Convex Optimization and Applications

2 - Convex geometry

Guillaume Sagnol

Technische Universität Berlin

Outline

Scalars, vectors, matrices

- $[n] := \{1, \ldots, n\}$

Scalars, vectors, matrices

- $[n] := \{1, \ldots, n\}$
- Scalar: plain lower case, e.g. $c \in \mathbb{R}$.

Scalars, vectors, matrices

- $[n] := \{1, \ldots, n\}$
- Scalar: plain lower case, e.g. $c \in \mathbb{R}$.
- Vectors: boldface lower case letters, e.g. $\boldsymbol{v} \in \mathbb{R}^n$, with elements $v_1, \ldots, v_n$.

Scalars, vectors, matrices

- $[n] := \{1, \ldots, n\}$
- Scalar: plain lower case, e.g. $c \in \mathbb{R}$.
- Vectors: boldface lower case letters, e.g. $\boldsymbol{v} \in \mathbb{R}^n$, with elements $v_1, \ldots, v_n$.
- Matrices: upper case letters, e.g. $A \in \mathbb{R}^{m \times n}$, with elements A_{ij} $(i \in [m],\ j \in [n])$.

Scalars, vectors, matrices

- $[n] := \{1, \ldots, n\}$
- Scalar: plain lower case, e.g. $c \in \mathbb{R}$.
- Vectors: boldface lower case letters, e.g. $\boldsymbol{v} \in \mathbb{R}^n$, with elements $v_1, \ldots, v_n$.
- Matrices: upper case letters, e.g. $A \in \mathbb{R}^{m \times n}$, with elements A_{ij} $(i \in [m],\ j \in [n])$.
- Column decomposition of a matrix:

$$A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] \in \mathbb{R}^{m \times n},$$

means that $\boldsymbol{a}_j \in \mathbb{R}^m$ is the jth column of A.

Scalars, vectors, matrices

- $[n] := \{1, \ldots, n\}$
- Scalar: plain lower case, e.g. $c \in \mathbb{R}$.
- Vectors: boldface lower case letters, e.g. $\boldsymbol{v} \in \mathbb{R}^n$, with elements $v_1, \ldots, v_n$.
- Matrices: upper case letters, e.g. $A \in \mathbb{R}^{m \times n}$, with elements A_{ij} ($i \in [m],\ j \in [n]$).
- Column decomposition of a matrix:

$$A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] \in \mathbb{R}^{m \times n},$$

  means that $\boldsymbol{a}_j \in \mathbb{R}^m$ is the jth column of A.
- Similarly,

$$A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^T \in \mathbb{R}^{m \times n}$$

  means that $\boldsymbol{a}_i^T$ is the ith row of A (with $\boldsymbol{a}_i \in \mathbb{R}^n$).

Vector notation

- $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$.
- $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$.

- $\mathbb{S}^n = \{X \in \mathbb{R}^{n \times n} : X = X^T\}$.
- $\mathbb{S}^n_+ = \{X \in \mathbb{S}^n :$ X is positive semidefinite$\}$.
- $\mathbb{S}^n_{++} = \{X \in \mathbb{S}^n :$ X is positive definite$\}$.

- Elementwise inequalities: $\boldsymbol{x} \leq \boldsymbol{y}$ means $x_i \leq y_i, \forall i$

Example

If $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m]^T$, then $A\boldsymbol{x} \leq \boldsymbol{b}$ means

$$\boldsymbol{a}_i^T \boldsymbol{x} \leq b_i \quad (\forall i \in [m]).$$

Vector notation

- $\boldsymbol{e}_i = i$th standard unit vector $[0, \dots, 0, 1, 0, \dots, 0]^T$
- $\mathbf{1}$ or $\mathbf{1}_n$= all-ones vector $[1, \dots, 1]^T$ (on blackboard: $\mathbb{1}$)
- Identity matrix I or I_n
- All-ones matrix $J_n = \mathbf{1}_n \mathbf{1}_n^T$
- $\mathbf{Diag}\,(\boldsymbol{u}) = \begin{pmatrix} u_1 & & \\ & \ddots & \\ & & u_n \end{pmatrix}$, $\mathbf{diag}\,(M) = [M_{11}, \dots, M_{nn}]^T$

Example

For $\boldsymbol{v} \in \mathbb{R}^n$ and $M \in \mathbb{R}^{m \times n}$ it holds:

- $v_i = \boldsymbol{e}_i^T \boldsymbol{v}$
- $M_{ij} = \boldsymbol{e}_i^T M \boldsymbol{e}_j$

Scalar products and norms

- For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$,

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle := \boldsymbol{u}^T \boldsymbol{v} = \sum_i u_i v_i$$

- For $A, B \in \mathbb{R}^{m \times n}$,

$$\langle A, B \rangle := \textbf{trace}\, A^T B = \sum_{i,j} A_{ij} B_{ij}$$

- In particular, if $A, B \in \mathbb{S}^n$, it holds $\langle A, B \rangle = \textbf{trace}\, AB$.

Example

- $\langle \mathbf{1}, \boldsymbol{v} \rangle = \mathbf{1}^T \boldsymbol{v}$ is the sum of all entries of $\boldsymbol{v}$
- $\langle J, M \rangle$ is the sum of all entries of M
- $\langle I, M \rangle$ is the trace of M

Scalar products and norms

- Euclidean norm $\|\boldsymbol{v}\| := \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$
- Frobenius norm of a matrix:

$$\|A\|_F := \sqrt{\langle A, A \rangle} = \Big(\sum_{i,j} A_{ij}^2\Big)^{1/2}.$$

- The *vectorization* of $A = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n] \in \mathbb{R}^{m \times n}$ is

$$\mathbf{vec}(A) := \begin{bmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} \in \mathbb{R}^{mn}.$$

Example

- $\langle A, B \rangle = \langle \mathbf{vec}(A), \mathbf{vec}(B) \rangle$
- $\|A\|_F = \|\mathbf{vec}(A)\|$

Affine functions

- Affine functions mapping $\mathbb{R}^n \to \mathbb{R}$ have the form
$$\boldsymbol{x} \mapsto \boldsymbol{a}^T \boldsymbol{x} + b.$$
- More generally, an affine function mapping $\mathbb{R}^n \to \mathbb{R}^m$ has the form
$$f : \boldsymbol{x} \mapsto A\boldsymbol{x} + \boldsymbol{b}.$$
for some matrix $A \in \mathbb{R}^{m \times n}$ and a vector $\boldsymbol{b} \in \mathbb{R}^m$.

Affine functions

- Affine functions mapping $\mathbb{R}^n \to \mathbb{R}$ have the form

$$\boldsymbol{x} \mapsto \boldsymbol{a}^T \boldsymbol{x} + b.$$

- More generally, an affine function mapping $\mathbb{R}^n \to \mathbb{R}^m$ has the form

$$f : \boldsymbol{x} \mapsto A\boldsymbol{x} + \boldsymbol{b}.$$

  for some matrix $A \in \mathbb{R}^{m \times n}$ and a vector $\boldsymbol{b} \in \mathbb{R}^m$.

- f *linear* usually means $\boldsymbol{b} = \mathbf{0}$, but we abuse the language, i.e. "*linear*$\simeq$*affine*"...
- To emphasize that $\boldsymbol{b} = \mathbf{0}$, we say that f is a *linear form*

Quadratic functions

- Quadratic functions mapping $\mathbb{R}^n \to \mathbb{R}$ have the form

$$x \mapsto x^T Q x + a^T x + b.$$

- A *quadratic form* is a quadratic function without linear part, i.e., $a = \mathbf{0}$, $b = 0$.

Quadratic functions

- Quadratic functions mapping $\mathbb{R}^n \to \mathbb{R}$ have the form

$$x \mapsto x^T Q x + a^T x + b.$$

- A *quadratic form* is a quadratic function without linear part, i.e., $a = \mathbf{0}, b = 0$.

- *Homogenization*: every quadratic function is a quadratic form over $\mathbb{R}^n \times \{1\}$:

$$x^T Q x + a^T x + b = \begin{bmatrix} x \\ 1 \end{bmatrix}^T \begin{pmatrix} Q & \frac{1}{2}a \\ \frac{1}{2}a^T & b \end{pmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}.$$

Gradient and Hessian

- The gradients and hessian of (sufficiently) differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ are

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{bmatrix} \in \mathbb{R}^n, \ \nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\boldsymbol{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\boldsymbol{x}) \end{pmatrix} \in \mathbb{S}^n.$$

Example

- $\nabla(x \mapsto \boldsymbol{a}^T x) = \boldsymbol{a}$
- $\nabla^2(x \mapsto \boldsymbol{a}^T x) = 0 \in \mathbb{S}^n$
- $\nabla(x \mapsto \frac{1}{2} x^T Q x) = Q x$
- $\nabla^2(x \mapsto \frac{1}{2} x^T Q x) = Q.$

Expressing a quadratic form...

as a linear function of the associated matrix

Lemma (aka the trace-trick).

The function $f : \mathbb{S}^n \to \mathbb{R},\ X \mapsto \boldsymbol{u}^T X \boldsymbol{u}$ is a linear function of X. Indeed,

$$\boldsymbol{u}^T X \boldsymbol{u} = \langle X, \boldsymbol{u}\boldsymbol{u}^T \rangle.$$

proof. Recall that $\textbf{trace}\, AB = \textbf{trace}\, BA$ (trace is *invariant to cyclic permutations*).

$$\begin{aligned}
\boldsymbol{u}^T X \boldsymbol{u} &= \textbf{trace}\, \boldsymbol{u}^T X \boldsymbol{u} && \text{(seen as a } 1 \times 1\text{-matrix)} \\
&= \textbf{trace}\, X \boldsymbol{u}\boldsymbol{u}^T && \text{(cyclic permutation)} \\
&= \langle X, \boldsymbol{u}\boldsymbol{u}^T \rangle && \text{(note that } \boldsymbol{u}\boldsymbol{u}^T \text{ is an } m \times n\text{-matrix)}
\end{aligned}$$

Outline

Lines, segments, rays

Definition (Lines, segments, rays).

Let $x_1, x_2 \in \mathbb{R}^n$.

- The *line* through x_1 and x_2 is

$$\{\theta x_1 + (1-\theta)x_2 : \ \theta \in \mathbb{R}\}.$$

- The *segment* between x_1 and x_2 is

$$\{\theta x_1 + (1-\theta)x_2 : \ \theta \in [0,1]\}.$$

- The *ray* through x_1 is

$$\{\ \theta x_1 : \theta \geq 0\}.$$

Affine, Convex, Conic

Definition (Affine, Convex, and Conic sets).

Let S be a subset of $\mathbb{R}^n$.

- S is *affine* if contains all lines joining points of S:

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in \mathbb{R} \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$

Affine, Convex, Conic

Definition (Affine, Convex, and Conic sets).

Let S be a subset of $\mathbb{R}^n$.

- S is *affine* if contains all lines joining points of S:

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in \mathbb{R} \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$

- S is *convex* if contains all segments joining points of S:

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in [0,1] \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$

Affine, Convex, Conic

Definition (Affine, Convex, and Conic sets).

Let S be a subset of $\mathbb{R}^n$.

- S is *affine* if contains all lines joining points of S:

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in \mathbb{R} \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$

- S is *convex* if contains all segments joining points of S:

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in [0,1] \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$

- S is a *cone* if contains the ray through any point of S:

$$\boldsymbol{x} \in S, \theta \geq 0 \implies \theta \boldsymbol{x} \in S.$$

Affine, Convex, Conic

Definition (Affine, Convex, and Conic sets).

Let S be a subset of $\mathbb{R}^n$.

- S is *affine* if contains all lines joining points of S:

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in \mathbb{R} \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$

- S is *convex* if contains all segments joining points of S:

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \theta \in [0,1] \implies \theta \boldsymbol{x}_1 + (1-\theta)\boldsymbol{x}_2 \in S.$$

- S is a *cone* if contains the ray through any point of S:

$$\boldsymbol{x} \in S, \theta \geq 0 \implies \theta \boldsymbol{x} \in S.$$

- S is a convex *cone* if:

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in S, \lambda_1, \lambda_2 \geq 0 \implies \lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 \in S.$$

Affine, Convex, Conic combinations

More generally, we can *combine* more than 2 points

Definition (Affine, Convex, Conic combinations).

Let $\boldsymbol{x}_i \in \mathbb{R}^n$ ($\forall i \in [k]$). The expression $\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ is called

- an *affine combination* of the $x_i's$ if $\sum_i \lambda_i = 1$.
- a *convex combination* of the $x_i's$ if $\sum_i \lambda_i = 1, \boldsymbol{\lambda} \geq \boldsymbol{0}$.
- a *conic combination* of the $x_i's$ if $\boldsymbol{\lambda} \geq \boldsymbol{0}$.

Affine, Convex, Conic combinations

More generally, we can *combine* more than 2 points

Definition (Affine, Convex, Conic combinations).

Let $\boldsymbol{x}_i \in \mathbb{R}^n$ $(\forall i \in [k])$. The expression $\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ is called

- an *affine combination* of the $x_i's$ if $\sum_i \lambda_i = 1$.
- a *convex combination* of the $x_i's$ if $\sum_i \lambda_i = 1, \boldsymbol{\lambda} \geq \mathbf{0}$.
- a *conic combination* of the $x_i's$ if $\boldsymbol{\lambda} \geq \mathbf{0}$.

Proposition

A set is affine/convex/a convex cone iff it is stable by affine/convex/conic combinations.

Affine, Convex, Conic hull

Definition (Affine, Convex, and Conic hull).

- The *vector space* spanned by $S \subseteq \mathbb{R}^n$ is:

$$\textbf{span}\, S = \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i : \; k \in \mathbb{N}, \quad \forall i \in [k], \boldsymbol{x}_i \in S, \quad \boldsymbol{\lambda} \in \mathbb{R}^k\}.$$

- The *affine hull* of S is:

$$\textbf{aff}\, S = \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i : \; k \in \mathbb{N}, \forall i \in [k], \boldsymbol{x}_i \in S, \; \boldsymbol{\lambda} \in \mathbb{R}^k, \; \mathbf{1}^T \boldsymbol{\lambda} = 1\}.$$

- The *convex hull* of S is:

$$\textbf{conv}\, S = \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i : \; k \in \mathbb{N}, \forall i \in [k], \boldsymbol{x}_i \in S, \; \boldsymbol{\lambda} \geq \mathbf{0}, \; \mathbf{1}^T \boldsymbol{\lambda} = 1\}.$$

- The *conic hull* of S is:

$$\textbf{cone}\, S = \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i : \; k \in \mathbb{N}, \quad \forall i \in [k], \boldsymbol{x}_i \in S, \quad \boldsymbol{\lambda} \geq \mathbf{0}\}.$$

Affine, Convex, Conic hull

The previous hull definitions coincinde with the intuitive meaning of *hull*:

Proposition

$$\textbf{aff}\ /\textbf{conv}\ /\textbf{cone}\, S = \bigcap_{\substack{T \supseteq S \\ T \text{affine/convex/convex cone}}} T$$

That is, the affine (convex, conic hull) of S is the smallest affine set (convex set, convex cone) that contains S.

Characterization of affine spaces

Affine spaces are *vector spaces plus a shift:*

Proposition

Let L be an affine space, and $x_0 \in L$. Then, $V = L - x_0$ is a vector space, and does not depend on the choice of x_0. Hence we can define $\dim L := \dim V$.

Using the fact that we can write $V = \mathbf{Im}\, A$ or $V = \mathbf{Ker}\, F$,

Proposition

L is an affine subspace of $\mathbb{R}^n$ of dimension $m \leq n$

$$\iff L = \{A\boldsymbol{y} + \boldsymbol{b} : \ \boldsymbol{y} \in \mathbb{R}^m\} \text{ for some } A \in \mathbb{R}^{n\times m}, \boldsymbol{b} \in \mathbb{R}^n.$$

$$\iff L = \{\boldsymbol{x} \in \mathbb{R}^n : F\boldsymbol{x} = \boldsymbol{g}\} \text{ for some } F \in \mathbb{R}^{m\times n}, \boldsymbol{g} \in \mathbb{R}^m\}.$$

Caratheodory theorem

Recall the definition of a convex hull:

$$\mathbf{conv}\, S = \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i : \ k \in \mathbb{N},\ \forall i \in [k], \boldsymbol{x}_i \in S,\ \boldsymbol{\lambda} \geq \mathbf{0},\ \mathbf{1}^T \boldsymbol{\lambda} = 1\}.$$

$$= \left\{ \boldsymbol{x} \in \mathbb{R}^n : \ \begin{array}{l} \boldsymbol{x} \text{ is convex combination of } k \text{ elements} \\ \text{of S, for some } k \in \mathbb{N} \end{array} \right\}.$$

Can we bound the number k of elements of S we need to combine to get any elements of $\mathbf{conv}\, S$?

Caratheodory theorem

Recall the definition of a convex hull:

$$\textbf{conv}\, S = \{\sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i : \; k \in \mathbb{N},\ \forall i \in [k], \boldsymbol{x}_i \in S,\ \boldsymbol{\lambda} \geq \mathbf{0},\ \mathbf{1}^T \boldsymbol{\lambda} = 1\}.$$

$$= \left\{ \boldsymbol{x} \in \mathbb{R}^n : \; \begin{array}{l} \boldsymbol{x} \text{ is convex combination of } k \text{ elements} \\ \text{of S, for some } k \in \mathbb{N} \end{array} \right\}.$$

Can we bound the number k of elements of S we need to combine to get any elements of $\textbf{conv}\, S$?

Theorem (Caratheodory).

Let $S \subseteq \mathbb{R}^n$ be of affine dimension $m := \dim \textbf{aff}\, S \leq n$, and $\boldsymbol{x} \in \textbf{conv}\, S$. Then, $\boldsymbol{x}$ can be expressed as a convex combination of $k \leq m+1$ points of S.

Caratheodory theorem

Theorem (Caratheodory).

Let $S \subseteq \mathbb{R}^n$ be of affine dimension $m := \dim \mathbf{aff}\ S \leq n$, and $x \in \mathbf{conv}\ S$. Then, x can be expressed as a convex combination of $k \leq m + 1$ points of S.

There is also an analog result for conic hulls:

Theorem (Caratheodory – conic version).

Let $S \subseteq \mathbb{R}^n$, such that $\dim \mathbf{span}\ S = m \leq n$, and let $x \in \mathbf{cone}\ S$. Then, x can be expressed as a conic combination of $k \leq m \leq n$ points of S.

Outline

Simple convex sets

1. Hyperplane: $\{x : \boldsymbol{a}^T x = b\}$ (affine, convex)

Simple convex sets

1. Hyperplane: $\{x : a^T x = b\}$ (affine, convex)
2. Halfspace: $\{x : a^T x \leq b\}$ (convex)

Simple convex sets

1. Hyperplane: $\{x : a^T x = b\}$ (affine, convex)
2. Halfspace: $\{x : a^T x \leq b\}$ (convex)
3. Polytope: $\textbf{conv}\,\{x_1, \ldots, x_k\}$ (convex)

Simple convex sets

1. Hyperplane: $\{x : \boldsymbol{a}^T \boldsymbol{x} = b\}$ (affine, convex)
2. Halfspace: $\{x : \boldsymbol{a}^T \boldsymbol{x} \leq b\}$ (convex)
3. Polytope: $\textbf{conv}\,\{x_1, \ldots, x_k\}$ (convex)
4. Polyhedron: $\{x : Ax \leq \boldsymbol{b}\}$ (convex)

Simple convex sets

1. Hyperplane: $\{x : \boldsymbol{a}^T x = b\}$ (affine, convex)
2. Halfspace: $\{x : \boldsymbol{a}^T x \leq b\}$ (convex)
3. Polytope: $\mathbf{conv}\,\{x_1, \ldots, x_k\}$ (convex)
4. Polyhedron: $\{x : Ax \leq \boldsymbol{b}\}$ (convex)
5. Ball: $\{x : \|x - x_0\| \leq r\}$ (convex, *for any norm*)

Simple convex sets

1. Hyperplane: $\{x : \boldsymbol{a}^T x = b\}$ (affine, convex)
2. Halfspace: $\{x : \boldsymbol{a}^T x \leq b\}$ (convex)
3. Polytope: $\mathbf{conv}\,\{x_1, \ldots, x_k\}$ (convex)
4. Polyhedron: $\{x : Ax \leq \boldsymbol{b}\}$ (convex)
5. Ball: $\{x : \|x - x_0\| \leq r\}$ (convex, *for any norm*)
6. Norm cone: $\{(x, t) : \|x - x_0\| \leq t\}$ (convex, *for any norm*)

Simple convex sets

1. Hyperplane: $\{x : \boldsymbol{a}^T \boldsymbol{x} = b\}$ (affine, convex)
2. Halfspace: $\{\boldsymbol{x} : \boldsymbol{a}^T \boldsymbol{x} \leq b\}$ (convex)
3. Polytope: **conv** $\{x_1, \ldots, x_k\}$ (convex)
4. Polyhedron: $\{\boldsymbol{x} : A\boldsymbol{x} \leq \boldsymbol{b}\}$ (convex)
5. Ball: $\{\boldsymbol{x} : \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq r\}$ (convex, *for any norm*)
6. Norm cone: $\{(\boldsymbol{x}, t) : \|\boldsymbol{x} - \boldsymbol{x}_0\| \leq t\}$ (convex, *for any norm*)
7. Unit simplex: $\Delta_n := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x} \geq \mathbf{0},\ \mathbf{1}^T \boldsymbol{x} \leq 1\}$

Simple convex sets

1. Hyperplane: $\{x : a^T x = b\}$ (affine, convex)
2. Halfspace: $\{x : a^T x \leq b\}$ (convex)
3. Polytope: $\textbf{conv}\,\{x_1, \ldots, x_k\}$ (convex)
4. Polyhedron: $\{x : Ax \leq \boldsymbol{b}\}$ (convex)
5. Ball: $\{x : \|x - x_0\| \leq r\}$ (convex, *for any norm*)
6. Norm cone: $\{(x, t) : \|x - x_0\| \leq t\}$ (convex, *for any norm*)
7. Unit simplex: $\Delta_n := \{x \in \mathbb{R}^n : x \geq \mathbf{0},\ \mathbf{1}^T x \leq 1\}$
8. Probability simplex: $\Delta_n^= := \{x \in \mathbb{R}^n : x \geq \mathbf{0},\ \mathbf{1}^T x = 1\}$

Simple convex sets

1. Hyperplane: $\{x : a^T x = b\}$ (affine, convex)
2. Halfspace: $\{x : a^T x \leq b\}$ (convex)
3. Polytope: $\mathbf{conv}\,\{x_1, \ldots, x_k\}$ (convex)
4. Polyhedron: $\{x : Ax \leq \boldsymbol{b}\}$ (convex)
5. Ball: $\{x : \|x - x_0\| \leq r\}$ (convex, *for any norm*)
6. Norm cone: $\{(x, t) : \|x - x_0\| \leq t\}$ (convex, *for any norm*)
7. Unit simplex: $\Delta_n := \{x \in \mathbb{R}^n : x \geq \mathbf{0},\ \mathbf{1}^T x \leq 1\}$
8. Probability simplex: $\Delta_n^= := \{x \in \mathbb{R}^n : x \geq \mathbf{0},\ \mathbf{1}^T x = 1\}$
9. Nonnegative orthant: $\mathbb{R}_+^n$ (convex cone)

Simple convex sets

1. Hyperplane: $\{x : a^T x = b\}$ (affine, convex)
2. Halfspace: $\{x : a^T x \leq b\}$ (convex)
3. Polytope: $\mathbf{conv}\,\{x_1, \ldots, x_k\}$ (convex)
4. Polyhedron: $\{x : Ax \leq \boldsymbol{b}\}$ (convex)
5. Ball: $\{x : \|x - x_0\| \leq r\}$ (convex, *for any norm*)
6. Norm cone: $\{(x, t) : \|x - x_0\| \leq t\}$ (convex, *for any norm*)
7. Unit simplex: $\Delta_n := \{x \in \mathbb{R}^n : x \geq \mathbf{0},\ \mathbf{1}^T x \leq 1\}$
8. Probability simplex: $\Delta_n^= := \{x \in \mathbb{R}^n : x \geq \mathbf{0},\ \mathbf{1}^T x = 1\}$
9. Nonnegative orthant: $\mathbb{R}_+^n$ (convex cone)
10. Symmetric matrices: $\mathbb{S}^n$ (vector space of dim. $\frac{1}{2}n(n+1)$)

Operations that preserve convexity

Let S, T be convex sets. Then, the following sets are convex:

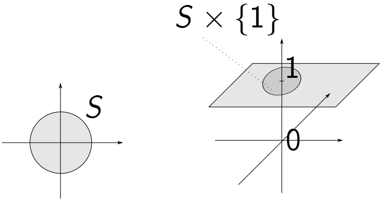
1. $S \cap T$ (also valid for intersection of infinite families)
2. $S \times T$ (cartesian product)
3. $\{A\boldsymbol{x} + \boldsymbol{b} : \boldsymbol{x} \in S\}$ (affine transformation of S)
   - ρS (scaling)
   - $S + \boldsymbol{b}$ (translation)
   - $\{(x_1, \ldots, x_k) : \boldsymbol{x} \in S\}$ (projection over some coordinates)
4. $S + T$ (Minkowski sum)
5. $\{\boldsymbol{x} : A\boldsymbol{x} + \boldsymbol{b} \in S\}$ (Reverse affine transformation)
6. **cl** S and **int** S (Closure and Interior)

Perspective transformation

Define the *perspective function*

$$P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n, \qquad (\boldsymbol{x}, t) \mapsto \frac{\boldsymbol{x}}{t}.$$

7 If $S \subseteq \mathbb{R}^n$ is convex, then
$P^{-1}(S) := \{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{++} : \ \frac{1}{t}\boldsymbol{x} \in S\}$ is convex.
Its closure is cl $P^{-1}(S) =$ **cone** $(S \times \{1\})$.



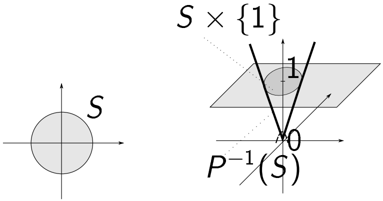
$S = \{\boldsymbol{x} : \|\boldsymbol{x}\| \leq 1\}$ (unit ball)

Perspective transformation

Define the *perspective function*

$$P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n, \qquad (\boldsymbol{x}, t) \mapsto \frac{\boldsymbol{x}}{t}.$$

7 If $S \subseteq \mathbb{R}^n$ is convex, then
$P^{-1}(S) := \{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{++} : \frac{1}{t}\boldsymbol{x} \in S\}$ is convex.
Its closure is $\mathbf{cl}\, P^{-1}(S) = \mathbf{cone}\,(S \times \{1\})$.



$S = \{\boldsymbol{x} : \|\boldsymbol{x}\| \leq 1\}$ (unit ball)

$P^{-1}(S) = \{(\boldsymbol{x}, t) : \|\boldsymbol{x}\| \leq t, t > 0\}$

$\mathbf{cl}\, P^{-1}(S) = \{(\boldsymbol{x}, t) : \|\boldsymbol{x}\| \leq t\}$ (Lorentz cone)

Positive semidefinite matrices

Proposition / Definition

Let $X \in \mathbb{S}^n$. The following statements are equivalent:

1. $X \in \mathbb{S}^n_+$ (S is positive semidefinite)
2. $\forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{u}^T X \boldsymbol{u} \geq 0.$
3. All eigenvalues of X are nonnegative.
4. $\exists H \in \mathbb{R}^{n \times m}, m \in \mathbb{N} : \ X = HH^T$
5. $X \in \textbf{conv}\,\{xx^T : x \in \mathbb{R}^n\} = \textbf{cone}\,\{xx^T : x \in \mathbb{R}^n\}.$

In particular, $\mathbb{S}^n_+$ is a convex cone.

Positive semidefinite matrices

Proposition / Definition

Let $X \in \mathbb{S}^n$. The following statements are equivalent:

1. $X \in \mathbb{S}^n_+$ (S is positive semidefinite)
2. $\forall \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{u}^T X \boldsymbol{u} \geq 0.$
3. All eigenvalues of X are nonnegative.
4. $\exists H \in \mathbb{R}^{n \times m}, m \in \mathbb{N} : \; X = HH^T$
5. $X \in \textbf{conv}\,\{\boldsymbol{x}\boldsymbol{x}^T : \boldsymbol{x} \in \mathbb{R}^n\} = \textbf{cone}\,\{\boldsymbol{x}\boldsymbol{x}^T : \boldsymbol{x} \in \mathbb{R}^n\}.$

In particular, $\mathbb{S}^n_+$ is a convex cone.

Other, direct proof of the convexity of $\mathbb{S}^n_+$:

$$\mathbb{S}^n_+ = \{X : \boldsymbol{u}^T X \boldsymbol{u} \geq 0, \forall \boldsymbol{u} \in \mathbb{R}^n\} = \{X : \langle X, \boldsymbol{u}\boldsymbol{u}^T \rangle \geq 0, \forall \boldsymbol{u} \in \mathbb{R}^n\}$$
$$= \bigcap_{\boldsymbol{u} \in \mathbb{R}^n} \{X \in \mathbb{S}^n : \langle X, \boldsymbol{u}\boldsymbol{u}^T \rangle \geq 0\}$$

Positive definite matrices

The interior of $\mathbb{S}^n_+$ is also a cone:

Proposition / Definition

Let $X \in \mathbb{S}^n$. The following statements are equivalent:

1. $X \in \mathbb{S}^n_{++}$ (S is positive definite)
2. $X \in \textbf{int}\, \mathbb{S}^n_+$
3. $\forall \boldsymbol{u} \in \mathbb{R}^n, \quad \boldsymbol{u} \neq \boldsymbol{0} \implies \boldsymbol{u}^T X \boldsymbol{u} > 0.$
4. All eigenvalues of X are positive.
5. $\exists H$ invertible such that $X = HH^T$.
6. Sylvester criterion: All leading principal minors of X are positive.

Properties of p.s.d. matrices

Lemma

Let $X \in \mathbb{S}_+^n$. Then,

1. The matrix AXA^T is positive semidefinite (for all A of appropriate size).
2. If I is a subset of $[n]$, the principal submatrix

$$X[I,I] = \{X_{i_1,i_2}\}_{i_1 \in I, i_2 \in I}$$

   is positive semidefinite.
3. For all $i,j \in [n]$, $|X_{ij}| \leq \sqrt{X_{ii}X_{jj}}$.
4. $X_{ii} = 0 \implies \forall j \in [n], X_{ij} = 0$.

Matrix decompositions

Proposition (Matrix square root).

Let $X \in \mathbb{S}_+^n$. Then, X has a square root, which we denote by $X^{\frac{1}{2}} \in \mathbb{S}_+^n$, and is the only positive semidefinite matrix that satisfies

$$X = \left(X^{\frac{1}{2}}\right)^2.$$

In particular, the eigenvalues of $X^{\frac{1}{2}}$ are the square roots of the eigenvalues of X.

Proposition (Cholesky decomposition).

$X \in \mathbb{S}_+^n$ admits a *Cholesky decomposition* of the form $X = LL^T$, where L is lower triangular.

If X is positive definite, then this decomposition is unique.

Ellipsoids

Definition (Ellipsoid)

An ellipsoid of $\mathbb{R}^n$ is a set of the form

$$\mathcal{E} = \{x \in \mathbb{R}^n : (x - x_0)^T Q^{-1}(x - x_0) \leq 1\},$$

where $x_0 \in \mathbb{R}^n$ and the matrix Q is positive definite.

All ellipsoids can be obtained as the **affine transformation** (or reverse image by some affine transformation) of a **unit ball**. Indeed,

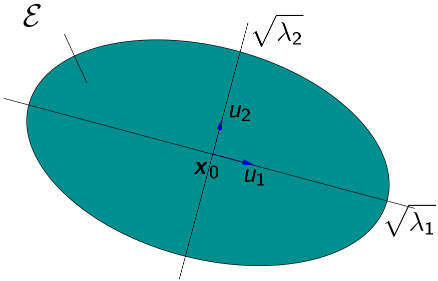
$$\begin{aligned}\mathcal{E} &= \{x \in \mathbb{R}^n : \|Q^{-1/2}x - Q^{-1/2}x_0\| \leq 1\} \\ &= \{Q^{1/2}y + x_0 : \|y\| \leq 1\}\end{aligned}$$

Ellipsoids vs. eigenvalue decomposition

$$\mathcal{E} = \{x \in \mathbb{R}^n : (x - x_0)^T Q^{-1} (x - x_0) \leq 1\}$$

Consider eigendecomposition $Q = U \Lambda U^T = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^T$.

Then, $\mathcal{E}$ is an elliposid centered at x_0, with semiaxis of length $\sqrt{\lambda_i}$ along $\boldsymbol{u}_i$.



Outline

Proper cone

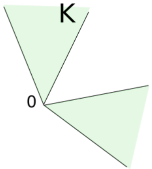
Definition (Proper cone).

A cone $K \subset \mathbb{R}^n$ is said to be *proper* if it is

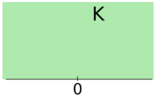
- closed;
- convex;
- pointed, i.e., it contains no lines. More precisely,

$$(x \in K, -x \in K) \implies x = \mathbf{0};$$

- and it has a nonempty interior.



not convex

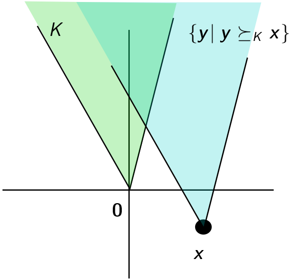


not pointed

Generalized conic inequality

- Given a proper cone K, we define a partial order $\preceq_K$:

$$x \preceq_K y \iff y - x \in K.$$
$$x \prec_K y \iff y - x \in \text{int}\, K.$$



Note: For matrices, $X \succeq Y$ means $X \succeq_{\mathbb{S}_+^n} Y$. In particular, $X \succeq 0$ means that X is positive semidefinite.

Properties of conic ordering

Proposition

Let K be a proper cone. The inequality $\preceq_K$ satisfies:

1. transitivity: $\boldsymbol{x} \preceq_K \boldsymbol{y}$ and $\boldsymbol{y} \preceq_K \boldsymbol{z} \implies \boldsymbol{x} \preceq_K \boldsymbol{z}$
2. reflexivity: $\boldsymbol{x} \preceq_K \boldsymbol{x}$.
3. antisymmetry: $\boldsymbol{x} \preceq_K \boldsymbol{y}$ and $\boldsymbol{y} \preceq_K \boldsymbol{x} \implies \boldsymbol{x} = \boldsymbol{y}$.
4. preservation under addition:
   $\boldsymbol{x} \preceq_K \boldsymbol{y}$ and $\boldsymbol{u} \preceq_K \boldsymbol{v} \implies \boldsymbol{x} + \boldsymbol{u} \preceq_K \boldsymbol{y} + \boldsymbol{v}$.
5. preservation under nonnegative scaling:
   $\boldsymbol{x} \preceq_K \boldsymbol{y}$ and $\alpha \geq 0 \implies \alpha \boldsymbol{x} \preceq_K \alpha \boldsymbol{y}$.

Note that $\preceq_K$ is a *partial order*, i.e.,

$$\boldsymbol{x} \not\preceq_K \boldsymbol{y} \quad \not\Longleftrightarrow \quad \boldsymbol{x} \succeq_K \boldsymbol{y}.$$

Examples

- $\preceq_{\mathbb{R}^n_+}$ is simply the standard elementwise inequality:

$$x \preceq_{\mathbb{R}^n_+} y \iff x \leq y.$$

Note that $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ are not comparable for $\preceq_{\mathbb{R}^n_+}$.

Examples

- $\preceq_{\mathbb{R}^n_+}$ is simply the standard elementwise inequality:

$$x \preceq_{\mathbb{R}^n_+} y \iff x \leq y.$$

  Note that $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ are not comparable for $\preceq_{\mathbb{R}^n_+}$.

- Let $K \subset \mathbb{R}^{d+1}$ be the cone of coefficients of polynomials of degree d that are nonnegative on $[0, 1]$:

$$K = \{\boldsymbol{\alpha} \in \mathbb{R}^{d+1} : \quad \forall x \in [0,1], \sum_{i=0}^{d} \alpha_i x^i \geq 0\}.$$

  Then,

$$\boldsymbol{\alpha} \preceq_K \boldsymbol{\beta} \iff \forall x \in [0,1], \quad \sum_{i=0}^{d} \alpha_i x^i \leq \sum_{i=0}^{d} \beta_i x^i$$

Dual cone

Definition (Dual cone).

The dual cone of K is

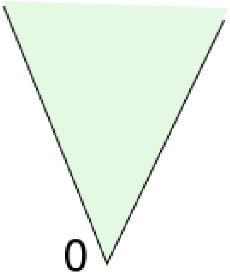
$$K^* = \{\boldsymbol{y} | \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq 0,\ \forall \boldsymbol{x} \in K\}.$$

Dual cone

Definition (Dual cone).

The dual cone of K is

$$K^* = \{\boldsymbol{y} | \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq 0, \ \forall \boldsymbol{x} \in K\}.$$

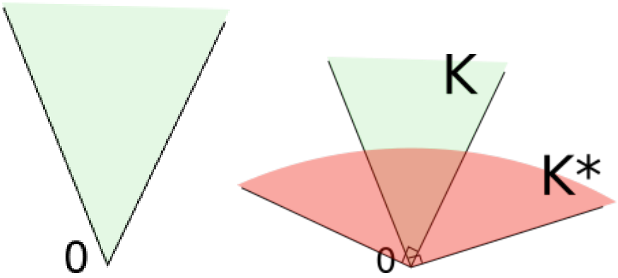


Dual cone

Definition (Dual cone).

The dual cone of K is

$$K^* = \{\boldsymbol{y} | \langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq 0, \ \forall \boldsymbol{x} \in K\}.$$



A fundamental result

Proposition

Let K be a cone. Then,

$$\inf_{\boldsymbol{x} \in K} \ \boldsymbol{c}^T \boldsymbol{x} = \begin{cases} 0 & \text{if } \boldsymbol{c} \in K^* \\ -\infty & \text{otherwise.} \end{cases}$$

Similarly,

$$\sup_{\boldsymbol{x} \in K} \ \boldsymbol{c}^T \boldsymbol{x} = \begin{cases} 0 & \text{if } \boldsymbol{c} \in -K^* \\ +\infty & \text{otherwise.} \end{cases}$$

Dual cone

Proposition (Properties of the dual cone).

Let K be a convex cone.

1. K^* is a convex cone.
2. K^* is closed (even if K is not).
3. $K_1 \subseteq K_2 \implies K_2^* \subseteq K_1^*$.
4. K has a nonempty interior $\implies K^*$ pointed.
5. $K^{**} = \mathbf{cl}\, K$ (so, in particular, K closed $\implies K = K^{**}$).
6. $\mathbf{cl}\, K$ is pointed $\implies K^*$ has a nonempty interior.

In particular,

$$K \text{ proper} \implies K^* \text{ proper}, \text{ and } K = (K^*)^*.$$

Outline

Separating hyperplane theorem

If two convex sets do not intersect, then they can be separated by some hyperplane:

Theorem (Separating hyperplane).

Let X, Y be two disjoint, nonempty convex sets of $\mathbb{R}^n$. Then, there $\exists c \in \mathbb{R}$, $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that

$$\forall \boldsymbol{x} \in X,\ \ \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq c \quad \text{and} \quad \forall \boldsymbol{y} \in Y,\ \ \langle \boldsymbol{y}, \boldsymbol{v} \rangle \geq c.$$

In other words, the hyperplane $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{v} \rangle = c\}$ *separates* X and Y.

Strict separation

When, in addition, both sets are closed and one of them is compact, it is possible to separate them *strictly*:

Theorem (Strict separating hyperplane).

Let X, Y be disjoint, nonempty, closed convex sets of $\mathbb{R}^n$. If X or Y is compact, then $\exists c \in \mathbb{R}$, $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that

$$\forall \boldsymbol{x} \in X, \ \ \langle \boldsymbol{x}, \boldsymbol{v} \rangle < c \quad \text{and} \quad \forall \boldsymbol{y} \in Y, \ \ \langle \boldsymbol{y}, \boldsymbol{v} \rangle > c.$$

In other words, the hyperplane $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{v} \rangle = c\}$ *strictly separates* X and Y.

Separation theorem for a cone

When one of the two sets is a cone, we can set $c = 0$:

Theorem (Separating hyperplane for a cone).

Let $C \subseteq \mathbb{R}^n$ be a nonempty convex cone, and $Y \subseteq \mathbb{R}^n$ be a nonempty convex set which does not intersect C.
Then, $\exists \boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that

$$\forall \boldsymbol{x} \in C, \ \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq 0 \quad \text{and} \quad \forall \boldsymbol{y} \in Y, \ \langle \boldsymbol{y}, \boldsymbol{v} \rangle \geq 0.$$

If in addition, C is closed and Y is compact, then:

$$\exists \boldsymbol{v} \in \mathbb{R}^n : \ \forall \boldsymbol{x} \in C, \ \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq 0 \quad \text{and} \quad \forall \boldsymbol{y} \in Y, \ \langle \boldsymbol{y}, \boldsymbol{v} \rangle > 0.$$

Separation theorem for a cone

When one of the two sets is a cone, we can set $c = 0$:

Theorem (Separating hyperplane for a cone).

Let $C \subseteq \mathbb{R}^n$ be a nonempty convex cone, and $Y \subseteq \mathbb{R}^n$ be a nonempty convex set which does not intersect C.
Then, $\exists \boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that

$$\forall \boldsymbol{x} \in C, \; \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq 0 \quad \text{and} \quad \forall \boldsymbol{y} \in Y, \; \langle \boldsymbol{y}, \boldsymbol{v} \rangle \geq 0.$$

If in addition, C is closed and Y is compact, then:

$$\exists \boldsymbol{v} \in \mathbb{R}^n : \; \forall \boldsymbol{x} \in C, \; \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq 0 \quad \text{and} \quad \forall \boldsymbol{y} \in Y, \; \langle \boldsymbol{y}, \boldsymbol{v} \rangle > 0.$$

In particular, if C is a closed convex cone and $\boldsymbol{y} \notin C$,

$$\exists \boldsymbol{v} \in \mathbb{R}^n : \; \forall \boldsymbol{x} \in C, \; \langle \boldsymbol{x}, \boldsymbol{v} \rangle \leq 0 \quad \text{and} \quad \langle \boldsymbol{y}, \boldsymbol{v} \rangle > 0.$$

Supporting hyperplane

A hyperplane separating S from some $\mathbf{y} \notin S$ is called a *supporting hyperplane* if *it touches* S:

Definition (Supporting hyperplane).

Let $S \subseteq \mathbb{R}^n$ be nonempty, $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $b \in \mathbb{R}$. We say that $H = \{\mathbf{x} : \mathbf{a}^T \mathbf{x} = b\}$ is a supporting hyperplane of S if

- S is contained in one of the two halfspaces defined by H, i.e,

$$\forall \mathbf{x} \in S, \ \mathbf{a}^T \mathbf{x} \leq b \quad \text{or} \quad \forall \mathbf{x} \in S, \ \mathbf{a}^T \mathbf{x} \geq b.$$

- S has at least one boundary point on the hyperplane, i.e., $H \cap \partial S \neq \emptyset$, where $\partial S := \mathbf{cl}\, S \setminus \mathbf{int}\, S$ is the boundary of S.

Supporting hyperplane theorem

Theorem (Supporting hyperplane).

Let S be a convex set and x_0 be a boundary point of S. Then, S has a supporting hyperplane at x_0, that is,

$$\exists \boldsymbol{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\} : \quad \forall \boldsymbol{x} \in S,\ \boldsymbol{a}^T \boldsymbol{x} \leq \boldsymbol{a}^T \boldsymbol{x}_0.$$

Conversely, if S is closed , has nonempty interior, and has (at least) one supporting hyperplane in each of its boundary points, then S is convex.