Convex Sets

Prof. Dan A. Simovici

UMB

Open and Closed Spheres

Definition

An open sphere of radius r centered in $\boldsymbol{x}_0$ is the set

$$B(\boldsymbol{x}_0, r) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \parallel \boldsymbol{x} - \boldsymbol{x}_0 \parallel < r\}.$$

A closed sphere of radius r centered in $\boldsymbol{x}_0$ is the set

$$B[\boldsymbol{x}_0, r] = \{\boldsymbol{x} \in \mathbb{R}^n \mid \parallel \boldsymbol{x} - \boldsymbol{x}_0 \parallel \leqslant r\}.$$

Closure of a Set

Definition

Let S be a subset of $\mathbb{R}^n$. A point $\boldsymbol{x}$ is in the closure of a set S if $S \cap B(\boldsymbol{x}, r) \neq \emptyset$ for every $r > 0$.
The closure of S is denoted by $\boldsymbol{K}(S)$.

If $S = \boldsymbol{K}(S)$, then S is said to be closed.

Interior of a Set

Definition

Let S be a subset of $\mathbb{R}^n$. A point $\boldsymbol{x}$ is in the interior of a set S if $B(\boldsymbol{x}, r) \subseteq S$ for some $r > 0$.
The interior of S is denoted by $\boldsymbol{I}(S)$.

If $S = \boldsymbol{I}(S)$, then S is said to be open.

Boundary of a Set

Definition

Let S be a subset of $\mathbb{R}^n$. A point $\boldsymbol{x}$ is in the border of a set S if we have both $B(\boldsymbol{x}, r) \cap S \neq \emptyset$ and $B(\boldsymbol{x}, r) \cap (\mathbb{R}^n - S) \neq \emptyset$ for every $r > 0$.
The border of S is denoted by $\partial(S)$.

Example

The set $S = B[\mathbf{0}_2, 1] = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leqslant 1\}$ is closed, that is, $S = \boldsymbol{K}(S)$.
The interior $\boldsymbol{I}(S)$ is

$$B(\mathbf{0}_2, 1) = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1^2 + x_2^2 < 1\},$$

while the border of S is

$$\partial S = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}.$$

Segments

Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. The *closed segment* determined by $\boldsymbol{x}$ and $\boldsymbol{y}$ is the set

$$[\boldsymbol{x}, \boldsymbol{y}] = \{a\boldsymbol{x} + (1-a)\boldsymbol{y} \mid 0 \leqslant a \leqslant 1\}.$$

The *half-closed segments* determined by $\boldsymbol{x}$ and $\boldsymbol{y}$ are the sets

$$[\boldsymbol{x}, \boldsymbol{y}) = \{a\boldsymbol{x} + (1-a)\boldsymbol{y} \mid 0 < a \leqslant 1\},$$

and

$$(\boldsymbol{x}, \boldsymbol{y}] = \{a\boldsymbol{x} + (1-a)\boldsymbol{y} \mid 0 \leqslant a < 1\}.$$

The *open segment* determined by $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$(\boldsymbol{x}, \boldsymbol{y}) = \{a\boldsymbol{x} + (1-a)\boldsymbol{y} \mid 0 < a < 1\}.$$

Definition

A subset C of $\mathbb{R}^n$ is *convex* if, for all $\boldsymbol{x}, \boldsymbol{y} \in C$ we have $[\boldsymbol{x}, \boldsymbol{y}] \subseteq C$.

Note that the empty subset and every singleton $\{\boldsymbol{x}\}$ of $\mathbb{R}^n$ is convex.

Example

Every linear subspace T of $\mathbb{R}^n$ is convex.

Example

The set $\mathbb{R}^n_{\geqslant 0}$ of all vectors of $\mathbb{R}^n$ having non-negative components is a convex set called the *non-negative orthant* of $\mathbb{R}^n$.

Example

The convex subsets of $(\mathbb{R}, +, \cdot)$ are the intervals of $\mathbb{R}$. Regular polygons are convex subsets of $\mathbb{R}^2$.

Example

An open sphere $C(\boldsymbol{x}_0, r) \subseteq \mathbb{R}^n$ is convex, where $\mathbb{R}^n$ is equipped with the Euclidean norm.
Indeed, suppose that $\boldsymbol{x}, \boldsymbol{y} \in C(\boldsymbol{x}_0, r)$, that is, $\parallel \boldsymbol{x} - \boldsymbol{x}_0 \parallel < r$ and $\parallel \boldsymbol{y} - \boldsymbol{x}_0 \parallel < r$.
Let $a \in [0,1]$ and let $\boldsymbol{z} = a\boldsymbol{x} + (1-a)\boldsymbol{y}$. We have

$$\begin{aligned} \parallel \boldsymbol{x}_0 - \boldsymbol{z} \parallel &= \parallel \boldsymbol{x}_0 - a\boldsymbol{x} - (1-a)\boldsymbol{y} \parallel \\ &= \parallel a(\boldsymbol{x}_0 - \boldsymbol{x}) + (1-a)(\boldsymbol{x}_0 - \boldsymbol{y}) \parallel \\ &\leqslant a \parallel \boldsymbol{x}_0 - \boldsymbol{x} \parallel + (1-a) \parallel \boldsymbol{x}_0 - \boldsymbol{y}) \parallel \leqslant r. \end{aligned}$$

so $\boldsymbol{z} \in C(\boldsymbol{x}_0, r)$.

Definition

Let U be a subset of $\mathbb{R}^n$ and let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in U$. A linear combination of U, $a_1\boldsymbol{x}_1 + \cdots + a_k\boldsymbol{x}_k$ is

- an affine combination of U if $\sum_{i=1}^{k} a_i = 1$;
- a *non-negative combination of U* if $a_i \geqslant 0$ for $1 \leqslant i \leqslant k$;
- a *positive combination of U* if $a_i > 0$ for $1 \leqslant i \leqslant k$;
- a *convex combination* of U if it is a non-negative combination of U and $a_1 + \cdots + a_k = 1$.

Theorem

A subset C of $\mathbb{R}^n$ is convex if and only if any convex combination of elements of C belongs to C.

Theorem

Let $C_1, \ldots, C_k$ be convex subsets of $\mathbb{R}^n$. If $b_1, \ldots, b_k \in \mathbb{R}$, then

$$\{\boldsymbol{y} = b_1\boldsymbol{x}_1 + \cdots + b_k\boldsymbol{x}_k \mid \boldsymbol{x}_i \in C_i \text{ for } 1 \leqslant i \leqslant k\}$$

is a convex set.

Theorem

If C is a convex subset of $\mathbb{R}^m$ and $f : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ is an affine mapping, then the set $f(C)$ is a convex subset of $\mathbb{R}^n$.
If D is a convex subset of $\mathbb{R}^n$, then $f^{-1}(D) = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \in D\}$ is a convex subset of $\mathbb{R}^m$.

Proof: Since f is an affine mapping we have $f(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}$, where $A \in \mathbb{R}^{n\times m}$ and $\boldsymbol{b} \in \mathbb{R}^n$ for $\boldsymbol{x} \in \mathbb{R}^m$. Therefore, if $\boldsymbol{y}_1, \boldsymbol{y}_2 \in f(C)$ we can write $\boldsymbol{y}_1 = A\boldsymbol{x}_1 + \boldsymbol{b}$ and $\boldsymbol{y}_2 = A\boldsymbol{x}_2 + \boldsymbol{b}$. This, in turn, allows us to write for $a \in [0, 1]$:

$$\begin{aligned} a\boldsymbol{y}_1 + (1-a)\boldsymbol{y}_2 &= a(A\boldsymbol{x}_1 + \boldsymbol{b}) + (1-a)(A\boldsymbol{x}_2 + \boldsymbol{b}) \\ &= A(a\boldsymbol{x}_1 + (1-a)\boldsymbol{x}_2) + \boldsymbol{b} \\ &= h(a\boldsymbol{x}_1 + (1-a)\boldsymbol{x}_2). \end{aligned}$$

The convexity of C implies $a\boldsymbol{x}_1 + (1-a)\boldsymbol{x}_2 \in C$, so $a\boldsymbol{y}_1 + (1-a)\boldsymbol{y}_2 \in f(C)$, which shows that $f(C)$ is convex.

Definition

A subset D of $\mathbb{R}^n$ is *affine* if, for all $\boldsymbol{x}, \boldsymbol{y} \in C$ and all $a \in \mathbb{R}$, we have $a\boldsymbol{x} + (1-a)\boldsymbol{y} \in D$.
In other words, D is an affine set if every point on the line determined by $\boldsymbol{x}$ and $\boldsymbol{y}$ belongs to C.

Note that D is a subspace of $\mathbb{R}^n$ if $\mathbf{0} \in D$ and D is an affine set.

Theorem

Let D be a non-empty affine set in $\mathbb{R}^n$. There exists translation $\mathrm{t}_{\boldsymbol{u}}$ and a unique subspace L of $\mathbb{R}^n$ such that $D = \mathrm{t}_{\boldsymbol{u}}(L)$.

Proof

Let $L = \{\boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{y} \in D\}$ and let $\boldsymbol{x}_0 \in D$. We have $\mathbf{0} = \boldsymbol{x}_0 - \boldsymbol{x}_0 \in L$ and it is immediate that L is an affine set. Therefore, L is a subspace.
Suppose that $D = \mathrm{t}_{\boldsymbol{u}}(L) = \mathrm{t}_{\boldsymbol{v}}(K)$, where both L and K are subspaces of $\mathbb{R}^n$. Since $\mathbf{0} \in K$, it follows that there exists $\boldsymbol{w} \in L$ such that $\boldsymbol{u} + \boldsymbol{w} = \boldsymbol{v}$. Similarly, since $\mathbf{0} \in L$, it follows that there exists $\boldsymbol{t} \in K$ such that $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{t}$. Consequently, since $\boldsymbol{w} + \boldsymbol{t} = \mathbf{0}$, both $\boldsymbol{w}$ and $\boldsymbol{t}$ belong to both subspaces L and K.
If $\boldsymbol{s} \in L$, it follows that $\boldsymbol{u} + \boldsymbol{s} = \boldsymbol{v} + \boldsymbol{z}$ for some $\boldsymbol{z} \in K$. Therefore, $\boldsymbol{s} = (\boldsymbol{v} - \boldsymbol{u}) + \boldsymbol{z} \in K$ because $\boldsymbol{w} = \boldsymbol{v} - \boldsymbol{u} \in K$. This implies $L \subseteq K$. The reverse inclusion can be shown similarly.

Theorem

Let $A \in \mathbb{R}^{m \times n}$ and let $\boldsymbol{b} \in \mathbb{R}^m$. The set $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} = \boldsymbol{b}\}$ is an affine subset of $\mathbb{R}^n$. Conversely, every affine subset of $\mathbb{R}^n$ is the set of solutions of a system of the form $A\boldsymbol{x} = \boldsymbol{b}$.

Proof

It is immediate that the set of solutions of a linear system is affine. Conversely, let S be an affine subset of $\mathbb{R}^n$ and let L be the linear subspace such that $S = \boldsymbol{u} + L$. Let $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$ be a basis of $L^\perp$. We have

$$L = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}_i'\boldsymbol{x} = 0 \text{ for } 1 \leqslant i \leqslant m\} = \{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} = \boldsymbol{0}\},$$

where A is a matrix whose rows are $\boldsymbol{a}_1', \ldots, \boldsymbol{a}_m'$. By defining $\boldsymbol{b} = A\boldsymbol{u}$ we have

$$S = \{\boldsymbol{u} + \boldsymbol{x} \mid A\boldsymbol{x} = \boldsymbol{0}\} = \{\boldsymbol{y} \in \mathbb{R}^n \mid A\boldsymbol{y} = \boldsymbol{b}\}.$$

Theorem

The intersection of any collection of convex (affine) sets in $\mathbb{R}^n$ is a convex (affine) set.

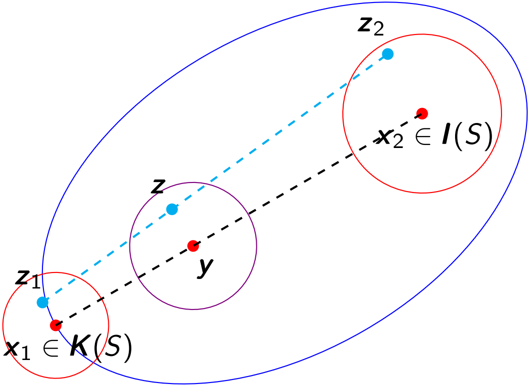
This allows us to define the convex closure $\boldsymbol{K}_{\mathrm{conv}}(S)$ of a subset S of $\mathbb{R}^n$ as the intersection of all convex sets that contain S. This is the least convex set that contains S.
Simlarly, $\boldsymbol{K}_{\mathrm{aff}}(S)$, the intersection of all affine sets that contain S is the least affine subset of $\mathbb{R}^n$ that contains S.

Theorem

Let S be a convex set in $\mathbb{R}^n$ with $\boldsymbol{I}(S) \neq \emptyset$. If $\boldsymbol{x}_1 \in \boldsymbol{K}(S)$ and $\boldsymbol{x}_2 \in \boldsymbol{I}(S)$, then $a\boldsymbol{x}_1 + (1-a)\boldsymbol{x}_2 \in S$ for $a \in (0,1)$.

Proof

Since $\boldsymbol{x}_2 \in \boldsymbol{I}(S)$ there exists $\epsilon > 0$ such that $B(\boldsymbol{x}_2, \epsilon) \subseteq S$. Let $\boldsymbol{y} = a\boldsymbol{x}_1 + (1-a)\boldsymbol{x}_2$. To show that $\boldsymbol{y} \in \boldsymbol{I}(S)$ it is sufficient to show that $B(\boldsymbol{y}, (1-a)\epsilon) \subseteq S$.



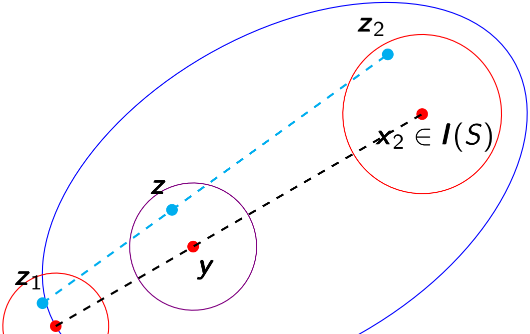
Proof (cont'd)

Since $\boldsymbol{x}_1 \in \boldsymbol{K}(S)$ we have

$$B\left(\boldsymbol{x}_1, \frac{(1-a)\epsilon - \parallel \boldsymbol{z}-\boldsymbol{y} \parallel}{a}\right) \cap S \neq \emptyset.$$

In particular, there exists $\boldsymbol{z}_1 \in S$ such that

$$\parallel \boldsymbol{z}_1 - \boldsymbol{x}_1 \parallel < \frac{(1-a)\epsilon - \parallel \boldsymbol{z}-\boldsymbol{y} \parallel}{a}.$$



Proof (cont'd)

Define $\boldsymbol{z}_1 = \frac{\boldsymbol{z}-a\boldsymbol{z}_1}{1-a}$. This allows us to write

$$\begin{aligned}
\parallel \boldsymbol{z}_2 - \boldsymbol{x}_2 \parallel &= \left\| \frac{\boldsymbol{z} - a\boldsymbol{z}_1}{1-a} - \boldsymbol{x}_2 \right\| \\
&= \frac{1}{1-a} \parallel (\boldsymbol{z} - \boldsymbol{y}) + a(\boldsymbol{x}_1 - \boldsymbol{z}_1) \parallel \\
&\leqslant \frac{1}{1-a}(\parallel (\boldsymbol{z} - \boldsymbol{y}) \parallel + a \parallel \boldsymbol{x}_1 - \boldsymbol{z}_1 \parallel < \epsilon,
\end{aligned}$$

so $\boldsymbol{z}_2 \in S$. By the definition of $\boldsymbol{z}_2$ note that $\boldsymbol{z} = a\boldsymbol{z}_1 + (1-a)\boldsymbol{z}_2$. Since $\boldsymbol{z}_1, \boldsymbol{z}_2 \in S$, we have $\boldsymbol{z} \in S$. Therefore, $\boldsymbol{y} \in I(\boldsymbol{S})$.

Corollary

For a convex set S, $\boldsymbol{I}(S)$ is convex.

Corollary

If S is a convex set and $\boldsymbol{I}(S) \neq \emptyset$, then $\boldsymbol{K}(S)$ is convex

Proof

Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{K}(S)$ and let $\boldsymbol{z} \in \boldsymbol{I}(S)$. By the theorem on slide 22, $a\boldsymbol{x}_2 + (1-a)\boldsymbol{z} \in \boldsymbol{I}(S)$ for each $a \in (0,1)$. For $b \in (0,1)$ we have $b\boldsymbol{x}_1 + (1-b)(a\boldsymbol{x}_2 + (1-a)\boldsymbol{z})\boldsymbol{I}(S) \subseteq S$. Since $\lim_{a\to 1} b\boldsymbol{x}_1 + (1-b)(a\boldsymbol{x}_2 + (1-a)\boldsymbol{z}) = b\boldsymbol{x}_1 + (1-b)\boldsymbol{x}_2 \in \boldsymbol{K}(S)$.

Corollary

Let S be a set with $\boldsymbol{I}(S) \neq \emptyset$. Then, $\boldsymbol{K}(\boldsymbol{I}(S)) = \boldsymbol{K}(S)$.

Proof

It is clear that $\boldsymbol{K}(\boldsymbol{I}(S)) \subseteq \boldsymbol{K}(S)$. Let $\boldsymbol{x} \in \boldsymbol{K}(S)$ and $\boldsymbol{y} \in \boldsymbol{I}(S)$ (since $\boldsymbol{I}(S) \neq \emptyset$). Then, $a\boldsymbol{x} + (1-a)\boldsymbol{y} \in \boldsymbol{I}(S)$ for each $a \in (0,1)$. Since $\boldsymbol{x} = \lim_{a \to 1} a\boldsymbol{x} + (1-a)\boldsymbol{y} \in \boldsymbol{K}(\boldsymbol{I}(S))$, the equality follows.

Corollary

Let S be a set with $\boldsymbol{I}(S) \neq \emptyset$. Then $\boldsymbol{I}(\boldsymbol{K}(S)) = \boldsymbol{I}(S)$.

Proof

We have $\boldsymbol{I}(S) \subseteq \boldsymbol{I}(\boldsymbol{K}(S))$. Let $\boldsymbol{x}_1 \in \boldsymbol{I}(\boldsymbol{K}(S))$. There exists $\epsilon > 0$ such that $B(\boldsymbol{x}_1, \epsilon) \subseteq \boldsymbol{K}(S)$. Let $\boldsymbol{x}_2 \neq \boldsymbol{x}_1$ that belongs to $\boldsymbol{I}(S)$ and let

$$\boldsymbol{y} = (1 + b)\boldsymbol{x}_1 - b\boldsymbol{x}_2,$$

where $b = \frac{\epsilon}{2\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|}$. Since $\| \boldsymbol{y} - \boldsymbol{x}_1 \| = \frac{\epsilon}{2}$, we have $\boldsymbol{y} \in \boldsymbol{K}(S)$. But $\boldsymbol{x}_1 = c\boldsymbol{y} + (1 - c)\boldsymbol{x}_2$, where $c = \frac{1}{1+b} \in (0, 1)$. Since $\boldsymbol{y} \in \boldsymbol{K}(S)$ and $\boldsymbol{x}_2 \in \boldsymbol{I}(S)$, then $\boldsymbol{x}_1 \in \boldsymbol{I}(S)$.

The Proximal Point

Lemma

Let C be a nonempty and closed convex set, $C \subseteq \mathbb{R}^n$ and let $\boldsymbol{x}_0 \notin C$. There exits a unique point $\boldsymbol{u} \in C$ such that $\| \boldsymbol{u} - \boldsymbol{x}_0 \|$ is the minimal distance from $\boldsymbol{x}_0$ to C.

Proof

Let $\mu = \min\{\| \mathbf{x} - \mathbf{x}_0 \| \mid \mathbf{x} \in C\}$. There exists a sequence of elements in C, $(\mathbf{z}_n)$ such that $\lim_{n\to\infty} \| \mathbf{z}_n - \mathbf{x}_0 \| = \mu$. By the law of the parallelogram,
$\| \mathbf{z}_k - \mathbf{z}_m \|^2 = 2 \| \mathbf{z}_k - \mathbf{x}_0 \|^2 + 2 \| \mathbf{z}_m - \mathbf{x}_0 \|^2 - 4 \| \frac{\mathbf{x}_k + \mathbf{x}_m}{2} - \mathbf{x}_0 \|^2$. Since C is convex, we have $\frac{\mathbf{x}_k + \mathbf{x}_m}{2} \in C$; the definition of μ implies that

$$\left\| \frac{\mathbf{x}_k + \mathbf{x}_m}{2} - \mathbf{x}_0 \right\|^2 \geqslant \mu^2,$$

so

$$\| \mathbf{z}_k - \mathbf{z}_m \|^2 \leqslant 2 \| \mathbf{z}_k - \mathbf{x}_0 \|^2 + 2 \| \mathbf{z}_m - \mathbf{x}_0 \|^2 - 4\mu^2.$$

Proof (cont'd)

Since $\lim_{n\to\infty} \| \boldsymbol{z}_n - \boldsymbol{x}_0 \| = \mu$, for every $\epsilon > 0$, there exists n_ϵ such that $k, m > n_\epsilon$ imply $\| \boldsymbol{z}_k - \boldsymbol{x}_0 \| < \mu\epsilon$ and $\| \boldsymbol{z}_m - \boldsymbol{x}_0 \| < \mu\epsilon$. Therefore, if $k, m > n_\epsilon$, it follows that

$$\| \boldsymbol{z}_k - \boldsymbol{z}_m \|^2 \leqslant 4\mu^2(\epsilon^2 - 1).$$

Thus, $(\boldsymbol{z}_n)$ is a Cauchy sequence. If $\lim_{n\to\infty} \boldsymbol{z}_n = \boldsymbol{u}$, then $\boldsymbol{u} \in C$ because C is a closed set.

Suppose $\boldsymbol{v} \in C$ with $\boldsymbol{v} \neq \boldsymbol{v}$ and $\parallel \boldsymbol{v} - \boldsymbol{x}_0 \parallel = \parallel \boldsymbol{u} - \boldsymbol{x}_0 \parallel$. Since C is convex, $\boldsymbol{w} = \frac{1}{2}(\boldsymbol{u} + \boldsymbol{v}) \in C$ and we have

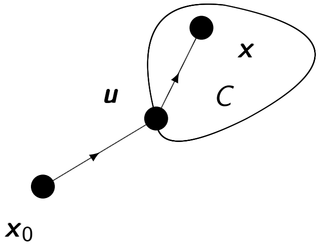
$$\left\| \frac{1}{2}(\boldsymbol{u} + \boldsymbol{v}) - \boldsymbol{x}_0 \right\| \leqslant \frac{1}{2} \parallel \boldsymbol{u} - \boldsymbol{x}_0 \parallel + \frac{1}{2} \parallel \boldsymbol{v} - \boldsymbol{x}_0 \parallel = \mu.$$

If $\left\| \frac{1}{2}(\boldsymbol{u} + \boldsymbol{v}) - \boldsymbol{x}_0 \right\| < \mu$, the definition of μ is violated. Therefore, we have

$$\left\| \frac{1}{2}(\boldsymbol{u} + \boldsymbol{v}) - \boldsymbol{x}_0 \right\| = \mu,$$

which implies $\boldsymbol{u} - \boldsymbol{x}_0 = k(\boldsymbol{v} - \boldsymbol{x}_0)$ for some $k \in \mathbb{R}$. This, in turn, implies $|k| = 1$. If $k = 1$ we would have $\boldsymbol{u} - \boldsymbol{x}_0 = \boldsymbol{v} - \boldsymbol{x}_0$, so $\boldsymbol{u} = \boldsymbol{v}$, which is a contradiction. Therefore, $a = 1$ and this implies $\boldsymbol{x}_0 = \frac{1}{2}(\boldsymbol{u} + \boldsymbol{v}) \in C$, which is again a contradiction. This implies that $\boldsymbol{u}$ is indeed unique.

The point $\boldsymbol{u}$ whose existence and uniqueness is was established is the C-proximal point to $\boldsymbol{x}_0$.



Lemma

Let C be a nonempty and closed convex set, $C \subseteq \mathbb{R}^n$ and let $\mathbf{x}_0 \notin C$. Then $\mathbf{u} \in C$ is the C-proximal point to $\mathbf{x}_0$ if and only if for all $\mathbf{x} \in C$ we have

$$(\mathbf{x} - \mathbf{u})'(\mathbf{u} - \mathbf{x}_0) \geqslant 0.$$

Proof

Let $\mathbf{x} \in C$. Since

$$\begin{aligned} \| \mathbf{x} - \mathbf{x}_0 \|^2 &= \| \mathbf{x} - \mathbf{u} + \mathbf{u} - \mathbf{x}_0 \|^2 \\ &= \| \mathbf{x} - \mathbf{u} \|^2 + \| \mathbf{u} - \mathbf{x}_0 \|^2 + (\mathbf{x} - \mathbf{u})'(\mathbf{u} - \mathbf{x}_0), \end{aligned}$$

$\| \mathbf{u} - \mathbf{x}_0 \|^2 \geqslant 0$ and $(\mathbf{x} - \mathbf{u})'(\mathbf{u} - \mathbf{x}_0) \geqslant 0$, it follows that $\| \mathbf{x} - \mathbf{x}_0 \| \geqslant \| \mathbf{x} - \mathbf{u} \|$, which means that $\mathbf{u}$ is the closest point in C to $\mathbf{x}_0$, and the condition of the lemma is sufficient.

Proof (cont'd)

Conversely, suppose that $\boldsymbol{u}$ is the proximal point in C to $\boldsymbol{x}_0$, that is, $\| \boldsymbol{x} - \boldsymbol{x}_0 \| \geqslant \| \boldsymbol{x}_0 - \boldsymbol{u} \|$ for $\boldsymbol{x} \in C$. If t is positive and sufficiently small, then $\boldsymbol{u} + t(\boldsymbol{x} - \boldsymbol{u}) \in C$ because $\boldsymbol{x} \in C$. Consequently,

$$\| \boldsymbol{x}_0 - \boldsymbol{u} - t(\boldsymbol{x} - \boldsymbol{u}) \|^2 \geqslant \| \boldsymbol{x}_0 - \boldsymbol{u} \|^2 .$$

Since

$$\| \boldsymbol{x}_0 - \boldsymbol{u} - t(\boldsymbol{x} - \boldsymbol{u}) \|^2 = \| \boldsymbol{x}_0 - \boldsymbol{u} \|^2 - 2t(\boldsymbol{x}_0 - \boldsymbol{u})'(\boldsymbol{x} - \boldsymbol{u}) + t^2 \| \boldsymbol{x} - \boldsymbol{u} \|^2$$

it follows that

$$-2t(\boldsymbol{x}_0 - \boldsymbol{u})'(\boldsymbol{x} - \boldsymbol{u}) + t^2 \| \boldsymbol{x} - \boldsymbol{u} \|^2 \geqslant 0,$$

which implies $(\boldsymbol{x} - \boldsymbol{u})'(\boldsymbol{u} - \boldsymbol{x}_0) \geqslant 0$, when we divide the previous equality by $-a \leqslant 0$.

Definition

Let S_1, S_2 be two subsets of $\mathbb{R}^n$ and let $H_{\boldsymbol{w},a}$ be a hyperplane in $\mathbb{R}^n$. $H_{\boldsymbol{w},a}$

- separates S_1 and S_2 if $\boldsymbol{w}'\boldsymbol{x} \geqslant a$ for $\boldsymbol{x} \in S_1$ and $\boldsymbol{w}'\boldsymbol{x} \leqslant a$ for $\boldsymbol{x} \in S_2$;
- strictly separates S_1 and S_2 if $\boldsymbol{w}'\boldsymbol{x} > a$ for $\boldsymbol{x} \in S_1$ and $\boldsymbol{w}'\boldsymbol{x} < a$ for $\boldsymbol{x} \in S_2$;
- strongly separates S_1 and S_2 if $\boldsymbol{w}'\boldsymbol{x} > a + \epsilon$ for $\boldsymbol{x} \in S_1$ and $\boldsymbol{w}'\boldsymbol{x} < a$ for $\boldsymbol{x} \in S_2$ and some $\epsilon > 0$.

Separation between a Convex Set and a Point

Theorem

Let S be a non-empty convex set in $\mathbb{R}^n$ and $\mathbf{y} \notin S$. There exists $\mathbf{w} \neq \mathbf{0}_n$ and $a \in \mathbb{R}$ such that $\mathbf{w}'\mathbf{y} > a$ and $\mathbf{w}'\mathbf{x} \leqslant a$ for $\mathbf{x} \in S$.

Proof

Since S is non-empty and closed and $\boldsymbol{y} \notin S$ there exists a unique closest point $\boldsymbol{x}_0 \in S$ such that $(\boldsymbol{x}-\boldsymbol{x}_0)'(\boldsymbol{y}-\boldsymbol{x}_0) \leqslant 0$ for each $\boldsymbol{x} \in S$. Equivalently,

$$-\boldsymbol{x}_0'(\boldsymbol{y}-\boldsymbol{x}_0) \leqslant -\boldsymbol{x}'(\boldsymbol{y}-\boldsymbol{x}_0).$$

Since

$$\begin{aligned} \parallel \boldsymbol{y}-\boldsymbol{x}_0 \parallel^2 &= (\boldsymbol{y}-\boldsymbol{x}_0)'(\boldsymbol{y}-\boldsymbol{x}_0) \\ &= \boldsymbol{y}'(\boldsymbol{y}-\boldsymbol{x}_0) - \boldsymbol{x}_0'(\boldsymbol{y}-\boldsymbol{x}_0) \\ &\leqslant \boldsymbol{y}'(\boldsymbol{y}-\boldsymbol{x}_0) - \boldsymbol{x}'(\boldsymbol{y}-\boldsymbol{x}_0) \\ &= (\boldsymbol{y}-\boldsymbol{x})'(\boldsymbol{y}-\boldsymbol{x}_0), \end{aligned}$$

for $\boldsymbol{w} = \boldsymbol{y}-\boldsymbol{x}_0 \neq \boldsymbol{0}_n$ we have

$$\boldsymbol{w}'(\boldsymbol{y}-\boldsymbol{x}) \geqslant \parallel \boldsymbol{y}-\boldsymbol{x}_0 \parallel^2,$$

so $\boldsymbol{w}'\boldsymbol{y} \geqslant \boldsymbol{w}'\boldsymbol{x} + \parallel \boldsymbol{y}-\boldsymbol{x}_0 \parallel^2$.
If $a = \sup\{\boldsymbol{w}'\boldsymbol{x} \mid \boldsymbol{x} \in S\}$ we have the desired inequalities.

A variation of the previous theorem, where C is just a convex set (not necessarily closed) is given next.

Theorem

Let C be a nonempty convex set, $C \subseteq \mathbb{R}^n$ and let $\boldsymbol{x}_0 \in \partial C$. There exists $\boldsymbol{w} \in \mathbb{R}^n - \{\mathbf{0}_n\}$ and $a \in R$ such that $\boldsymbol{w}'(\boldsymbol{x} - \boldsymbol{x}_0) \leqslant 0$ for $\boldsymbol{x} \in \boldsymbol{K}(C)$.

Proof

Since $\boldsymbol{x}_0 \in \partial C$, there exists a sequence $(\boldsymbol{z}_m)$ such that $\boldsymbol{z}_m \notin \boldsymbol{K}(C)$ and $\lim_{m\to\infty} \boldsymbol{z}_m = \boldsymbol{x}_0$. By Theorem on slide 41, for each $m \in \mathbb{N}$ there exists $\boldsymbol{w}_m \in \mathbb{R}^n - \{\boldsymbol{0}_n\}$ such that $\boldsymbol{w}'_m\boldsymbol{z}_m > \boldsymbol{w}'_m\boldsymbol{x}$ for each $\boldsymbol{x} \in \boldsymbol{K}(C)$. Without loss of generality we may assume that $\parallel \boldsymbol{w}_m \parallel = 1$. Since the sequence $(\boldsymbol{w}_m)$ is bounded, it contains a convergent subsequence $\boldsymbol{w}_{i_p}$ such that $\lim_{p\to\infty} \boldsymbol{w}_{i_p} = \boldsymbol{w}$ and we have $\boldsymbol{w}'_{i_p}\boldsymbol{z}_{i_p} > \boldsymbol{w}'_{i_p}\boldsymbol{x}$ for each $\boldsymbol{x} \in \boldsymbol{K}(C)$. Taking $p \to \infty$ we obtain $\boldsymbol{w}'\boldsymbol{x}_0 > \boldsymbol{w}'\boldsymbol{x}$ for $\boldsymbol{x} \in \boldsymbol{K}(C)$.

Theorem

Let C be a nonempty convex set, $C \subseteq \mathbb{R}^n$ and let $\mathbf{x}_0 \notin C$. There exists $\mathbf{w} \in \mathbb{R}^n - \{\mathbf{0}_n\}$ and $a \in R$ such that $\mathbf{w}'(\mathbf{x} - \mathbf{x}_0) \leqslant 0$ for $\mathbf{x} \in \mathbf{K}(C)$.

Proof: If $\mathbf{x} \notin \mathbf{K}(C)$, the statement follows from the Theorem on slide 41. Otherwise, $\mathbf{x}_0 \in \mathbf{K}(C) - C \subseteq \partial C$, so $\mathbf{x}_0 \in \partial C$ and the statement is a consequence of Theorem from slide 43.

Theorem

Let $C \subseteq \mathbb{R}^n$ be a closed and convex set. Then, C equals the intersection of all half-spaces that contain C.

Proof: It is immediate that C is included in the intersection of all half-spaces that contain C. Conversely, suppose that $\boldsymbol{z}$ be a point contained in all halfspaces that contain C such that $\boldsymbol{z} \notin C$. There exists a half-space that contains C but not $\boldsymbol{z}$, which contradicts the definition of $\boldsymbol{z}$. Thus, the intersection of all half-spaces that contain C equals C.

Definition

Let S be a non-empty subset of $\mathbb{R}^n$ and let $\boldsymbol{x}_0 \in \partial(S)$. A supporting hyperplane of S at $\boldsymbol{x}_0$ is a hyperplane $H_{\boldsymbol{w},a}$ such that either $S \subseteq H^+_{\boldsymbol{w},a}$ where $\boldsymbol{w}'(\boldsymbol{x} - \boldsymbol{x}_0) \geqslant 0$ for each $\boldsymbol{x} \in S$, or $S \subseteq H^-_{\boldsymbol{w},a}$ where $\boldsymbol{w}'(\boldsymbol{x} - \boldsymbol{x}_0) \leqslant 0$ for each $\boldsymbol{x} \in S$.

Equivalently, $H_{\boldsymbol{w},a}$ is a supporting hyperplane at $\boldsymbol{x}_0 \in partial(S)$ if either $\boldsymbol{w}'\boldsymbol{x}_0 = \inf\{\boldsymbol{w}'\boldsymbol{x} \mid \boldsymbol{x} \in S\}$, or $\boldsymbol{w}'\boldsymbol{x}_0 = \sup\{\boldsymbol{w}'\boldsymbol{x} \mid \boldsymbol{x} \in S\}$.

Theorem

Let $C \subseteq \mathbb{R}^n$ be a nonempty convex set and let $\boldsymbol{x}_0 \in \partial C$. There exists a supporting hyperplane of C at $\boldsymbol{x}_0$.

Proof: Since $\boldsymbol{x}_0 \in \partial C$, there exists a sequence $(\boldsymbol{z}_n)$ of elements of $\mathbb{R}^n - C$ such that $\lim_{n\to\infty} \boldsymbol{z}_n = \boldsymbol{x}_0$.
For each $\boldsymbol{z}_n$ there exists $\boldsymbol{w}_n$ such that $\boldsymbol{w}_n'\boldsymbol{z}_n > a$ and $\boldsymbol{w}_n'\boldsymbol{x} \leqslant a$ for $\boldsymbol{x} \in C$.
Without loss of generality we may assume that $\| \boldsymbol{w}_n \| = 1$. Since the sequence $(\boldsymbol{w}_n)$ is bounded, it contains a convergent subsequence $(\boldsymbol{w}_{i_m})$ such that $\lim_{m\to\infty} \boldsymbol{w}_{i_m} = \boldsymbol{w}$.
For this subsequence we have $\boldsymbol{w}'\boldsymbol{z}_{i_m} > a$ and $\boldsymbol{w}'\boldsymbol{x} \leqslant a$. Taking $m \to \infty$ we obtain $\boldsymbol{w}'\boldsymbol{x}_0 > a$ and $\boldsymbol{w}'\boldsymbol{x} \leqslant a$ for all $\boldsymbol{x} \in C$, which means that $H_{\boldsymbol{w},a}$ is a support plane of C at $\boldsymbol{x}_0$.

Theorem

Let S, T be two non-empty convex subsets of $\mathbb{R}^n$ that are disjoint. There exists $\boldsymbol{w} \in \mathbb{R}^n - \{\mathbf{0}_n\}$ such that

$$\inf\{\boldsymbol{w}'\boldsymbol{s} \mid \boldsymbol{s} \in S\} \geqslant \sup\{\boldsymbol{w}'\boldsymbol{t} \mid \boldsymbol{t} \in T\}.$$

Proof: It is easy to see that the set $S - T$ defined by

$$S - T = \{\boldsymbol{s} - \boldsymbol{t} \mid \boldsymbol{s} \in S \text{ and } \boldsymbol{t} \in T\}$$

is convex. Furthermore $\mathbf{0}_n \notin S - T$ because the sets S and T are disjoint. Thus, there exists in $S - T$ a proximal point $\boldsymbol{w}$ to $\mathbf{0}_n$, for which we have $(\boldsymbol{x} - \boldsymbol{w})'\boldsymbol{w} \geqslant 0$ for every $\boldsymbol{x} \in S - T$, that is, $(\boldsymbol{s} - \boldsymbol{t} - \boldsymbol{w})'\boldsymbol{w} \geqslant 0$, which is equivalent to

$$\boldsymbol{s}'\boldsymbol{w} \geqslant \boldsymbol{t}'\boldsymbol{w} + \parallel \boldsymbol{w} \parallel^2$$

for $\boldsymbol{s} \in S$ and $\boldsymbol{t} \in T$. This implies the inequality of the theorem.

Corollary

For any two non-empty convex subsets that are disjoint, there exists a non-zero vector $\boldsymbol{w} \in \mathbb{R}^n$ *such that*

$$\inf\{\boldsymbol{w}'\boldsymbol{s} \mid \boldsymbol{s} \in S\} \geqslant \sup\{\boldsymbol{w}'\boldsymbol{t} \mid \boldsymbol{t} \in \boldsymbol{K}(T)\}.$$

Notation

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we write

$$\boldsymbol{x} > \boldsymbol{y}$$

if $x_i > y_i$ for $1 \leqslant i \leqslant n$,

$$\boldsymbol{x} \geqslant \boldsymbol{y}$$

if $x_i \geqslant y_i$ for $1 \leqslant i \leqslant n$, and

$$\boldsymbol{x} \geq \boldsymbol{y}$$

if $x_i \geqslant y_i$ for $1 \leqslant i \leqslant n$ and at least of these inequalities is strict, that is, there exists i such that $x_i > y_i$.

Separation results have two important consequences for optimization theory, namely Farkas' and Gordan's alternative theorems.

Theorem

(Farkas' Alternative Theorem) *Let* $A \in \mathbb{R}^{m \times n}$ *and let* $\boldsymbol{c} \in \mathbb{R}^n$. *Exactly one of the following linear systems has a solution:*

(i) $A\boldsymbol{x} \leqslant \mathbf{0}_m$ *and* $\boldsymbol{c}'\boldsymbol{x} > 0$;

(ii) $A'\boldsymbol{y} = \boldsymbol{c}$ *and* $\boldsymbol{y} \geqslant \mathbf{0}_m$.

Proof

First System	Second System
$A\boldsymbol{x} \leqslant \boldsymbol{0}_m$	$A'\boldsymbol{y} = \boldsymbol{c}$
$\boldsymbol{c}'\boldsymbol{x} > 0$	$\boldsymbol{y} \geqslant \boldsymbol{0}_m$

If the second system has a solution, then $A'\boldsymbol{y} = \boldsymbol{c}$ and $\boldsymbol{y} \geqslant \boldsymbol{0}_m$ for some $\boldsymbol{y} \in \mathbb{R}^m$. Suppose that $\boldsymbol{x}$ is a solution of the first system. Then, $\boldsymbol{c}'\boldsymbol{x} = \boldsymbol{y}'A\boldsymbol{x} \leqslant 0$, which contradicts the inequality $\boldsymbol{c}'\boldsymbol{x} > 0$. Thus, if the second system has a solution, the first system has no solution.

First System	Second System
$A\boldsymbol{x} \leqslant \boldsymbol{0}_m$	$A'\boldsymbol{y} = \boldsymbol{c}$
$\boldsymbol{c}'\boldsymbol{x} > 0$	$\boldsymbol{y} \geqslant \boldsymbol{0}_m$

Suppose now that the second system has no solution. Note that the set $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = A'\boldsymbol{y}, \boldsymbol{y} \geqslant \boldsymbol{0}_m\}$ is a closed convex set. Furthermore, $\boldsymbol{c} \not\in S$ because, otherwise, $\boldsymbol{c}$ would be a solution of the second system. Thus, there exists $\boldsymbol{w} \in \mathbb{R}^n - \{\boldsymbol{0}_n\}$ and $a \in R$ such that $\boldsymbol{w}'\boldsymbol{c} > a$ and $\boldsymbol{w}'\boldsymbol{x} \leqslant a$ for $\boldsymbol{x} \in S$. In particular, since $\boldsymbol{0}_n \in S$ we have $a \geqslant 0$ and, therefore, $\boldsymbol{w}'\boldsymbol{c} > 0$. Also, for $\boldsymbol{y} \geqslant \boldsymbol{0}_m$ we have $a \geqslant \boldsymbol{w}'A'\boldsymbol{y} = \boldsymbol{y}'A\boldsymbol{w}$. Since $\boldsymbol{y}$ can be made arbitrarily large we must have $A\boldsymbol{w} \leqslant \boldsymbol{0}_m$. Then $\boldsymbol{w}$ is a solution of the first system.

Theorem

(Gordan's Alternative Theorem) *Let $A \in \mathbb{R}^{m \times n}$ be a matrix. Exactly one of the following linear systems has a solution:*

- $A\mathbf{x} < \mathbf{0}_m$ *for* $\mathbf{x} \in \mathbb{R}^n$;
- $A'\mathbf{y} = \mathbf{0}_n$ *and* $\mathbf{y} \geq \mathbf{0}_m$ *for* $\mathbf{y} \in \mathbb{R}^m$.

Proof

First System	Second System
$A\mathbf{x} < \mathbf{0}_m$	$A'\mathbf{y} = \mathbf{0}_n$
	$\mathbf{y} \geq \mathbf{0}_m$

Let A be a matrix such that the first system, $A\mathbf{x} < \mathbf{0}_m$ has a solution $\mathbf{x}_0$. Suppose that a solution $\mathbf{y}_0$ of the second system exists. Since $A\mathbf{x}_0 < \mathbf{0}_m$ and $\mathbf{y}_0 \geq \mathbf{0}_m$ (which implies that at least one component of $\mathbf{y}_0$ is positive) it follows that $\mathbf{y}_0' A\mathbf{x}_0 < 0$, which is equivalent to $\mathbf{x}_0' A'\mathbf{y} < 0$. This contradicts the assumption that $A'\mathbf{y} = \mathbf{0}_n$. Thus, the second system cannot have a solution if the first has one.

Proof (cont'd)

First System	Second System
$A\mathbf{x} < \mathbf{0}_m$	$A'\mathbf{y} = \mathbf{0}_n$
	$\mathbf{y} \geq \mathbf{0}_m$

Suppose now that the first system has no solution and consider the non-empty convex subsets S, T of $\mathbb{R}^m$ defined by

$$S = \{\mathbf{s} \in \mathbb{R}^m \mid \mathbf{s} = A\mathbf{x}, \mathbf{x} \in \mathbb{R}^n\} \text{ and } T = \{\mathbf{t} \in \mathbb{R}^m \mid \mathbf{t} < \mathbf{0}_m\}.$$

These sets are disjoint by the previous supposition. Then, there exists $\mathbf{w} \neq \mathbf{0}_m$ such that $\mathbf{w}'A\mathbf{s} \geqslant \mathbf{w}'\mathbf{t}$ for $\mathbf{s} \in S$ and $\mathbf{t} \in \mathbf{K}(T)$. This implies that $\mathbf{w} \geqslant \mathbf{0}_m$ because otherwise the components of $\mathbf{t}$ that correspond to a negative component of $\mathbf{w}$ could be made arbitrarily negative (and large in absolute value) and this would contradict the above inequality. Thus, $\mathbf{w} \geq \mathbf{0}_m$.

Since $\mathbf{0}_m \in \mathbf{K}(T)$, we also have $\mathbf{w}'A\mathbf{s} \geqslant 0$ for every $\mathbf{s} \in \mathbb{R}^m$. In particular, for $\mathbf{s} = -A'\mathbf{w}$ we obtain $\mathbf{w}'A(-A'\mathbf{w}) = - \parallel A'\mathbf{w} \parallel^2 = 0$, so $A'\mathbf{w} = \mathbf{0}_m$, which means that the second system has a solution.