ..., s_m\}$ is that the system

$$b = \sum \alpha_j s_j, \alpha_j \geq 0 \; j = 1, 2, 3, ..., m \text{ has a solution}$$

$$\alpha = (\alpha_1, \alpha_2, ..., \alpha_m) \in \mathbf{R}^m. \tag{3.17}$$

Farkas Lemma is often formulated as a set of two statements such that each one is false when the other is true. In view of this we get the following.

Lemma 3.2.7: (Farkas' II)

Let $b, s_1, s_2, ..., s_m$ be given in $\mathbf{R}^n$. Then exactly one of the following statements is true

a) $b = \sum \alpha_j s_j, \alpha_j \geq 0 \; j = 1,2,3, ..., m$ has a solution $\alpha = (\alpha_1, \alpha_2, ..., \alpha_m) \in \mathbf{R}^m$.

b) The system $\langle b, x \rangle > 0$, $\langle s_j, x \rangle \leq 0$ for $j = 1, 2, ..., m$ has a solution $x \in \mathbf{R}^n$.

Still another approach is quiet geometric. If K stands to the convex cone generated by $\{s_1, s_2, ..., s_m\}$, then K^0 is exactly set A (see chapter 2.2). What Farkas Lemma says is thus $b \in K$ if and only if $\langle b, x \rangle \leq 0$ whenever $x \in K^0$. That means $b \in K^{00}$. Shortly Farkas Lemma is therefore $K = K^{00}$. However, from (III) we know that this holds only if K is closed. Accordingly the proof of Farkas Lemma reduces in proving the following result.

Lemma 3.2.8: (Farkas' III)

Let $s_1, s_2, ..., s_m$ be given in $\mathbf{R}^n$. Then the convex cone

$$K: = \text{cone } \{s_1, s_2, ..., s_m\} = \{ \sum_{j=1}^{m} \alpha_j s_j, \alpha_j \geq 0, \text{ for } j = 1, 2, ..., m\}$$

is closed.

Proof: First consider the case when the s_j's are linearly independent and hence m = n. Let (x_k) be a sequence in K that converges in $\mathbf{R}^n$. That is

$$(x_k) = \sum_{j=1}^{m} \alpha_j^k s_j \xrightarrow{k\to\infty} x = \sum_{j=1}^{m} \beta_j s_j \in R^n, \alpha_j^k \geq 0, j = 1, 2, ..., m. \tag{3.18}$$

But convergence in norm in a finite dimensional space is identical to the component wise convergence. Thus the convergence of (x_k) is equivalent to the convergence of (α_j^k) to some β_j. Furthermore, since each (α_j^k) in (3.18) is non-negative so does each β_j.

It follows that

$$x \in K.$$

On the other hand suppose the s_j's are not linearly independent. Then there is a system $\sum_{j=1}^{m} \beta_j s_j = 0$ having a non-zero solution $\beta = (\beta_1, \beta_2, ..., \beta_m) \in \mathbf{R}^m$.

Assume $\beta_j < 0$ for some j (change β to $-\beta$ if need be). For $x \in K$, $x = \sum_{j=1}^{m} \alpha_j s_j$,

Set $\alpha_j' := \alpha_j + t^*(x)\beta_j$ where

$$t^*(x) = -\frac{\alpha_j(x)}{\beta_j(x)}, \ i(x) \in \{ j \mid \min_{\beta_j<0} -\frac{\alpha_j}{\beta_j} \text{ for } j = 1, 2, ..., m \}.$$

Then each α_j' is non-negative and

$$x = \sum_{j=1}^{m} \alpha_j s_j = \sum_{j=1}^{m} [\alpha_j + t^*(x)\beta_j] s_j = \sum_{j \neq i(x)} \alpha_j' s_j.$$

Letting x vary in K we have thus constructed a decomposition $K = \bigcup_{i=1}^{m} K_i$ where K_i is the conical hull of at most the m -1 generators s_j. To this end if there is some i such that the generators of K_i are linearly dependent we repeat the argument for a further decomposition of this K_i. After finitely many such operations we end up with a decomposition of K as a finite union of polyhedron convex cones each having linearly independent generators. By the first part of the proof all these cones are closed and hence K is closed. □

We are now in the foreground to state the general version of Farkas Lemma with non-homogeneous terms and infinitely many inequalities. The proof directly uses that of the Separation Theorem.

Theorem 3.2.9: (The Generalized Farkas')

Let J be an index set and $(s_j, \rho_j) \in R^n \times R$, $j \in J$. Assume that the system of inequalities

$$\langle s_j, x \rangle \leq \rho_j, \ j \in J \tag{3.19}$$

has a solution $x \in \mathbf{R}^n$. Then the following two propositions are equivalent.

(a) $\langle b, x \rangle \leq r$ for all x satisfying (3.19).

(b) $(b, r) \in \mathrm{cl}K$ where K is the convex conical hull of $S := \{(0,1)\} \cup \{(s_j, \rho_j)\}_{i \in J}$.

Proof: (b) $\Rightarrow$ (a):

Let $(b, r) \in \mathrm{cl}\ K$. Then there exist a finite set $\{1, 2, \ldots, m\} \subseteq J$ non-negative scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$ such that

$$b = \lim_{k\to\infty} \sum_{j=1}^{m} \alpha_j^k s_j \ \text{ and } r = \lim_{k\to\infty} (\alpha_0^k + \sum_{j=1}^{m} \alpha_j^k \rho_j).$$

Making use of continuity of the scalar product and for each x satisfying (3.19) we have

$$\langle b, x \rangle = \langle \lim_{k\to\infty} \sum_{j=1}^{m} \alpha_j^k s_j, x \rangle = \lim_{k\to\infty} \sum_{j=1}^{m} \alpha_j^k \langle s_j, x \rangle,$$

$$\leq \lim_{k\to\infty} \sum_{j=1}^{m} \alpha_j^k \sigma_j = \lim_{k\to\infty} (r - \alpha_0^k) \leq r.$$

This establishes the conclusion (a).

(a) $\Rightarrow$ (b):

We shall prove the contrapositive.

First of all $\mathbf{R}^n \times \mathbf{R}$ is a pre-Hilbert space equipped with the scalar product

$$\langle (y, r_1), (z, r_2) \rangle = \langle y, z \rangle + r_1 r_2 .$$

Suppose (b, r) $\notin$ clK. Then by the Separation Theorem clK and (b, r) can be separated and hence there exists

$$(d, -t) \in \mathbf{R}^n \times \mathbf{R} \text{ such that } \langle (b, r), (d, -t) \rangle = \langle b, d \rangle - rt > \sup_{(s,\rho)\in K} \langle s, d \rangle - \rho t . \qquad (3.20)$$

It follows that the right hand supremum is a finite number q. Then by the conical character of K we have that

$$\alpha q \leq q \text{ for all } \alpha > 0 \text{ and hence } q \leq 0.$$

Actually q = 0 since (0, 0) $\in$K.

To sum up we have signed out an element (d, t) $\in \mathbf{R}^n \times \mathbf{R}$ such that

(i) $t \geq 0$ as (0, 1) $\in$K.

(ii) $\langle s_j, d \rangle - \rho t \leq 0$ for all j $\in$ J as (s_j, ρ_j) $\in$K.

(iii) $\langle b, d \rangle - rt > 0$ due to (3.20) and definition of a cone.

To this end we consider two different cases.

Case1: If t > 0, then dividing (ii) and (iii) by t results

$$\langle s_j, \frac{d}{t} \rangle \leq \rho_j \quad \text{for all j} \in \text{J and } \langle b, \frac{d}{t} \rangle > r.$$

Then setting the point x $= \frac{d}{t}$ automatically violates (a).

Case2: If t = 0, then take any x_0 satisfying (3.19) (such x_0 exists since the system is consistent) and consider

$$x(\alpha) = x_0 + \alpha d \qquad \text{for all} \quad \alpha > 0.$$

Then using (ii) we have

$$\langle s_j, x(\alpha) \rangle = \langle s_j, x_0 + \alpha d \rangle = \langle s_j, x_0 \rangle + \alpha \langle s_j, d \rangle \leq \rho_j \quad \text{for all j} \in \text{J}.$$

This implies that x(α) also satisfies (3.19) for all $\alpha > 0$.

But letting $\alpha \to \infty$ *in* $\langle b, x(\alpha)\rangle = \langle b, x_0 \rangle + \alpha \langle b, d\rangle$ and (iii) realizes that x(α) again violates (a). We thus have proved in both cases that not (b) implies not (a) which is trivially equivalent to (a) $\Rightarrow$ (b). □

We end up this chapter with two points relating Theorem 3.2.9 with the previous forms of Farkas' Lemma. At the first glance take the homogeneous case where both r and ρ_j 's are all zero. Then the system is automatically consistent by the trivial solution x = 0 and then the Theorem says

(a') $\langle s_j, x \rangle \le 0$ implies $\langle b, x \rangle \le 0$ is

equivalent to

(b') b$\in$ cl K where K is the cone generated by { s_j | j $\in$ J }. This is indeed Lemma 3.2.6.

Secondly, suppose that J = {1, 2, ..., m } be a finite set. Consequently the set $\{ x \in \mathbf{R}^n \mid \langle s_j, x\rangle \le \rho_j, \ \ \forall j \in J \}$ becomes a closed convex polyhedron provided that it is non-empty. Possibly by a more handy matrix notation we have that $A^T x \le \rho$ where A is the matrix with columns s_j's and $\rho \in \mathbf{R}^n$ with coordinates $\rho_1, \rho_2, \ldots, \rho_m$. Then Theorem 3.2.9 again says:

(a") $\{x \in \mathbf{R}^n \mid A^T x \le \rho \} \subseteq \{x \in \mathbf{R}^n \mid b^T x \le r\}$ is

equivalent to

(b") There exists $\alpha \in \mathbf{R}^m$ such that $\alpha \ge 0$, $A\alpha = b$, $\rho^T \alpha \le r$. Here again it only suffices to recall Lemma 3.2.8. The conical hull involved in (b) of Theorem 3.2.9 is already closed and the relation in (b") is really an inequality.

4. Conical Approximation of Convex sets

Let a set $S \subseteq \mathbf{R}^n$ and a point $x \in S$ be given. A profitable idea is the chance to approximate S by a "simpler" set near to x. In differential geometry we know that a " smooth " surface S can be approximated by an affine manifold "tangent" to S. This notion is mostly investigated in the differentiation of a "smooth" function

$f: \mathbf{R}^n \to \mathbf{R}$ where its graph is tangent to a hyperplane in $\mathbf{R}^n \times \mathbf{R}$:

$$\text{Graph of } f = \text{grf} \cong \{ (y,r) \in R^n \times R \mid r - f(x) = \langle \nabla f(x), y-x \rangle \}.$$

Nevertheless, as there is no reason for convex sets to be "smooth", a corresponding substitute to an affine manifold must be suggested. Truly, affine manifolds are translations of subspaces. We thus approximate S near to x by

$S \cong H_s(x) = \{x\} + V_s(x)$ where $V_s(x)$ is the subspace tangent to S at x.

Coming into the space of convex analysis, we recall that (see chapter 2.2) the closed convex cones are the usual substitutes for subspaces. Like just that of the Moreau's Theorem (see chapter 2.2) here too polarity plays the role of orthogonality.

4.1 Convenient Definitions of Tangent Cones

To bring the convenient objects in the foreground, it is imperative first to make a fresh glance about the general concept of tangency. Let $S \subseteq \mathbf{R}^n$ be an arbitrary closed set. Then classically a direction $d \in \mathbf{R}^n$ is said to be tangent to S at $x \in S$ when it is the derivative at x of some curve drawn on S. Here being rather interested by cones, we require only half derivative of the curve in consideration. What follows is thus the new definition of tangency that replaces curves by sequences so as to have a tangent direction for a given non-void set.

Definition 4.1.1: Let $S \subseteq \mathbf{R}^n$ be a non -empty set. Then $d \in \mathbf{R}^n$ is said to be a direction tangent to S at $x \in S$ if there exist sequences $\{x_k\} \subseteq S$ and $\{t_k\} \subseteq \mathbf{R}$ such that

$$\text{When } k\to\infty, \quad x_k \to x, \quad t_k \downarrow 0, \quad \frac{x_k - x}{t_k} \to d. \tag{4.1}$$

Furthermore, the set

$$T_S(x) = \{d \in \mathbf{R}^n \mid d \text{ is a direction tangent to S at } x \in S\}$$

is called the **tangent cone** to S at x.

Consider $0 \in \mathbf{R}^n$ and the sequences $x_k := x \in S$ for all $k \in \mathbf{N}$.

Obviously, $x_k \xrightarrow{k\to\infty} x$, $\frac{x_k - x}{t_k} \xrightarrow{k\to\infty} 0$ for any sequence $t_k \downarrow 0, \quad t_k \neq 0, \quad k \in N$.

It follows that

$$0 \in T_S(x).$$

This realizes that each non-void subset S of $\mathbf{R}^n$ adimits a tangent direction.

Let $d \in T_S(x)$. Then there exist sequences $\{x_k\} \subseteq S$ and $\{t_k\}$satisfying (4.1).

For all $\alpha>0$, take the sequnces $r_k := \frac{t_k}{\alpha}$. Hence we have that

$$x_k \xrightarrow{k\to\infty} x, \quad \frac{x_k - x}{r_k} \xrightarrow{k\to\infty} \alpha d, \quad r_k \downarrow 0.$$

Accordingly, αd is a direction tangent. This actually legalizes the terminology " tangent cone ."

Consider again a point $x \in \text{int}S$. Then there exists a $\delta > 0$ such that $B(x, \delta) \subseteq S$. By the axiom of Archimedes there is $k_0 \in \mathbf{N}$ where

$$\delta > \frac{1}{k} \text{ for all } k \geq k_0.$$

Let $d \in \mathbf{R}^n$, $x_k := \begin{cases} x + \frac{d}{k}, & \textit{for all } k \geq k_0 \\ x, & \textit{otherwise} \end{cases}$

and sequnce $t_k := \frac{1}{k}, \quad k \in N$. Then the sequnce (x_k) is in S.

Obviously, when $k \to \infty, \quad x_k \to x, \quad t_k \downarrow 0, \quad \textit{and} \quad \frac{x_k - x}{t_k} \to d$ (replace $\frac{1}{k}$ *by* $-\frac{1}{k}$ if need be). We thus get

$d \in T_S(x)$.

Consequently, for $x \in \text{int}S$, $T_S(x)$ becomes the whole space $\mathbf{R}^n$. The only interesting points are therefore those on the boundary of S. The following theorem gives an equivalent formulation of definition 4.1.1.

Theorem 4.1.1: A direction d is tangent to S at $x \in S$ if and only if there exist sequences $\{d_\kappa\}$and $\{t_\kappa\}$ such that

$$x + t_k d_k \in S \text{ for all } k \in \mathbf{N} \text{ and } d_k \to d, \quad t_k \downarrow 0$$

when $k \to \infty$.

Proof: Let $d \in T_S(x)$. Then there exist sequences $\{x_\kappa\} \subseteq S$ and $\{t_\kappa\}$satisfying (4.1).

Set $d_k := \frac{x_k - x}{t_k}$. Then

$$x_k = x + t_k d_k \in S \text{ for all } k \in \mathbf{N} \text{ and } d_k \to d, \quad t_k \downarrow 0$$

when $k \to \infty$.

Conversely, put $x_k := x + t_k d_k \in S$. Then definition 4.1.1 is easily satisfied for d to be in $T_S(x)$. □

Theorem 4.1.2: The tangent cone is closed.

Proof: Let (d_n) be a sequence in $T_S(x)$ converging to $d \in \mathbf{R}^n$. Thenfor each n take the sequnces $\{x_{n,k}\}_k$ and $\{t_{n,k}\}_k$ attached to d_n in the sense of definition 4.1.1.

Now, for a fixed $n_0 \in \mathbf{N}$, we can find k_{n_0} such that

$$\left\| \frac{x_{n,k_{n0}} - x}{t_{n,k_{n0}}} - d_{n_0} \right\| \leq \frac{1}{n_0}. \tag{4.2}$$

Letting $n_0 \to \infty$ in (4.2) gives

$$\frac{x_{n,k_{n0}} - x}{t_{n,k_{n0}}} \to d.$$

Therefore we obtain sequences $\{x_{n,k_{n0}}\}_{n_0}$ and $\{t_{n,k_{n0}}\}_{n_0}$ which assure d to be an element of $T_S(x)$.

As it only depends on the behavior of S near to x, tangency is basically a local concept. From its very definition $T_S(x)$ appears as the set of all possible cluster points of the set

$$\{\frac{y-x}{t}\} \text{ where } y \in S \text{ and } t \downarrow 0.$$

As a result, using set valued notations we have that

$$T_S(x) = \lim_{t \downarrow 0} ext\,(\frac{S-x}{t}) \tag{4.3}$$

where lim ext stands to the set of all cluster points. Still another formulation is possible using the distance function $d_S(x) := \min_{y \in S} \|y - x\|$

$$T_S(x) = \{d \in \mathbf{R}^n \mid \lim_{t \downarrow 0} \inf \frac{d_S(x+td)}{t} = 0\}. \tag{4.4}$$

If $x \in S$, then $d_S(x) = 0$. Thus the infimand in (4.4) can be also interpreted as

$$\inf_{t \in R} \frac{d_S(x+td) - d_S(x)}{t}.$$

4.2 The Tangent and the normal Cones to a convex set.

From now on instead of an arbitrary set S, we restrict our considerations to a closed convex set $C \subseteq \mathbf{R}^n$. In this very limited situations tangent cones can be given in a more handy expressions. The central point is to observe that the role of the property $t_k \downarrow 0$ in (4.4). If both x and $x + t_k d_k$ are in C, then convexity of C allows $x + \alpha d_k$ to be in C for all $\alpha \in (0, t_k)$ and in particular

$$C - \{x\} \subseteq T_C(x). \tag{4.5}$$

Theorem 4.2.1: The tangent cone to C at x is the closure of the cone generated by C - {x}, i.e.

$$T_C(x) = \overline{cone}\ (C - x),$$

$$= \mathrm{cl}\ \{d \in \mathbf{R}^n \mid d = \alpha\,(y-x), y \in C, \alpha \geq 0, \alpha \in R\}.$$

Proof: From (4.5) we have that $C - \{x\} \subseteq T_C(x)$. As $T_C(x)$ is a closed cone (Theorem 4.1.2), it immediately follows that

$$\mathrm{cl}\,(C - x) \subseteq T_C(x).$$

Conversely, for $d \in T_S(x)$, take the sequences $\{x_\kappa\}$ and $\{t_\kappa\}$ in the sense of definition 4.1.1. Then the sequence

$$\frac{x_k - x}{t_k} \in \mathrm{Cone}\,(C - x).$$

And hence its limit

$$d \in \overline{cone}\,(C - x). \qquad \square$$

This new equivalent formulation of the tangent cone is perhaps easier to work with. Furthermore, we can easily observe that

$$\mathrm{Cone}(C - x) = \bigcup_{t>0} \frac{C - x}{t}. \tag{4.6}$$

Due to the monotonocity (in the sense of inclusion) property of the mapping

$$t \mapsto \frac{C - x}{t},$$

(4,6) can have also the next limit formulation:

$$\mathrm{Cone}(C - x) = \lim_{t \downarrow 0} \frac{C - x}{t}. \tag{4.7}$$

As immediate consequences of Theorem 4.2.1 we get the following.

Corollary 4.2.2: (a) aff cone(C- $\{x\}$)= aff $T_C(x)$ = affC-$\{x\}$.

(b) $T_C(x)$ = affC-$\{x\}$ when $x \in$ riC.

(c) the sets cone(C-$\{x\}$) and its closure $T_C(x)$ have the same relative interior.

In particular when $x \in$ intC, then $T_C(x) = \mathbf{R}^n$. As a result, approximating C by $x + T_C(x)$ presents interest only when $x \in$ rbdC.

Recalling Corollary 3.2.4, $T_C(x)$ being a closed and convex set can also be described as the intersection of all closed half spaces containing it. In the present conical situation as what is coming up still some more can be said.

Definition 4.2.1: Let $C \subseteq \mathbf{R}^n$ be a non -empty, closed and convex set. Then a direction $s \in \mathbf{R}^n$ is said to be normal to C at $x \in C$ when:

$$\langle s, y - x \rangle \leq 0 \qquad \text{for all } y \in C. \tag{4.8}$$

From (4.8) we have that

$$\langle s, y - x \rangle = \|s\| \|y - x\| \cos\theta \leq 0 \text{ for all } y \in C.$$

This implies that

$$\theta \in [\frac{\pi}{2}, \pi).$$

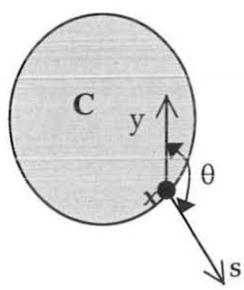


Fig. 4.2.1, a normal to a set

A normal is thus a vector s such that the angle between s and y - x is obtuse for all $y \in C$.

The set

$$N_C(x) := \{s \in \mathbf{R}^n \mid \langle s, y - x \rangle \leq 0 \text{ for all } y \in C\}$$

is called the **normal cone** to C at $x \in C$.

Obviously, $0 \in N_C(x)$ and thus every non -empty, closed and convex subset C of $\mathbf{R}^n$ has a normal vector.

Let x be a boundary point of C. Then by Lemma 3.2.2 there is a hyperplane $H_{s,r}$, $s \neq 0$ supporting C at x where

$$C \subseteq H^-_{s,r} \text{ and } r = \langle s, x \rangle.$$

Consequently we get that

$$\langle s, y \rangle \leq \langle s, x \rangle \text{ for all } y \in C.$$

or

$\langle s, y - x \rangle \leq 0$ for all y$\in$ C.

This again ensures the existence of a non-zero normal s at each boundary point x of C. In general, for x$\in\mathbf{R}^n$ by the Characterization Theorem we have that

$\langle x - P_C(x), y - P_C(x) \rangle \leq 0$ for all y$\in$ C.

To this effect

x - Pc(x) $\in$ Nc(Pc(x)) for each x$\in$ $\mathbf{R}^n$.

In contrast to that of Tc(x) , Nc(x) = {o}for each x$\in$ intC.

Example: Let C = $H^-_{s,r} := \{y \in R^n \mid \langle s, y\rangle \leq r\}$, a closed half spaces defined by a given (s, r)$\in R^n \times R$, $s \neq 0$. Then

Tc(x) = {d$\in\mathbf{R}^n$ | d = αs, $\alpha \geq 0$, $\alpha \in R$ }.

Theorem 4.2.3: Let C $\subseteq$ $\mathbf{R}^n$ be a non -empty, closed and convex set. Then Nc(x) is also a closed convex set.

Proof: Let s, z$\in$Nc(x). Then it follows that

$\langle s, y - x \rangle \leq 0$ and $\langle s, y - x \rangle \leq 0$ for all y $\in$ C.

For $\alpha \in$ (0, 1), consider

$\langle \alpha s + (1-\alpha) z, \alpha \langle s, y - x\rangle \leq 0$ for all y $\in$ C.

Thus

$\alpha s + (1-\alpha) z \in$Nc(x) for each $\alpha \in$ (0, 1).

For the completion of the proof, consider a sequence (s_K) in Nc(x) converging to s$\in\mathbf{R}^n$. Then by continuity of the scalar product we get:

$\langle s, y - x\rangle = \langle \lim_{k\to\infty}, y - x \rangle = \lim_{k\to\infty} \langle s_k, y - x \rangle \leq 0$ for all y $\in$ C.

Therefore

s$\in$ Nc(x). □

The next important theorem gives the relation between the normal and the tangent cones.

Theorem 4.2.4: The normal cone is the polar of the tangent cone, i.e.

Nc(x) = [Tc(x)]^0.

Proof: Let s $\in$ $N_C(x)$. Then

$$\langle s, y-x \rangle \leq 0 \quad \text{for all } y \in C.$$

or

$$\langle s, z \rangle \leq 0 \quad \text{for all } z \in C - x.$$

Take $d \in \text{Cone}(C - x)$ where $d = \alpha\, z$ for some $\alpha \geq 0$ and $z \in C - x$.

Then

$$\langle s, d \rangle = \alpha \langle s, z \rangle \leq 0 \quad \text{for all } d \in \text{Cone}(C - x).$$

Furthermore, if $r \in \overline{cone}\,(C - x)$, then

$$r = \lim_{k \to \infty} d_k, \quad d_k \in \text{Cone}(C - x) \quad \text{for all } k \in \mathbf{N}.$$

Applying continuity of the scalar product results:

$$\langle r, s \rangle = \lim_{k \to \infty} \langle d_k, s \rangle \leq 0 \quad \text{for all } r \in \overline{cone}\,(C - x).$$

But from Theorem 4.2.1, $T_C(x) = \overline{cone}\,(C - x)$ and thus $\langle r, s \rangle \leq 0$ for all $r \in T_C(x)$,

That is

$$s \in [\, T_C(x)]^0.$$

Conversely, let $s \in [\, T_C(x)]^0$. Then

$$\langle s, d \rangle \leq 0 \quad \text{for all } d \in T_C(x).$$

And it follows that

$$\langle s, d \rangle \leq 0 \quad \text{for all } d \in C - x \subseteq T_C(x).$$

Setting $d := z - x$, $z \in C$ gives:

$$\langle s, d \rangle = \langle s, z - x \rangle \leq 0 \quad \text{for all } z \in C.$$

That is to mean $s \in N_C(x)$. □

Combing this result with Thorem 4.1.3 and Thorem 3.2.5 gives the following another formulation of the tangent cone.

Corollary 4.2.5: The tangent cone is the polar of the normal cone.i.e,

$$T_C(x) = [\, N_C(x)]^0.$$

This corollary characterizes the tangent cone as intersection of homogeneous half spaces and comparing with chapter 3 of section (III) we have $r_s = 0$ for each s. Besides the index set $\sum T_C(x)$ is nothing more than $N_C(x)$. More interesting is like that of tangency normality is also a local concept. Truly, the normal cone to $C \cap B(x, \delta)$ at x coincides with $N_C(x)$. Indeed if $C - \{x\} \subseteq C' - \{x\} \subseteq T_C(x)$, then

$$N_{C'}(x) = N_C(x) \text{ and } T_{C'}(x) = T_C(x) .$$

For a fixed $x \in C$, $T_C(x)$ increases (in the sense of inclusion) if and only if $N_C(x)$ decreases. This in particular happen when set C increases.

Remark: The tangent and the normal cones to a non-closed convex set C (if need be) can be defined through replacing C by cl C in their respective definitions.

Example 1: Let C be a closed convex cone. Then using Theorem 3.2.5 and Corollary 4.2.5 we have that:

$$T_C(0) \;= [N_C(0)]0 \;= C^{00} = C.$$

Hence by Theorem 4.2.4

$$N_C(0) = [T_C(0)]^0 = C^0.$$

The polar of a closed convex cone is therefore its normal cone at 0.

On the other hand, if $x \in C$, $x \neq 0$, then observe that

$$T_C(x) \supset \{x\}.$$

It means that

$$N_C(x) \subseteq \{x\}^{\perp}.$$

Thus for $s \in T_C(x)$ we have that

$$\langle s, y - x \rangle = \langle s, y \rangle - \langle s, x \rangle = \langle s, y \rangle \leq 0 \quad \text{for all } y \in C.$$

As a result,

$$N_C(x) = \{ s \in C^0 \mid \langle s, x \rangle \;= 0 \}.$$

Example 2: Let $s_j \in \mathbf{R}^n$, $j \in \{1, 2, ..., m\}$ and $r_j \in \mathbf{R}$.

If C := $\{x \in \mathbf{R}^n \mid \langle s_j, x \rangle \le r_j \quad for\ j = 1, 2, \cdots, m\}$, Then C is a closed convex polyhedron. Let J(x) := $\{ j \in \{1, 2, ..., m\} \mid \langle s_j, x \rangle = r_j \}$. Then

$$N_C(x) = \text{Cone}\{ s_j \mid j \in J(x) \} = \{ \sum_{j \in J(x)} \alpha_j s_j \quad , \alpha_j \ge 0 \}.$$

$$T_C(x) = [N_C(x)]^0 = \{ d \in \mathbf{R}^n \mid \langle s_j, d \rangle \le 0 \quad for\ j \in J(x) \}.$$

4.3. Some more properties of the tangent and the normal cones

The tangent and the normal cones have more worth mentioning properties. Here on we exploit some of those that amalgamate various notions seen heretofore.

Theorem 4.3.1: Let $C_i \subseteq \mathbf{R}^n$, i =1, 2 be a non -empty, closed and convex sets. Then

(a) $T_{C_1 \cap C_2}(x) \subseteq T_{C_1}(x) \cap T_{C_2}(x)$ and

$N_{C_1}(x) + N_{C_2}(x) \subseteq N_{C_1 \cap C_2}(x)$ for each $x \in C_1 \cap C_2$.

(b) $T_{C_1 + C_2}(x_1 + x_2) = cl[T_{C_1}(x) + T_{C_2}(x_2)]$, and

$N_{C_1 + C_2}(x_1 + x_2) = N_{C_1}(x_1) \cap N\ (x_2)$ for $x_1 \in C_1$, $x_2 \in C_2$.

Proof: (a) Let $d \in T_{C_1 \cap C_2}(x)$. Then there exist sequences $\{x_K\} \subseteq C_1 \cap C_2$ and $\{t_K\}$ satisfying definition 4.1.1.

Obviously, $\{x_K\} \subseteq C_i$ and together with $\{t_K\}$ satisfy definition 4.1.1and hence assured to be an element of $T_{C_i}(x)$ for i = 1, 2.

We thus have

$$d \in T_{C_1}(x) \cap T_{C_2}(x).$$

For the other, consider $d \in N_{C_1}(x) + N_{C_2}(x)$. Then

$$d = d_1 + d_2 \text{ where } d_1 \in N_{C_1}(x) \text{ and } d_2 \in N_{C_2}(x).$$

Consequently for all $y_1 \in C_1$and $y_2 \in C_2$ we have that:

$$\langle d_1, y_1 - x \rangle \le 0 \text{ and } \langle d_2, y_2 - x \rangle \le 0.$$

or

$$\langle d_1, y - x \rangle + \langle d_2, y - x \rangle = \langle d_1 + d_2, y - x \rangle \le 0 \text{ for all } y \in C_1 \cap C_2.$$

In order to formulate equality in (a) additional supposition is necessary. One is of course when

$$0 \in \text{ri}(C_1 - C_2) \text{ or } \text{cl}(C_1 \cap C_2) = \text{cl}\, C_1 \cap \text{cl}\, C_2.$$

(b) $T_{C_1+C_2}(x_1+x_2) = \overline{Cone}\ (C_1 + C_2 - (x_1+x_2)),$

$$= \text{cl}\,\{d \in \mathbf{R}^n \mid d = \alpha(y_1+y_2) - \alpha(x_1+x_2),\ \ y_i \in C_i,\ \ \alpha \ge 0,\ \ i = 1, 2\},$$

$$= \text{cl}\,\{d \in \mathbf{R}^n \mid d \in \text{Cone}\,(C_1 - x_1) + \text{Cone}\,(C_2 - x_2)\},$$

$$= \text{cl}\,(T_{C_1}(x_1) + T_{C_2}(x_2)\,).$$

To show the other, we proceed as:

$$N_{C_1+C_2}(x_1+x_2) = \{s \in \mathbf{R}^n \mid \langle s, y-(x_1+x_2)\rangle \le 0 \quad \forall y \in C_1 + C_2\},$$

$$= \{s \in \mathbf{R}^n \mid \langle s, y_1 - x_1\rangle + \langle s, y_2 - x_2)\rangle \le 0 \quad \forall y \in C_1 + C_2\}.$$

This implies that

$$\{s \in \mathbf{R}^n \mid \langle s, y_1 - x_1\rangle \le 0 \quad \forall y_1 \in C_1\} \text{ and } \{s \in \mathbf{R}^n \mid \langle s, y_2 - x_2\rangle \le 0 \quad \forall y_2 \in C_2\}.$$

Thus $s \in N_{C_1}(x_1) \cap N\ (x_2)$. The other inclusion trivially follows.

Definition 4.3.1: Let $C \subseteq \mathbf{R}^n$ be a non -empty convex set. A point $x \in C$ is said to be an **extreme** point of C if and only if

x is not an interior point of a segment $[x_1, x_2]$ for all $x_1, x_2 \in C$.

The set of extreme points of C is denoted by extC.

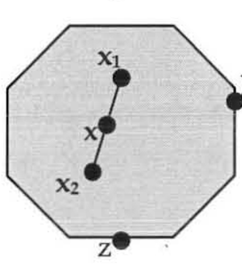


Fig.4.3.1

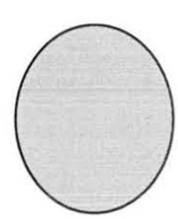


Fig.4.3.2

Extreme points necessarily lie on the boundary of the set. But every boundary point may not be an extreme point. For example In Fig.4.3.2 all boundary points are extreme points while in Fig.4.3.1 z is a boundary point but not an extreme point. In Fig.4.3.1 y is an extreme point while x is not an extreme point. The set extC is a closed set when $n \geq 2$. But in general extC has no topological or linear properties. Obviously, if x is an extreme point of C, then the set $C \setminus \{x\}$ is still a convex set. A convex set C is always the convex hull of its extreme points.

Example 1: Consider $C = \{x \in \mathbf{R}^n \mid \|x\| \leq 1\}$.

Then for $x_1, x_2 \in C$, we have the parallelogram equation

$$\frac{1}{2}\|x_1 + x_2\|^2 = \|x_1\|^2 + \|x_2\| - \frac{1}{2}\|x_1 - x_2\|^2.$$

Multiplying the equation by $\frac{1}{2}$ we get

$$\left\|\frac{x_1 + x_2}{2}\right\|^2 + \left\|\frac{x_1 - x_2}{2}\right\|^2 = \frac{1}{2}\left(\|x_1\|^2 + \|x_2\|^2\right).$$

This exhibits that every x of norm 1 is an extreme point of C.

Example2: Let C be a convex cone and $x \in C$. Then we have that

$$x = \frac{1}{3}x + \frac{2}{3}x.$$

This realizes that a non-zero $x \in C$ has no possibility to be an extreme point of C. Seemingly

affine manifolds and half spaces have no extreme points.

A higher dimensional generalization of extreme points is the following definition.

Definition 4.3.2: Let F and C be non-empty convex sets where

$$F \subseteq C \subseteq \mathbf{R}^n.$$ Then

F is said to be a **face** of C whenever $(x_1, x_2) \in C \times C$ and there exist $\alpha \in (0, 1)$ such that

$$\alpha x_1 + \alpha x_2 \in F \text{ implies } (x_1, x_2) \subseteq F.$$

Like other convex sets, a face has also its own affine hull, closure and relative interior. If $x \in \text{ext}C$, then $\{x\}$ is a face of C. That means extreme points appear as faces that are singletons.

Definition 4.3.3: Let $C \subseteq \mathbf{R}^n$ be a closed convex set. Then

$$F \subseteq C$$

is said to be an **exposed face** of C if there is a supporting hyperplane $H_{s,r}$ of C such that

$$F = C \cap H_{s,r}.$$

A point $x \in C$ at which there is a supporting hyperplane $H_{s,r}$ of C such that

$$C \cap H_{s,r} = \{x\}$$

is called an exposed point.

By definition the empty set and the set C itself are exposed faces of C. If a supporting hyperplane touches C, then the set of contact points is also an exposed face. Being the intersection of convex sets, an exposed face is also a face. Let F be an exposed face and $H_{s,r}$ be its associated supporting hyperplane. Then from the definition it immediately follows that

$$\langle s, y \rangle \leq \langle s, x \rangle \text{ for all } y \in C \text{ and all } x \in F.$$

The following is therefore an equivalent formulation of Definition 4.3.3.

The set $F \subseteq C$ is said to be an exposed face of C when there is a non- zero $s \in \mathbf{R}^n$ such that

$$F = \{x \in C \mid \langle s, x \rangle = \sup_{y \in C} \langle s, y \rangle \}.$$

The exposed face of C associated with $s \in \mathbf{R}^n$ is denoted by $F_C(s)$.

Theorem 4.3.2: Let $C \subseteq \mathbf{R}^n$ be a closed convex set and $x \in C$. Then the following properties are equivalent.

(i) $s \in N_C(x)$;

(ii) x is in the exposed face $F_C(s)$: $\langle s, x \rangle = \max_{y \in C} \langle s, y \rangle$;

(iii) $x = P_C(x + s)$.

Proof: (i) $\Rightarrow$ (ii):

Let $s \in N_C(x)$. Then

$$\langle s, y - x \rangle \leq 0 \quad \text{for all } y \in C.$$

or

$$\langle s, x \rangle \geq \langle s, y \rangle \quad \text{for all } y \in C.$$

Which implies that

$$\langle s, x \rangle \geq \max_{y \in C} \langle s, y \rangle .$$

Since $x \in C$, it follows that

$$\langle s, x \rangle = \max_{y \in C} \langle s, \ y \rangle .$$

(ii) $\Rightarrow$ (iii):

Suppose $\langle s, x \rangle = \max_{y \in C} \langle s, \ y \rangle$. Then

$$\langle s, x \rangle \geq \langle s, y \rangle \quad \text{for all } y \in C.$$

or

$$\langle s, y - x \rangle \leq 0 \quad \text{for all } y \in C.$$

Consequently,

$$\langle (s + x) - x, y - x \rangle \leq 0 \quad \text{for all } y \in C.$$

Then by Characterization Theorem we have that

$$x = P_C(x + s).$$

(iii) $\Rightarrow$ (i):

Assume $x = P_C(x + s)$. Then by Characterization Theorem it follows that

$$\langle s + x - x, y - x \rangle = \langle s, y \rangle \leq 0 \quad \text{for all } y \in C.$$

That implies

$$s \in N_C(x). \qquad \square$$

From $x = P_C\ (x + s)$, in particular we get

$$P_C^{-1}(x) = \{x\} + N_C(x) \text{ for all } x \in C.$$

Furthermore, since the projection is a single valued function for $x \neq x'$,

$$[\{x\} + N_C(x)\,] \cap [\{x'\} + N_C(x')\,] = \phi .$$

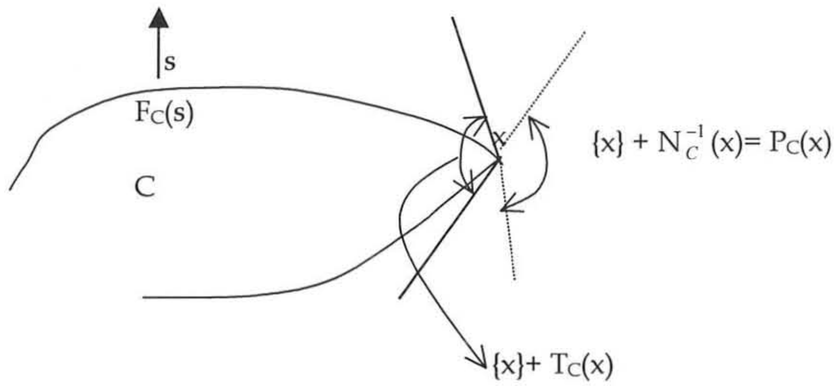


Fig.4.3.3 , Normal cones ,Projections and exposed face

Fix $x \in C$ and consider only those supporting hyperplanes that pass through x. Then the corresponding intersection of the half spaces construct $T_C(x)$ and ends up with

$$\{x\} + T_C(x) \supset C.$$

If need be we apply the closure operation especially when the relative boundary of C presents some curvature near to x. Doing this operation for all $x \in C$ yields

$$C \subseteq \bigcap_{x \in C} [\, x + T_C(x) \,]. \tag{4.9}$$

The inclusion in (4.9) is actually equality. To observe this we proceed as below. Consider a point $y \notin C$. Then write y as $y = P_C(y) + s$ where $s = y - P_C(y)$. Using Theorem 4.3.2 it follows that $s \in N_C[P_C(y)]$. Hence a non-zero s cannot be in $T_C[P_C(y)]$.That means

$$y \notin P_C(y) + T_C[P_C(y)].$$

It means that y is not in the right hand side of (4.9). Furthermore, if $x \in \text{ri}C$, then $T_C(x)$ expands to the whole aff$C$-$\{x\}$ and contains all other tangent cones. Thus x can be restricted to the relative boundary of C. To this effect

$$C = \bigcap_{x \in rbdC} [\, x + T_C(x) \,]$$

The above result is formulated in the following.

Theorem 4.3.3: Let $C \subseteq \mathbf{R}^n$ be a closed convex set and $x \in C$. Then

$$C = \bigcap_{x \in rbdC} [\, x + T_C(x) \,]$$

In chapter 2 we pointed out that the projection onto a closed convex set is a continuous operator. In conical situations still some more can be said. The normal and the tangent cones help to estimate the variation of this operator. In line with this we enclose this chapter with the following approximation result.

Theorem 4.3.4: Let $C \subseteq \mathbf{R}^n$ be a non -empty, closed and convex set. Then for $x \in C$ and $d \in \mathbf{R}^n$:

$$\lim_{t \downarrow 0} \frac{P_C(x+td)-x}{t} = P_{T_C(x)}(d). \tag{4.10}$$

Proof: By definition $P_C(x + td)$ is the projection of $x + td$ onto C. Applying the Characterization Theorem we have that

$$\langle x+td - P_C(x+td),\ z - P_C(x+td) \rangle \ \le 0 \quad \text{for all } z \in C.$$

Let $y = \frac{z-x}{t} \in \frac{C-x}{t}$. Then consider the following;

$$\left\langle\ d - \frac{P_C(x+td)-x}{t},\ y - \frac{P_C(x+td)-x}{t}\ \right\rangle =$$

$$= \frac{1}{t^2}\left\langle\ dt + x - P_C(x+td\,),\ z - P_C(x+td) - x \right\rangle \ \le 0 \text{ for all } z \in C.$$

This means that

$$\left\langle\ d - \frac{P_C(x+td)-x}{t},\ y - \frac{P_C(x+td)-x}{t}\ \right\rangle$$

$$\le 0 \quad \text{for all } y \in \frac{C-x}{t}.$$

Then by the Characterization Theorem we get

$$P_{\frac{C-x}{t}}(d) = \frac{P_C(x+td)-x}{t}.$$

Letting $t \downarrow 0$ and applying both (4.3) and (4.7) gives the desired result (4.10).

Considering (4.10) and using property of the projection operator we have that

$$\lim_{t \downarrow 0} \frac{P_C(x+td)-x}{t} \;=\; \lim_{t \downarrow 0} \frac{P_C(x+td)-P_C(x)}{t} \;=\; P_{T_C(x)}(d).$$

Thus $P_{T_C(x)}(d)$ is the directional derivative of P_C at x. □

Evidently, the notation $P'_{T_C(x)}(.) = P'_C(x,.)$ can be equally used. Moreover due to the approximation role of the tangent cones a possible notation is also

$$T_C(x) = C'(x).$$

In view of what has been said (4.10) can be rephrased as the projection and derivation operations commute each other. i.e. $P'_C(x,.) \;=\; P_{C'(x)}(.)$.

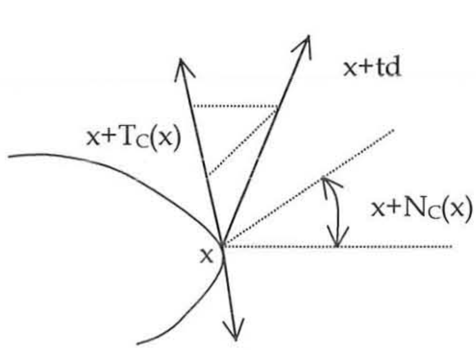


Fig.4.3.4, Differentiating a projection

References

[1] Deumlich, R. : Functional Analysis I .
Text Book for the lecture, Addis Ababa 1997.

[2] Deumlich, R. : Optimization and Theory of Approximation.
Text Book for the lecture, Addis Ababa 1997.

[3] Fletcher, R. : Practical Methods of optimization Volume I.
John Wiley and Sons, Cheichester, New York, Toronto, 1980.

[4] Hirit-Urruty,J.B, Lemarechal,C.: Convex Analysis and Optimization Algorithms Volume I.
Springer Verlag Berlin, Heidelberg, NewYork, 1993.

[5] Karmov,V. G.: Mathematical Programming.
Mir Publishers, Mosco, 1989.

[6] Macki, J. and Strausse, A. : Introduction to Optimal Control Theory.
Springer-Verlag, New York, Heidelberg, Berlin, 1982.

[7] Mital,K.V. : Optimization Methods in Operation Research and System Analysis.
Wiley Easter Limited, New Delhi, Bangalore, Bombay, Calcutta, 1983.

[8] Stoer,J. ,Witzgall, Ch.: Convexity and Optimization in Finite Dimensions
Springer-Verlag, Berlin, Heidelberg, NewYork, 1970.

[9] Treves, F. : Locally Convex Spaces and Linear Practical Differential Equation.
Springer-Verlag, Berlin, Heidelberg, New York, 1967.

[10] Walsh, G.R.: Methods of Optimization.
London, New York, Sydney, Toronto, 1975.