Convex Optimization Theory and Applications

Topic 1 - Convex Set

Li Li

Department of Automation
Tsinghua University

Fall, 2009-2021.

1.0. Outline

1.-1. Preliminary 预备知识

1.1. Definition and Examples 定义和例子

1.2. Generalized Inequalities 广义不等式

1.3. Operations that Preserve Convexity 保凸运算

1.4. Separating Hyperplane Theorem 凸集分离

1.5. Some Quiz Problems

1.-1. Preliminary

非正式地说，向量 x 的范数 $|x|_p$ 是这个向量的某种长度度量，以第二范数欧几里得范数为例，$|x|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$ ，写成向量形式为 $|x|_2^2 = x^T x$ 。

正式地说，范数是满足以下四个条件的任何函数 $f : \mathbb{R}^n \to \mathbb{R}$：

- 非负性：对于所有 $x \in \mathbb{R}^n$，有 $f(x) \geq 0$
- 正定性：当且仅当 $x = 0$ 的时候 $f(x) = 0$
- 齐次性：对于所有 $x \in \mathbb{R}^n,\ t \in \mathbb{R}$，$f(tx) = |t|\, f(x)$
- 三角不等式：对于所有 $x, y \in \mathbb{R}^n$，$f(x+y) \leq f(x) + f(y)$

1.-1. Preliminary

仿射子空间（affine subspaces）：给定矩阵$A\in\mathbb{R}^{m\times n}$以及向量$b\in\mathbb{R}^m$，仿射子空间为集合$\{x\in\mathbb{R}^n : Ax=b\}$

仿射变换（affine transformation）：从$\mathbb{R}^n$到$\mathbb{R}^m$的变换$y=Ax+c$被称为仿射变换，其中$y\in\mathbb{R}^m$，$A\in\mathbb{R}^{m\times n}$，$c\in\mathbb{R}^m$为常向量

仿射函数（affine function）：仿射函数为仿射变换的一种特殊情况，当仿射变换中的$m=1$时，$A^T\in\mathbb{R}^n$，y,c都为常数；此外，如果将此处$y=Ax+c$中的截距c去掉，剩下的部分$y=Ax$被称为线性函数

1.1. Definition and Examples

为什么我们从凸集开始

布尔巴基学派认为“数学结构勾画了数学体系内部的关系，数学结构简称为结构，结构由若干集合，定义在集合上或集合间的一些关系，以及一组作为条件的公理组成”

--《数学辞海》编辑委员会《数学辞海（第一卷）》

希望大家在学习的过程中，能逐渐熟悉相对抽象的性质和定理，同时又掌握足够多的相对具象的实例

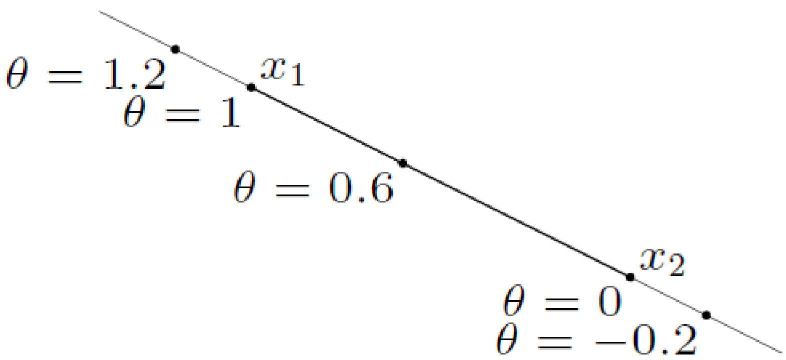
1.1. Definition and Examples

点Point，在我们课中，如无特殊说明，指欧式空间中的点

线Line和放射集Affine set

line through x_1, x_2: all points

$$x = \theta x_1 + (1-\theta)x_2 \qquad (\theta \in \mathbf{R})$$



affine set: contains the line through any two distinct points in the set

It is easy to show that every affine set is convex

1.1. Definition and Examples

Every affine set can be expressed as solution set of system of linear equations

Let S be the solution of a linear equation. By definition, we have $S = \{x \in \mathbb{R}^n : Ax = b\}$.

Let $x_1, x_2 \in S \Rightarrow Ax_1 = b$ and $Ax_2 = b$.

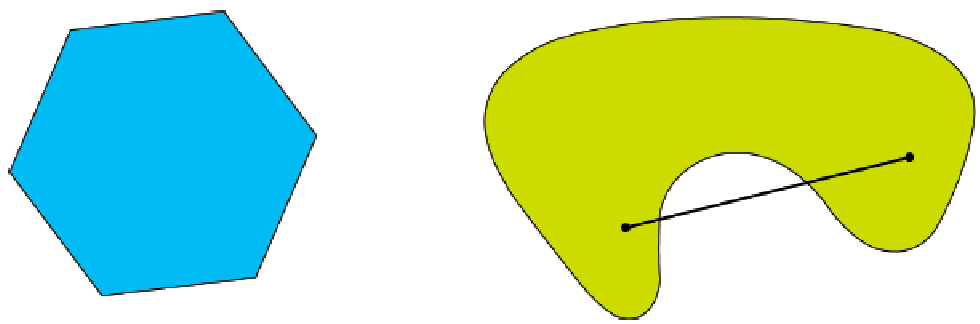
To prove: $A[\theta x_1 + (1-\theta)x_2] = b, \forall \theta \in \mathbb{R}$, let us check

$$A[\theta x_1 + (1-\theta)x_2] = \theta Ax_1 + (1-\theta)Ax_2 = \theta b + (1-\theta)b = b$$

Thus S is an affine set.

1.1. Definition and Examples

Set C is a convex set, if for any $x, y \in C$, and $\theta \in [0,1]$, we have $\theta x + (1-\theta) y \in C$



典型的凸集包括：

- 点 Point，线 Line，Ray，面 Hyperplane，空间 Half Space
- 锥 Cone
- 球 Ball，多面体 Polyhedral $P_1 = \{x \mid Ax \le b\}$

1.1. Definition and Examples

球 Ball、超平面 Hyperplane、半平面 Half Space 都是凸集

↓

作为超平面、半平面的交集的可行域是多面体 Polyhedron，多面体是凸集

↓

基于球的概念定义边界点 Boundary Point，顶点 Extreme Point

1.1. Definition and Examples

球 Ball

Definition

A *open ball* centered at a point $\boldsymbol{x}^*$ with radius $r \in \mathbb{R}^+$ is defined as

$$B_r(\boldsymbol{x}^*) = \{\boldsymbol{x} \mid |\boldsymbol{x} - \boldsymbol{x}^*|_2 < r\}$$

Definition

A *closed ball* centered at a point $\boldsymbol{x}^*$ with radius $r \in \mathbb{R}^+$ is defined as

$$\bar{B}_r(\boldsymbol{x}^*) = \{\boldsymbol{x} \mid |\boldsymbol{x} - \boldsymbol{x}^*|_2 \leq r\}$$

1.1. Definition and Examples

球 Ball

Both open balls and closed balls are convex.
We only show that the statement holds for open balls, since the other proof is similar. Suppose there are two points $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ in a open ball $B_r(\boldsymbol{x}^*)$. For any $\lambda \in [0, 1]$, we have

$$\begin{aligned} |(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) - \boldsymbol{x}^*|_2 &= |\lambda(\boldsymbol{x}_1 - \boldsymbol{x}^*) + (1-\lambda)(\boldsymbol{x}_2 - \boldsymbol{x}^*)|_2 \\ &\leq \lambda|\boldsymbol{x}_1 - \boldsymbol{x}^*|_2 + (1-\lambda)|\boldsymbol{x}_2 - \boldsymbol{x}^*|_2 \\ &< \lambda r + (1-\lambda)r = r \end{aligned}$$

By definition, the open balls are convex.

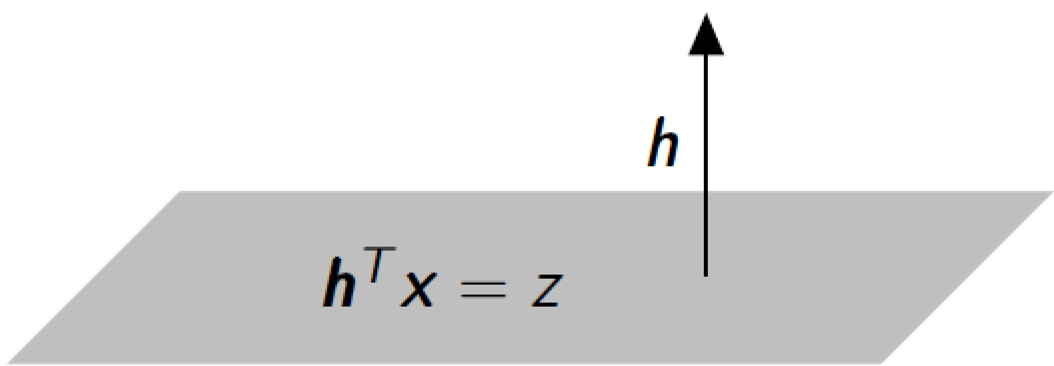
1.1. Definition and Examples

超平面 Hyperplane

A *hyperplane* in $\mathbb{R}^n$ is a set of all points satisfying $\{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} = z\}$, where $\boldsymbol{x}$ is an n-dimensional column vector in $\mathbb{R}^n$, $\boldsymbol{h}$ is a non-zero n-dimensional column vector in $\mathbb{R}^n$ and $z \in \mathbb{R}$. $\boldsymbol{h}$ is said to be the *normal* of hyperplane $H = \{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} = z \neq 0\}$.

Apparently, if and only if $z = 0$, the hyperplane $\{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} = z\}$ passes through the origin.

Moreover, $\boldsymbol{h}$ is orthogonal to any vectors lying in the hyperplane H. Indeed, for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in H$, we have $\boldsymbol{h}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2) = z - z = 0$.



1.1. Definition and Examples

半平面 Halfspace

A *closed half-space* in $\mathbb{R}^n$ is the set of all points satisfying $\{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} \leq z\}$, where $\boldsymbol{x}$ is an n-dimensional column vector in $\mathbb{R}^n$, $\boldsymbol{h}$ is an n-dimensional column vector in $\mathbb{R}^n$ and $z \in \mathbb{R}$.

A *open half-space* in $\mathbb{R}^n$ is the set of all points satisfying $\{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} < z\}$, where $\boldsymbol{x}$ is an n-dimensional column vector in $\mathbb{R}^n$, $\boldsymbol{h}$ is an n-dimensional column vector in $\mathbb{R}^n$ and $z \in \mathbb{R}$.

Some literatures use $\geq$ $(>)$ instead of $\leq$ $(<)$, but their meaning are indeed the same.

1.1. Definition and Examples

半平面 Half Space

1. Any a hyperplane is convex.
2. The closed half-space $\{\mathbf{x} \mid \mathbf{h}^T\mathbf{x} \le z\}$ *and* $\{\mathbf{x} \mid \mathbf{h}^T\mathbf{x} \ge z\}$ *are convex.*
3. The open half-space $\{\mathbf{x} \mid \mathbf{h}^T\mathbf{x} < z\}$ *and* $\{\mathbf{x} \mid \mathbf{h}^T\mathbf{x} > z\}$ *are convex.*

For statement 2, let $\Omega = \{\mathbf{x} \mid \mathbf{h}^T\mathbf{x} \ge z\}$ be the studied half-space. For all $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$ and $\lambda \in [0,1]$, we have

$$\mathbf{h}^T[\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2] = \lambda\mathbf{h}^T\mathbf{x}_1 + (1-\lambda)\mathbf{h}^T\mathbf{x}_2 \ge \lambda z + (1-\lambda)z = z.$$

So $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \Omega$, which means Ω is convex. Similarly we can prove the other half-space is convex.

1.1. Definition and Examples

边界点 Boundary Point

Let Ω be a subset of $\mathbb{R}^n$. A point $\boldsymbol{x}$ is a *boundary point* of Ω if every open ball centered at $\boldsymbol{x}$ contains both a point in Ω and a point in $\mathbb{R}^n - \Omega$. The set of all boundary points of Ω, denoted by $\partial\Omega$, is the *boundary* of Ω.

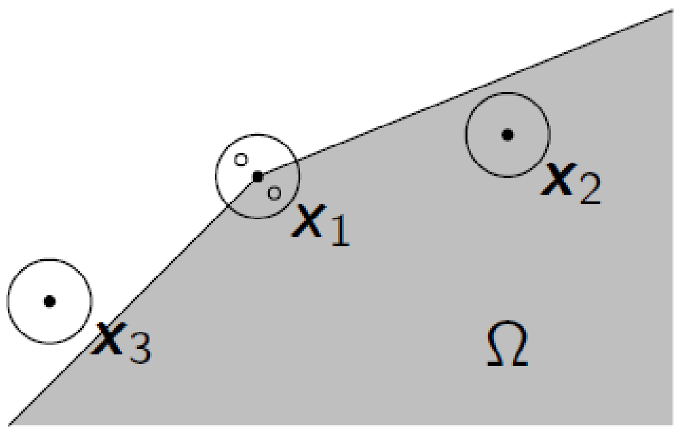


Figure: An illustration of the boundary points, where $\boldsymbol{x}_1$ is a boundary point and $\boldsymbol{x}_2$, $\boldsymbol{x}_3$ are not.

1.1. Definition and Examples

顶点Extreme Point，如果凸集内的一点不在凸集内任何不同的两点的连线上，则称该点为该凸集的顶点

A point $\boldsymbol{x}$ is an *extreme point* of a convex set C if there exist no two distinct points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2 \in C$ such that $\boldsymbol{x} = \lambda \boldsymbol{x}_1 + (1 - \lambda)\boldsymbol{x}_2$ for some $\lambda \in (0, 1)$.

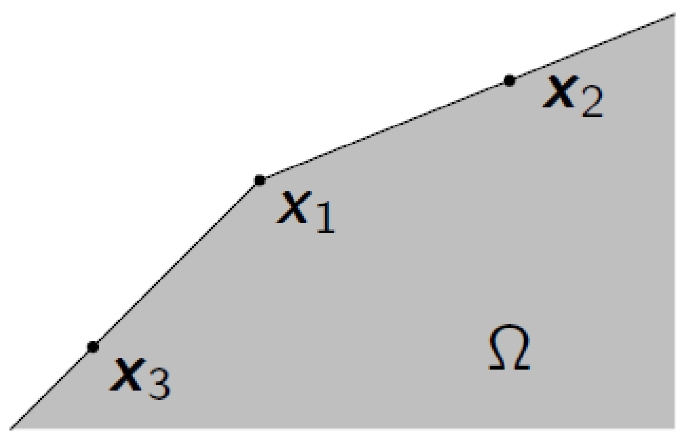


Figure: An illustration of the extreme points, where $\boldsymbol{x}_1$ is an extreme point and $\boldsymbol{x}_2$, $\boldsymbol{x}_3$ are not.

1.1. Definition and Examples

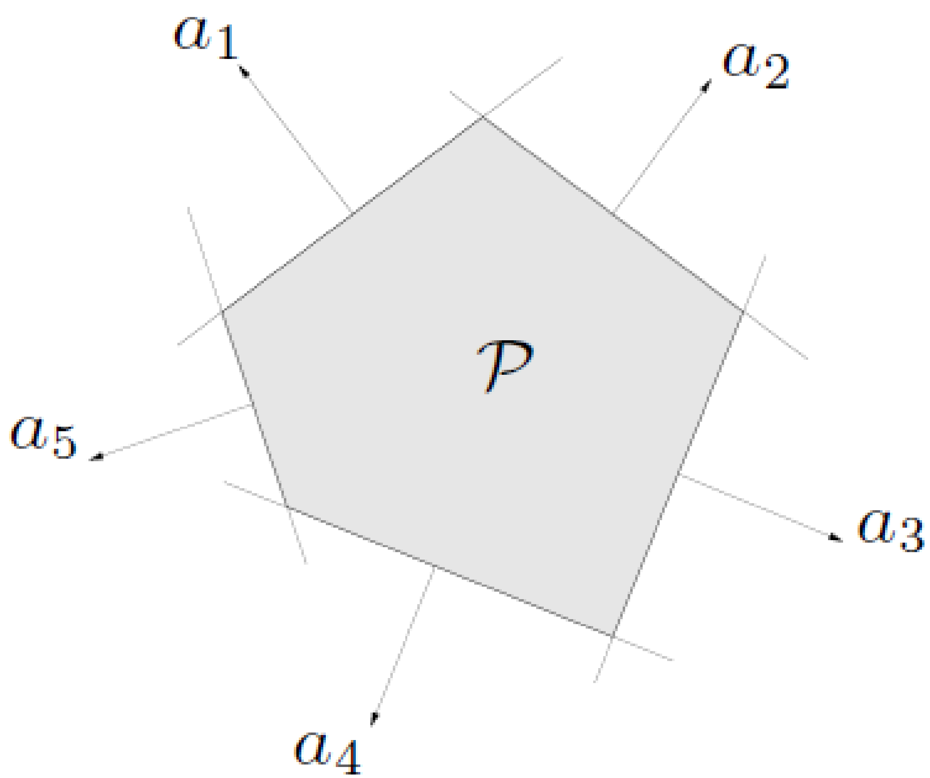
多面体Polyhedron

The intersection of convex sets is still convex.

Let us define the intersection of k convex sets as $\Omega = \bigcap_{i=1,\ldots,k} \Omega_i$. For any $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega_i$, $i = 1, \ldots, k$, we have $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \Omega_i$, $i = 1, \ldots, k$. So, $\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \Omega$. Therefore, Ω is convex.

A *polyhedron* is defined as the solution set of a finite number of linear equalities and inequalities:
$\Omega = \left\{ \boldsymbol{x} \mid \boldsymbol{a}_i^T \boldsymbol{x} \leq b_i, i = 1, \ldots, m, \boldsymbol{c}_j^T \boldsymbol{x} = d_j, j = 1, \ldots, p \right\}$. A polyhedron is thus the intersection of a finite number of halfspaces and hyperplanes.

1.1. Definition and Examples



A polyhedron is intersection of a finite number of halfspaces and hyperplanes

1.1. Definition and Examples

多面体Polyhedron是凸集也可以用下面的方法来证明：

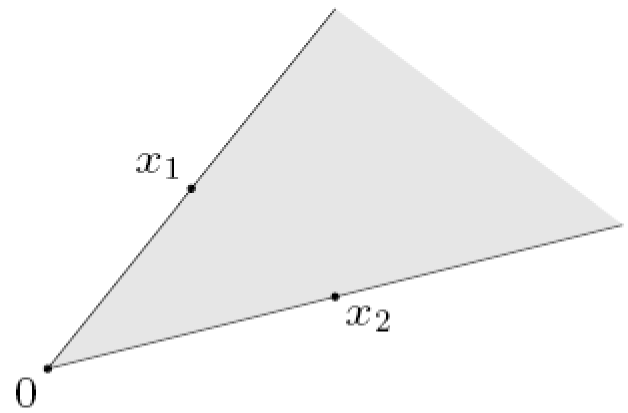
大前提：线性规划问题的约束条件为若干个线性等式或者不等式，而这些集合都是凸集

小前提：凸集的交集也是凸集

结论：作为这些凸集交集的线性规划问题定义域也是凸集（我们约定空集为凸集）

1.1. Definition and Examples

Conic (nonnegative) combination of x_1 and x_2: any point of the form $x = \theta_1 x_1 + \theta_2 x_2$, with θ_1, $\theta_2 \geq 0$.



凸锥Convex cone C: set that contains all conic combinations of points in the set and convex. Thus, $y \in C$ if and only if there is a solution of θ_1, $\theta_2 \geq 0$ for $y = \theta_1 x_1 + \theta_2 x_2$.

1.1. Definition and Examples

A set $\mathcal{K} \subset \mathbb{R}^n$ is a *cone* if

$$x \in \mathcal{K} \quad \Rightarrow \quad \theta x \in \mathcal{K}, \ \forall \theta \geq 0$$

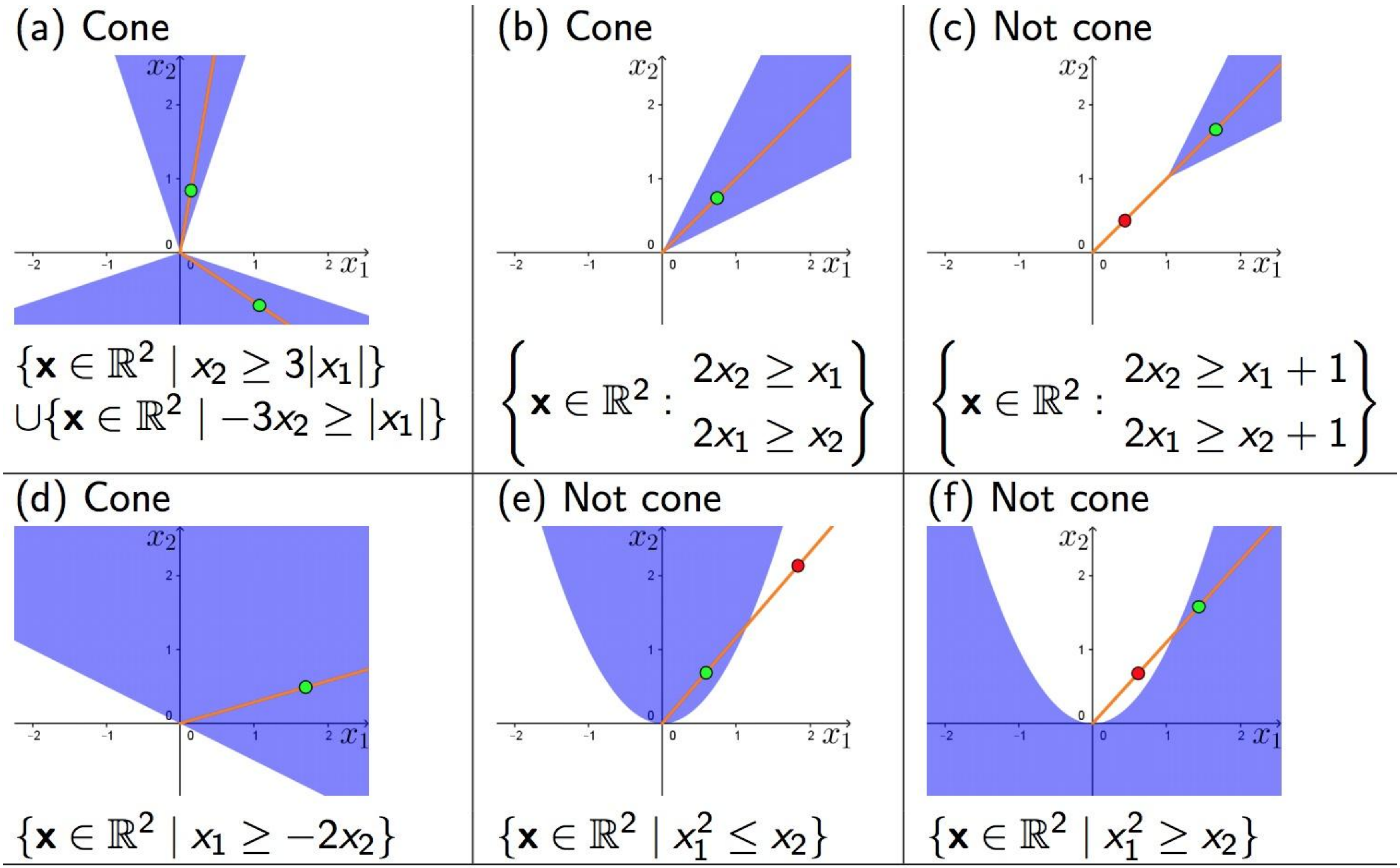
A set $\mathcal{K}$ is a *convex cone* if $\mathcal{K}$ is a cone and $\mathcal{K}$ is convex.

If $\mathcal{K}_1$ and $\mathcal{K}_2$ are cones then $\mathcal{K} = \mathcal{K}_1 \bigcap \mathcal{K}_2$ is a cone.

A cone $\mathcal{K}$ is a *proper cone* if

- $\mathcal{K}$ is convex
- $\mathcal{K}$ is closed
- $\mathcal{K}$ is *solid*, i.e. it has non-empty interior
- $\mathcal{K}$ is *pointed*, i.e. if $x \in \mathcal{K}$ and $-x \in \mathcal{K}$ then $x = 0$

1.1. Definition and Examples



1.1. Definition and Examples

Definition

A cone is said to be a *pointed cone* if it includes the origin, $\mathbf{0}$; otherwise, it is said to be *blunt cone*.

Notice that if $\boldsymbol{x} \in \Omega$, we also have $\lambda \boldsymbol{x}, (1-\lambda)\boldsymbol{x} \in \Omega$, for any $\lambda \in (0, 1)$. Thus, $\mathbf{0}$ is the only possible extreme point in a cone.

Definition

A polyhedron in the form of $\Omega = \{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} \geq \mathbf{0}\}$ is a cone called *polyhedral cone*.

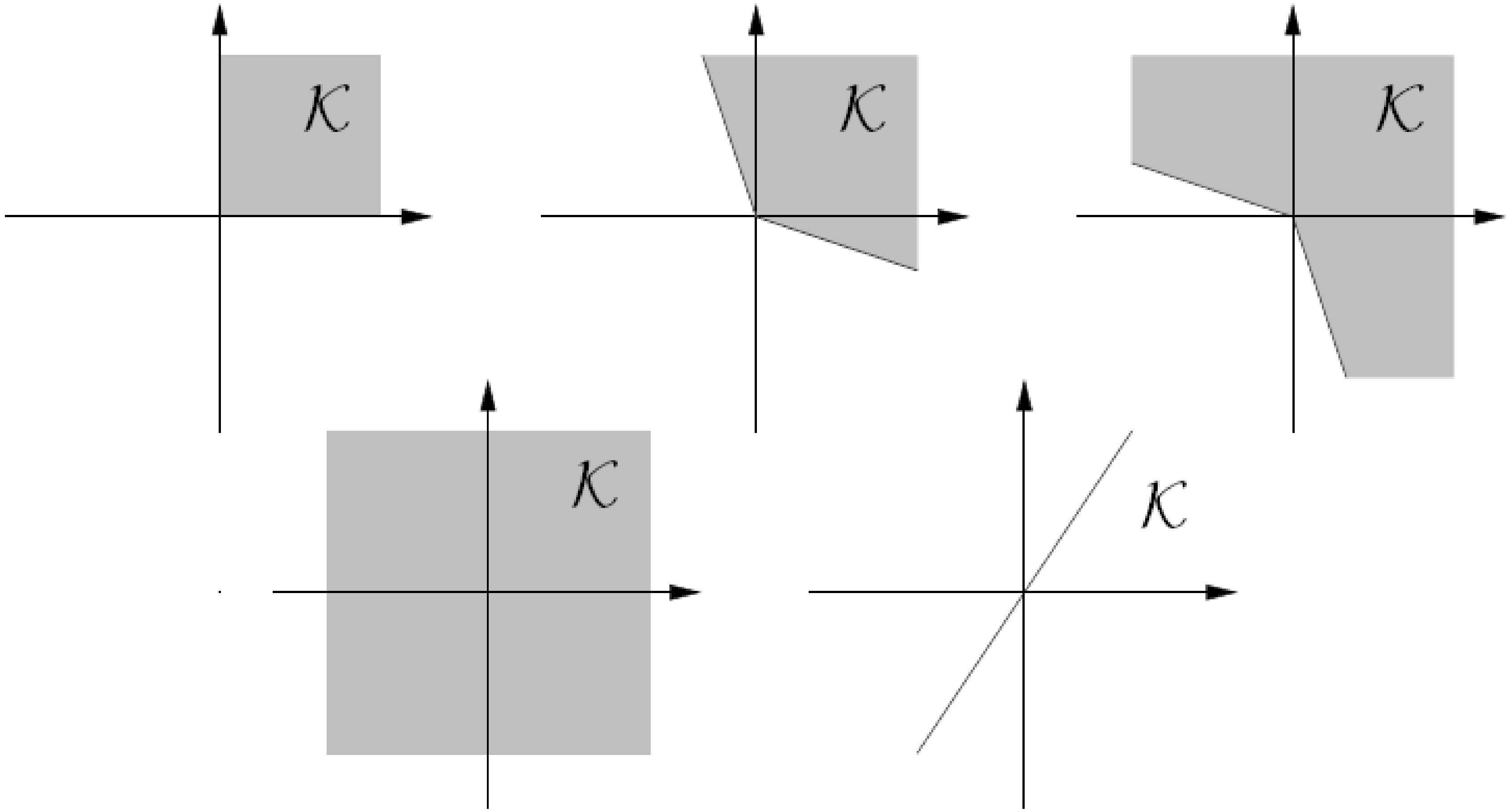
Theorem

$\mathbf{0}$ is the extreme point of a polyhedron cone $\{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} \geq \mathbf{0}\}$, if and only if $rank(A) = n$.

Proof.

First, when $\mathbf{0}$ is an extreme point, we can prove $\mathrm{rank}(A) = n$ by contradiction.
Suppose we have a point $\boldsymbol{y} \neq \mathbf{0}$ satisfying $A\boldsymbol{y} = \mathbf{0}$. We can easily show that $-\boldsymbol{y}$ also satisfies $A(-\boldsymbol{y}) = \mathbf{0}$. Thus, both $\boldsymbol{y}$ and $-\boldsymbol{y}$ are in this cone; while $\boldsymbol{y} + (-\boldsymbol{y}) = \mathbf{0}$. This contradicts our assumption that $\mathbf{0}$ is the extreme point.
Second, when $\mathrm{rank}(A) = n$, we can prove $\mathbf{0}$ is an extreme point by contradiction.
If $\mathrm{rank}(A) = n$, $\mathbf{0}$ is clearly the only solution of $A\boldsymbol{x} = 0$. Assume two different points $\boldsymbol{x}_1, \boldsymbol{x}_2 > \mathbf{0}$ satisfying $\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 = \mathbf{0}$, $\lambda \in (0, 1)$. Thus, we have $A\left[\lambda\boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2\right] = \mathbf{0}$. Since $A\boldsymbol{x}_1 \geq \mathbf{0}$ and $A\boldsymbol{x}_2 \geq \mathbf{0}$, we have $A\boldsymbol{x}_1 = A\boldsymbol{x}_2 = \mathbf{0}$. This contradicts our assumption that $\mathbf{0}$ is the only solution of $A\boldsymbol{x} = 0$. $\square$

1.1. Definition and Examples



Are they cones? Are they convex and/or proper cones? Remember that a proper cone contains no line.

1.1. Definition and Examples

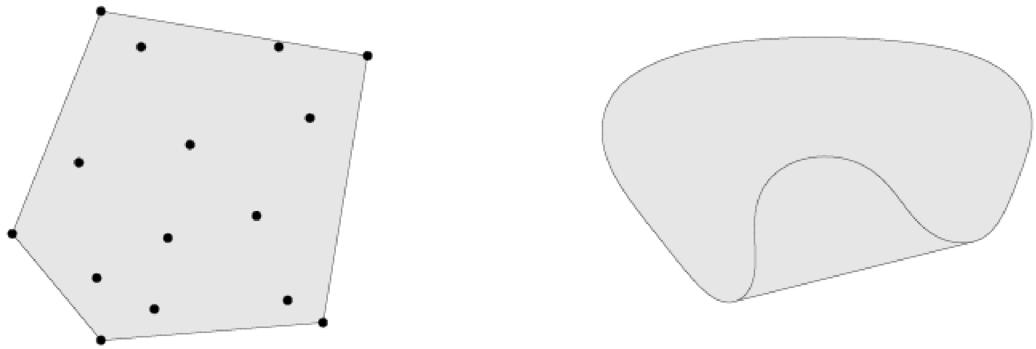
凸组合Convex combination of $x_1, \ldots, x_k$: any point x of the form

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

with $\theta_1 + \theta_2 + \cdots + \theta_k = 1,\ \theta_i \geq 0$.

等价定义1：The convex hull of a set Ω is defined to be the intersection of all convex sets that contain Ω.

等价定义2：The convex hull $\operatorname{conv} S$ is the set of all convex combinations of points in S.



1.1. Definition and Examples

我们可以证明这两个定义的等价性

Theorem

The convex hull of a finite number of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ is the set of all convex combinations of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$.

Proof.

Let us define the set of all convex combinations of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ as

$$\Omega_c = \left\{ \boldsymbol{x} \mid \boldsymbol{x} = \sum_{i=1}^{k} M_i \boldsymbol{x}_i,\ M_i \geq 0,\ \sum_{i=1}^{k} M_i = 1 \right\}$$

and Ω_h be the convex hull of $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$. We will prove $\Omega_c = \Omega_h$.

First, we show Ω_c is a convex set. For any two points $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \Omega_c$, we have

$$\boldsymbol{y}_1 = \sum_{i=1}^{k} M_i \boldsymbol{x}_i, \ \boldsymbol{y}_2 = \sum_{i=1}^{k} N_i \boldsymbol{x}_i$$

with $M_i, N_i \geq 0$ and

$$\sum_{i=1}^{k} M_i = 1, \ \sum_{i=1}^{k} N_i = 1$$

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

$$\begin{aligned} \boldsymbol{y} &= \lambda \boldsymbol{y}_1 + (1-\lambda)\boldsymbol{y}_2 = \lambda \sum_{i=1}^{k} M_i \boldsymbol{x}_i + (1-\lambda) \sum_{i=1}^{k} N_i \boldsymbol{x}_i \\ &= \sum_{i=1}^{k} [\lambda M_i + (1-\lambda) N_i] \boldsymbol{x}_i \end{aligned}$$

Clearly, we have $\sum_{i=1}^{k} [\lambda M_i + (1-\lambda) N_i] = \lambda + (1-\lambda) = 1$. That is, we have $\boldsymbol{y} \in \Omega_c$. So, Ω_c is convex.

Second, we can easily prove $\Omega_h \subseteq \Omega_c$, since $\boldsymbol{x}_i = \sum_{j=1}^{k} \delta_j^i \boldsymbol{x}_j$, with $\delta_j^i = 0$ for any $j \neq i$ and $\delta_j^i = 1$ for $j = i$.
Third, we can use mathematical induction to show that $\Omega_c \subseteq \Omega_h$, too.
For $k = 1$, we have $\Omega_h = \{\boldsymbol{x}_1\} = \Omega_c$. It is easy to show $\Omega_c^k \subseteq \Omega_h^k$ holds for the case $k = 1$.
Assuming $\Omega_c^k \subseteq \Omega_h^k$ holds for the case k. For any a $\boldsymbol{x} \in \Omega_c^{k+1}$, we have

$$\boldsymbol{x} = \sum_{i=1}^{k+1} M_i \boldsymbol{x}_i$$

with $M_i \geq 0$ and $\sum_{i=1}^{k+1} M_i = 1$.
1) If $\boldsymbol{x} = \boldsymbol{x}_{k+1}$, we have $\boldsymbol{x} \in \Omega_h$ and the statement holds.
2) If $\boldsymbol{x} \neq \boldsymbol{x}_{k+1}$, we have $M_{k+1} \neq 1$, thus

$$\sum_{i=1}^{k} \frac{M_i}{1 - M_{k+1}} = 1$$

Let us define $\boldsymbol{y}$ as the projection of $\boldsymbol{x}$ on the convex set Ω_c^k; see Fig.1.

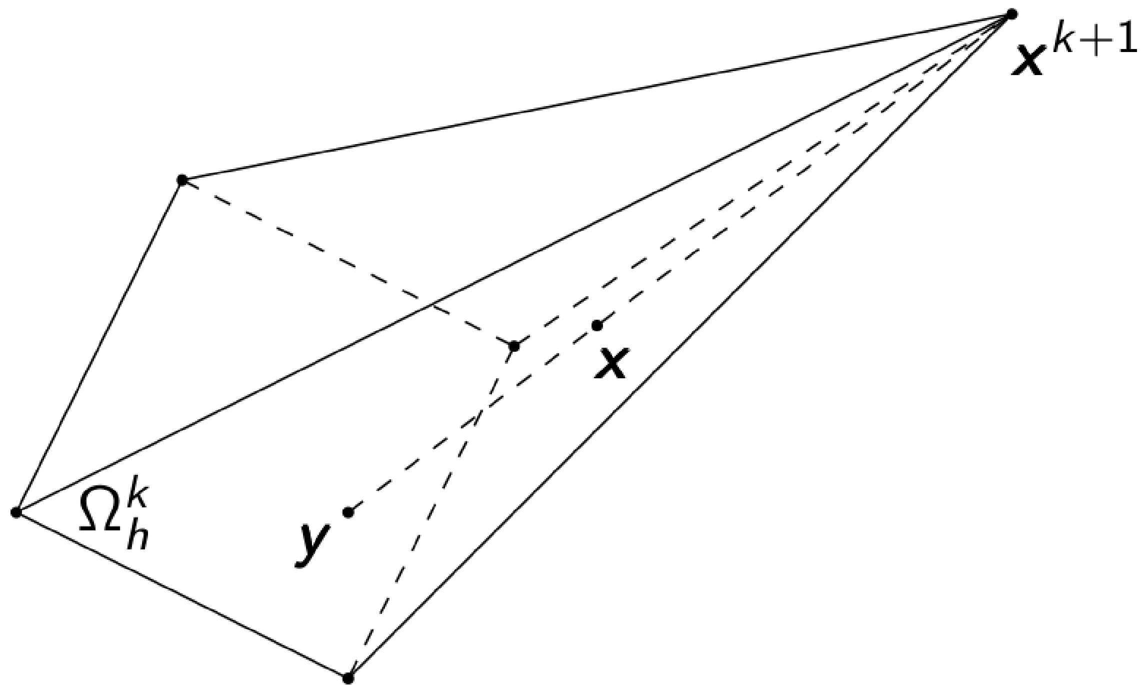


Figure: An illustration on the projection of $\boldsymbol{x}$ on the convex set Ω_c^k.

By the induction hypothesis, we have

$$\boldsymbol{y} = \sum_{i=1}^{k} \frac{M_i}{1 - M_{k+1}} \boldsymbol{x}_i \in \Omega_h^k \subseteq \Omega_h^{k+1}$$

1.1. Definition and Examples

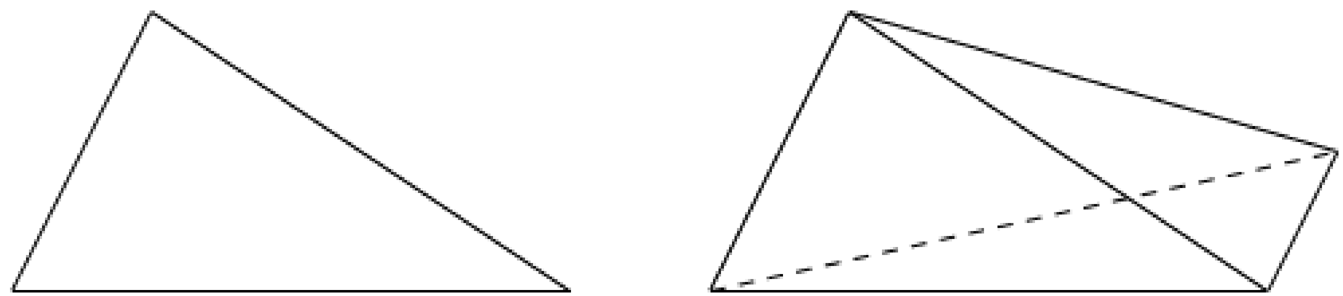
Because $\boldsymbol{x}_{k+1} \in \Omega_h^{k+1}$ and Ω_h^{k+1} is a convex set by definition, we have

$$\boldsymbol{x} = \sum_{i=1}^{k+1} M_i \boldsymbol{x}_i = (1 - M_{k+1})\boldsymbol{y} + M_{k+1}\boldsymbol{x}_{k+1} \in \Omega_h^{k+1}$$

So, $\Omega_c^{k+1} \subseteq \Omega_h^{k+1}$ for the case $k+1$. By induction property, we have $\Omega_c \subseteq \Omega_h$.
Since $\Omega_c \subseteq \Omega_h$ and $\Omega_h \subseteq \Omega_c$ holds simultaneously, we reach the statement.

1.1. Definition and Examples

单纯形Simplex：We call the convex hull of any set of $n+1$ points in R^n which do not lie on a hyperplane a simplex.



A simplex in $\mathbb{R}^2$ and a simplex in $\mathbb{R}^3$.

多胞形A bounded polyhedron is called a polytope.

1.1. Definition and Examples

Positive semidefinite cone

The set of real symmetric $n \times n$ matrices is denoted S^n. A matrix $A \in S^n$ is called positive semidefinite if $x^T Ax \geq 0$ for all $x \in R^n$, and is called positive definite if $x^T Ax > 0$ for all nonzero $x \in R^n$. The set of positive semidefinite matrices is denoted S_+^n and the set of positive definite matrices is denoted by S_{++}^n. S_+^n is a closed, convex, pointed, and solid cone.

There are several equivalent conditions for a matrix to be positive (semi)definite.

In the follows, we will focus on semipositive matrices but the conclusions can be extended to positive matrices.

1.1. Definition and Examples

Proposition 1.1: The following statements are equivalent:

1) matrix $A \in S^n$ is positive semidefinite
2) for all $x \in R^n$, $x^T A x \geq 0$
3) all the eigenvalues of A are nonnegative
4) all principal minors of A are nonnegative
5) there exists a factorization $A = B^T B$

1.1. Definition and Examples

Proof: 1) and 2) are equivalent by definition. 3) and 4) are obvious according to *Lemma 1.1* below.

Define $A = B^T B$. For any $x \in R^n$, $x^T Ax = x^T B^T Bx = \tilde{x}^T \tilde{x} \geq 0$, where $\tilde{x} = Bx$. So we can derive 5) from 1).

On the other side, based on on 5), there exist $A = P^T P = Q^T \Lambda Q$, where $P = \Lambda^{1/2} Q^T$, Λ is the diagonal matrix containing all the eigenvalues of A. This indicates 5) leads to 1).

Corollary 5.1: Every psd (pd) matrix A has a unique psd (pd) square root $A^{1/2}$ such that $A^{1/2} A^{1/2} = A$ and every pd matrix is nonsingular.

1.1. Definition and Examples

Proposition 1.2: The following statements are equivalent:

1) matrix $A \in S^n$ is positive

2) for all $x \in R^n$, $x^T A x > 0$

3) all the eigenvalues of A are positive

4) all principal minors of A are positive

5) there exists a factorization $A = B^T B$ (any else condition on B?)

1.1. Definition and Examples

Lemma 1.1: *Any a real Hermitian matrix* A *can be diagonalized by an orthogonal matrix.*

Proof:

1) all the eigenvalues and eigenvectors are real, notice that $\lambda(A)=\lambda(A^T)$

2) the eigenvectors of different eigenvalues are orthogonal. Suppose $\lambda_1 X_1 = AX_1$, $\lambda_2 X_2 = AX_2$, $\lambda_1 \neq \lambda_2$, $X_1^T X_2 = 0$, because $\lambda_1 X_1^T X_2 = X_1^T AX_2 = X_2^T AX_1 = \lambda_2 X_2^T X_1$, $(\lambda_1 - \lambda_2) X_1^T X_2 = 0$

3) if λ is a r th multiple root of the characteristic equation of A, we can prove that $rank(A-\lambda I) = n-r$. Thus, λ has r linear independent eigenvectors.

1.1. Definition and Examples

4) based on 2) and 3), we can see that A has n linear independent eigenvectors. Via Gram-Schmidt orthogonalization, we can make these eigenvectors orthogonal. So it is proved.

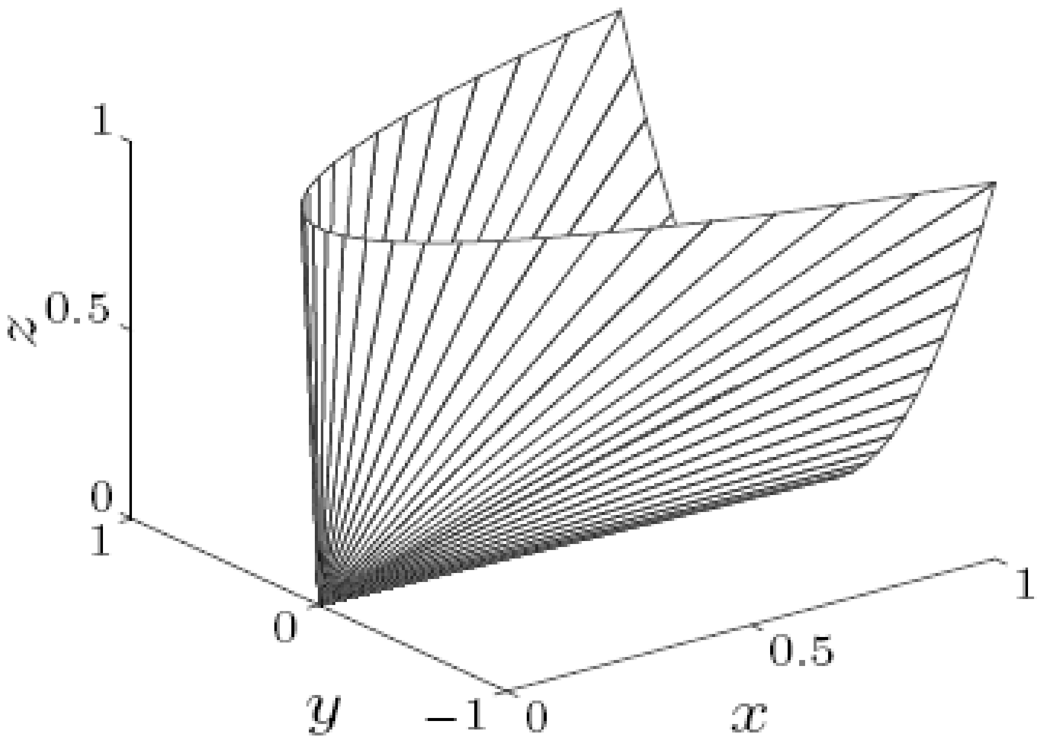
Another Proof:

According to the Schur triangularization theorem there is a unitary matrix U such that $U^T AU = R$, where R is upper triangular. But, moreover, we have $R^T = (U^T AU)^T = U^T AU = R$. Therefore, R is both upper and lower triangular. This makes R a diagonal matrix.

1.1. Definition and Examples

The set of all PSD matrices $A \in R^{n \times n}$ forms a convex cone.

For example, let $A = \begin{bmatrix} x & y \\ y & z \end{bmatrix}$, we have



1.1. Definition and Examples

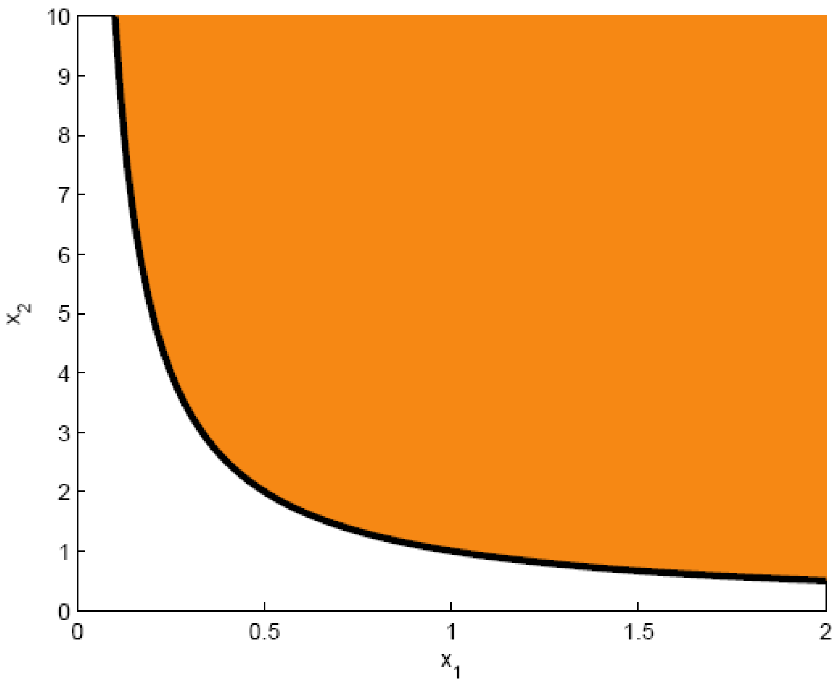
Is the following set convex?

$$x_1 \geq 0, \quad x_1 x_2 \geq 1$$

1.1. Definition and Examples

Yes

$$\begin{bmatrix} x_1 & 1 \\ 1 & x_2 \end{bmatrix} \geq 0$$



1.1. Definition and Examples

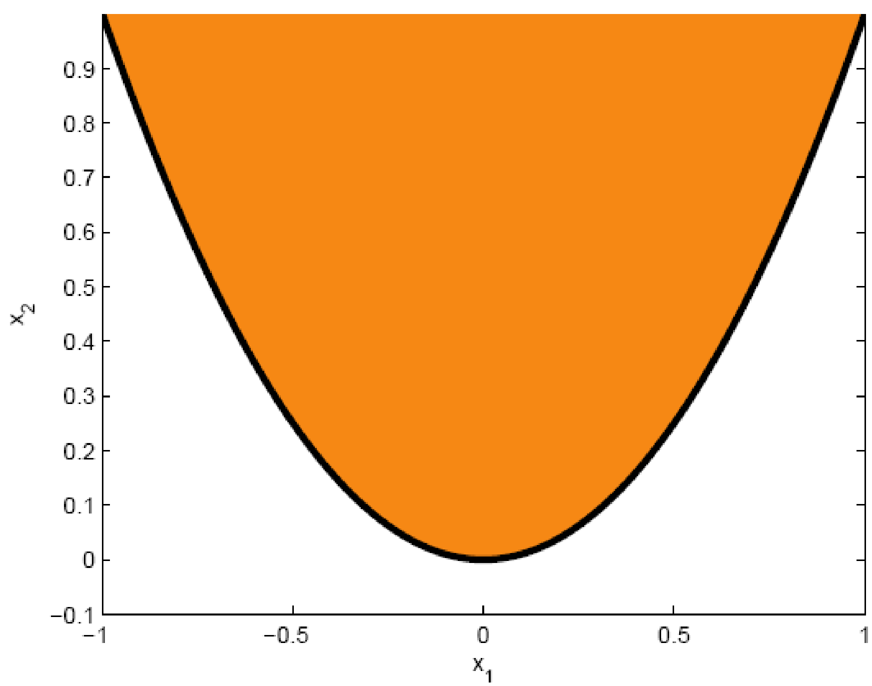
Is the following set convex?

$$x_2 \geq x_1^2$$

1.1. Definition and Examples

Yes

$$\begin{bmatrix} 1 & x_1 \\ x_1 & x_2 \end{bmatrix} \geq 0$$



1.1. Definition and Examples

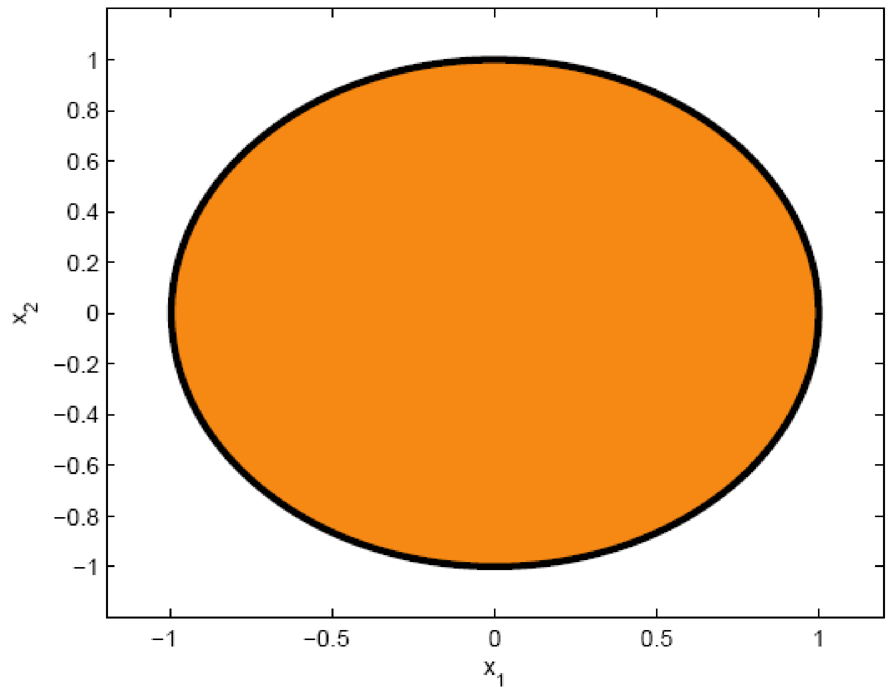
Is the following set convex?

$$x_1^2 + x_2^2 \leq 1$$

1.1. Definition and Examples

Yes

$$\begin{bmatrix} 1+x_1 & x_2 \\ x_2 & 1-x_1 \end{bmatrix} \geq 0$$



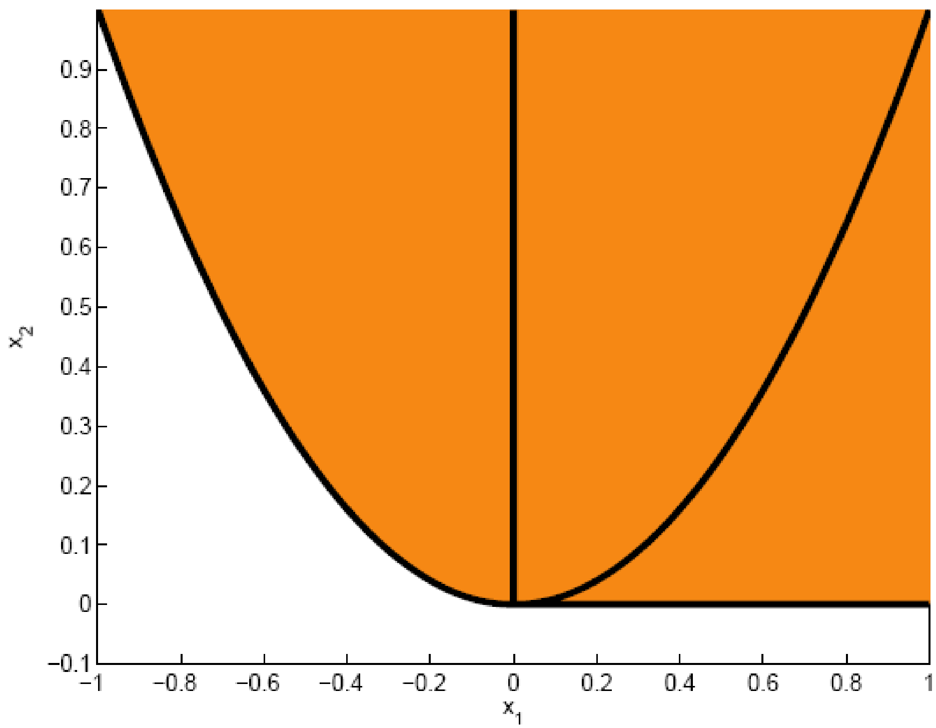
1.1. Definition and Examples

Is the following set convex?

$$x_2 \geq x_1^2 \quad \text{or} \quad x_1 \geq 0, \quad x_2 \geq 0$$

1.1. Definition and Examples

Yes, but cannot be written into a SDP (LMI) form, because not basic semialgebraic (a semialgebraic set is a subset S of real n-dimensional space defined by a finite sequence of polynomial equations and inequalities; or any finite union of such sets.)



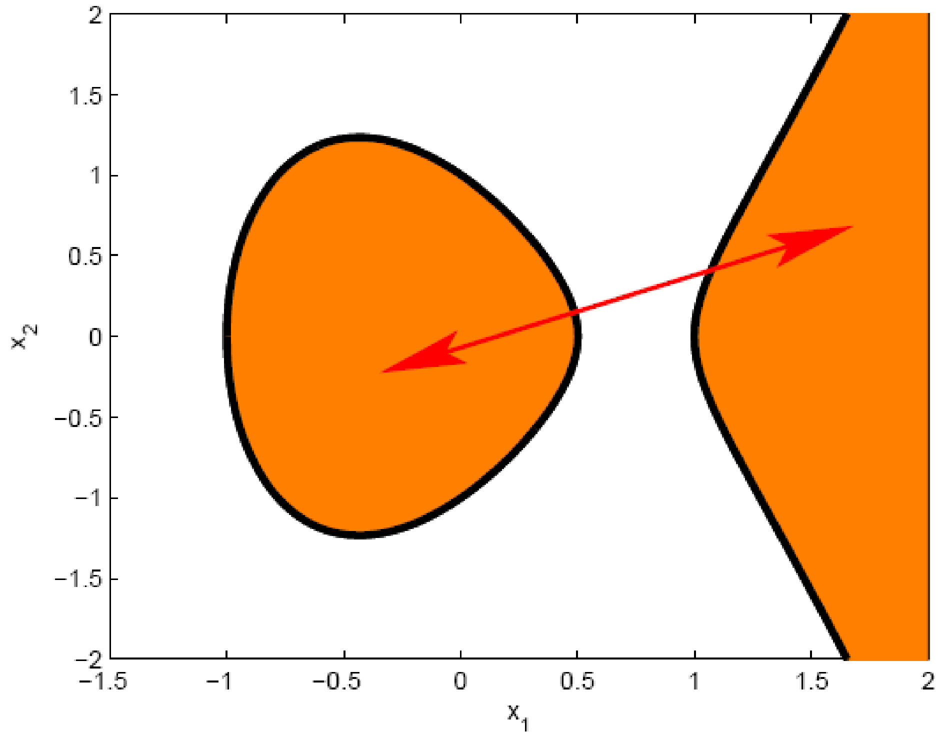
1.1. Definition and Examples

Is the following set convex?

$$1 - 2x_1 - x_1^2 - x_2^2 + 2x_1^3 \geq 0$$

1.1. Definition and Examples

No because not connected



1.1. Definition and Examples

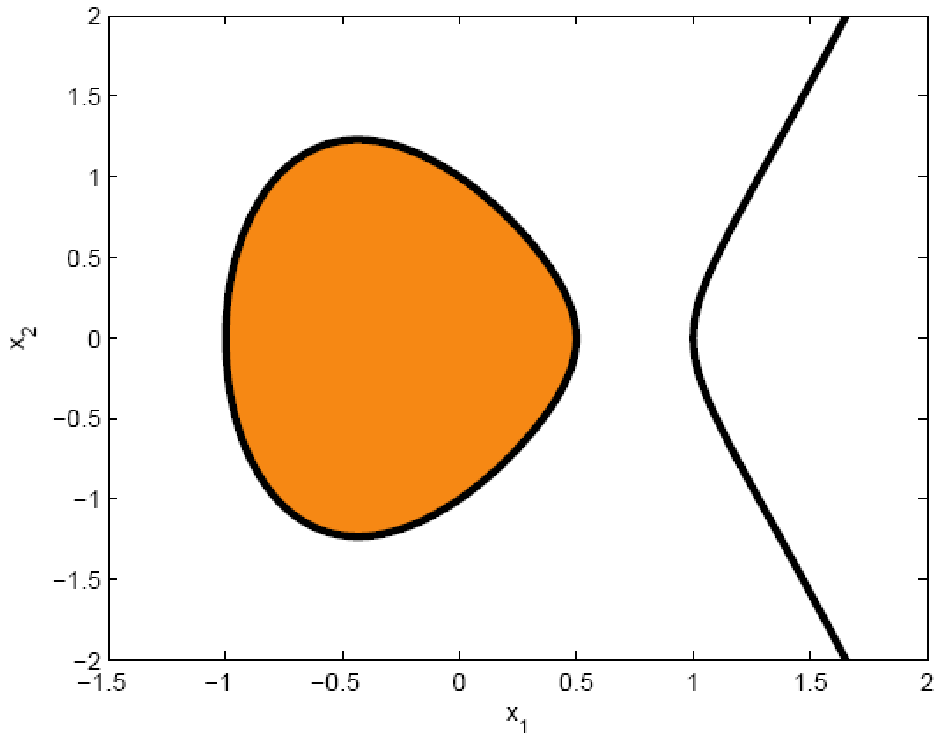
Is the following set convex?

$$1-2x_1-x_1^2-x_2^2+2x_1^3\geq 0 \quad \text{and} \quad x_1\leq 0.5$$

1.1. Definition and Examples

Yes

$$\begin{bmatrix} 1 & x_1 & 0 \\ x_1 & 1 & x_2 \\ 0 & x_2 & 1-2x_1 \end{bmatrix} \geq 0$$



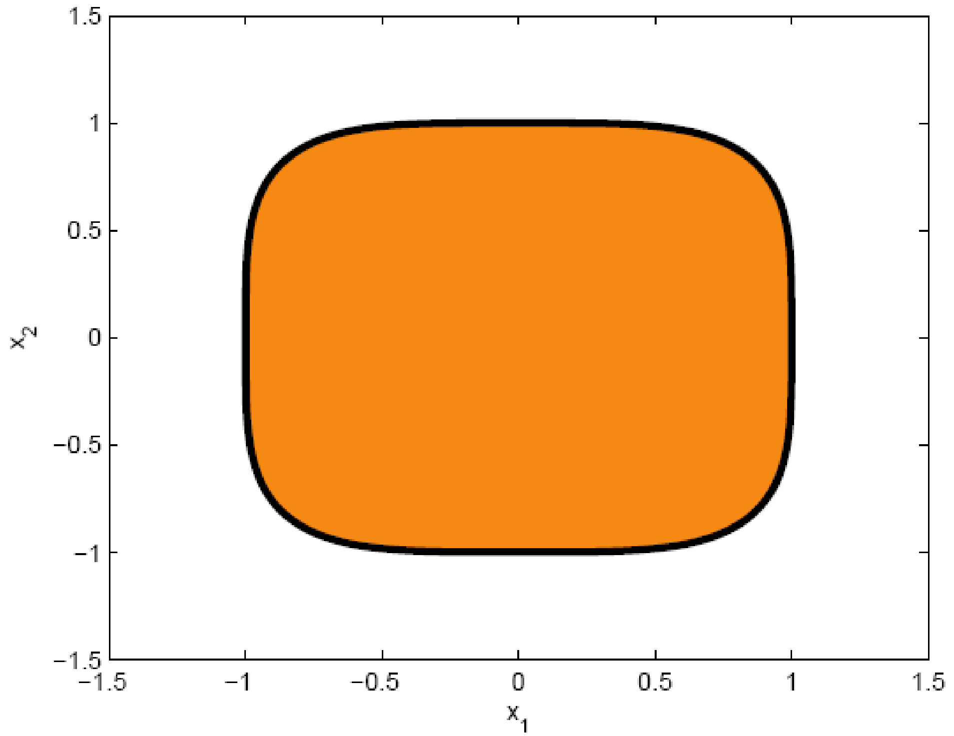
1.1. Definition and Examples

Is the following set convex?

$$x_1^4 + x_2^4 \leq 1$$

1.1. Definition and Examples

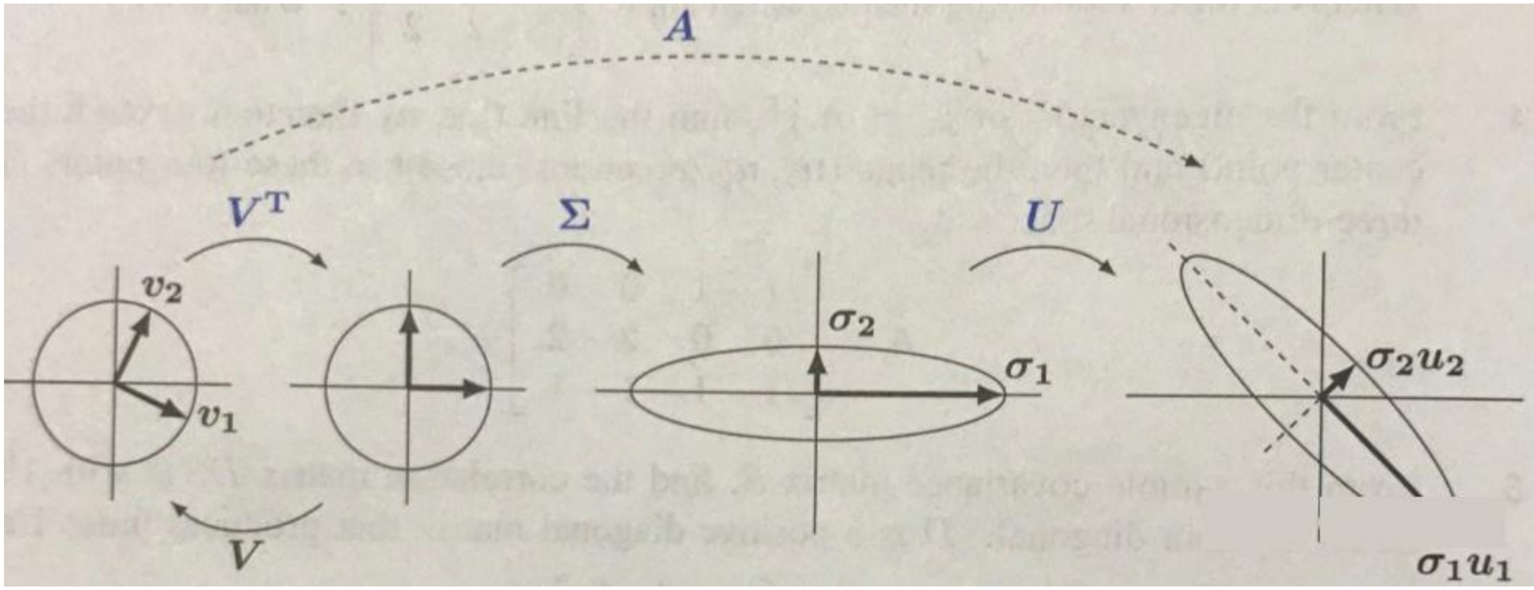
Yes, but cannot be written into a SDP (LMI) form.



1.1. Definition and Examples

Any a PSD matrix A is corresponding to a special ellipse $x^T Ax = I$.

Suppose A can be diagonalized by U as $U^T AU = \Lambda$. We have the ellipse rewritten as the standard ellipse form $y^T \Lambda y = I$, with $y = Ux$.



1.2. Generalized Inequalities

布尔巴基学派将数学结构分为三个元结构，分别是序结构、代数结构和拓扑结构。其中序结构即由集合及在其上规定的序关系组成的数学结构。

特别对于凸集，我们也可以定义序结构，比如对于凸锥，有了射线而不是直线，则有了方向！

Every proper cone K in R^n induces a partial ordering $\geq_K$ defining generalized inequalities on R^n

$$a \geq_K b \Leftrightarrow a-b \in K$$

positive orthant R^n: standard coordinatewise ordering or componentwise inequality

$$x \geq_{R_+^n} y \Leftrightarrow x_i \geq y_i$$

1.2. Generalized Inequalities

We can also define the partial order $A > B$, where matrix $H = A - B \in S^n$ is positive, or equivalently for all $x \in R^n$, $x^T(A-B)x > 0$. This is known as Löwner partial order.

Properties of partial order

- *Additive property*: If $x \preceq y$ and $u \preceq v$, then $x + u \preceq y + v$
- *Transitive property*: If $x \preceq y$ and $y \preceq z$, then $x \preceq z$
- *Non-negative scaling*: If $x \preceq y$, then $\theta x \preceq \theta y$ for all $\theta \geq 0$
- *Reflexive property*: $x \preceq x$
- *Antisymmetric property*: If $x \preceq y$ and $y \preceq x$, then $x = y$

1.2. Generalized Inequalities

$\preceq_K$ is not in general a *linear ordering*: we can have $x \not\preceq_K y$ and $y \not\preceq_K x$

$x \in S$ is **the minimum element** of S with respect to $\preceq_K$ if

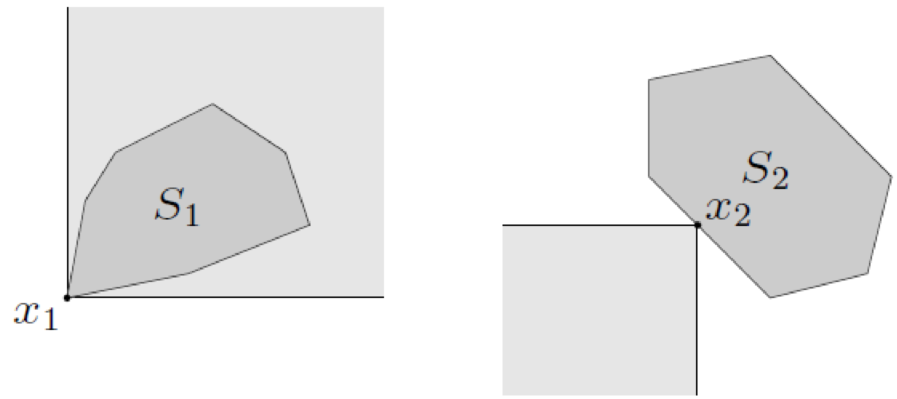
$$y \in S \quad \Longrightarrow \quad x \preceq_K y$$

$x \in S$ is **a minimal element** of S with respect to $\preceq_K$ if

$$y \in S, \quad y \preceq_K x \quad \Longrightarrow \quad y = x$$

example $(K = \mathbf{R}^2_+)$

x_1 is the minimum element of S_1
x_2 is a minimal element of S_2



1.3. Operations that Preserve Convexity

除了上述常见实例，如何分析手头的集合是否凸集？

practical methods for establishing convexity of a set C

1. apply definition

$$x_1, x_2 \in C, \quad 0 \le \theta \le 1 \quad \Longrightarrow \quad \theta x_1 + (1-\theta) x_2 \in C$$

2. show that C is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, . . .) by operations that preserve convexity

   - intersection
   - affine functions
   - perspective function
   - linear-fractional functions

1.3. Operations that Preserve Convexity

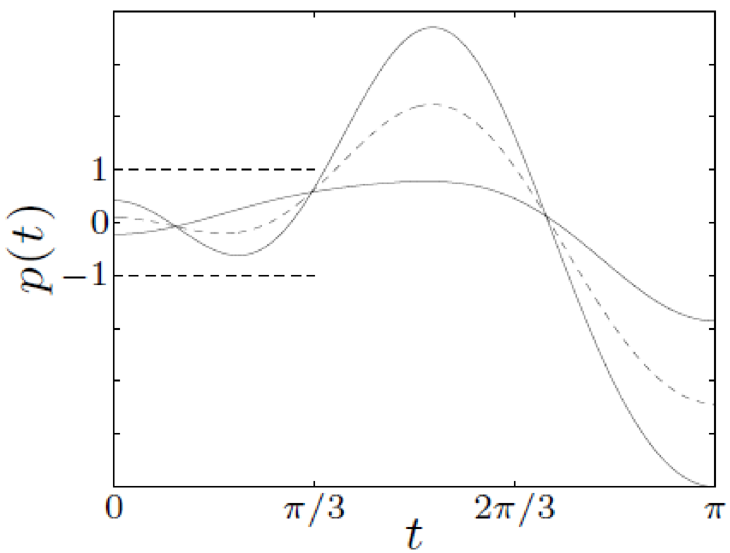
the intersection of (any number of) convex sets is convex

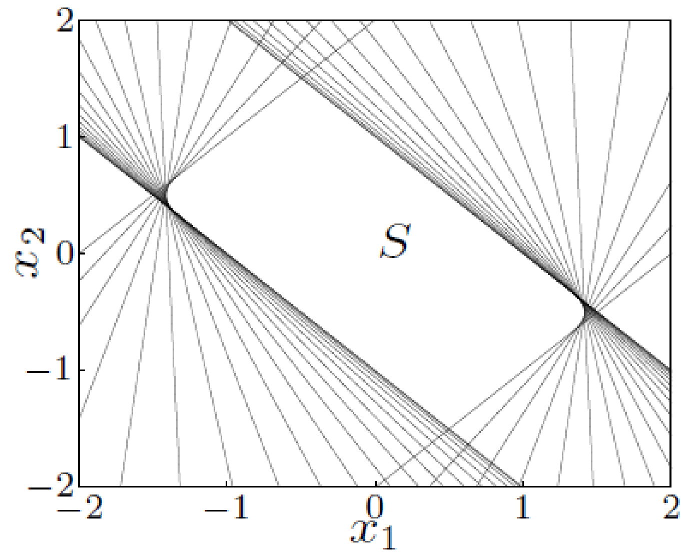
example:

$$S = \{x \in \mathbf{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}$$

where $p(t) = x_1 \cos t + x_2 \cos 2t + \cdots + x_m \cos mt$

for $m = 2$:





1.3. Operations that Preserve Convexity

suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ is affine ($f(x) = Ax + b$ with $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$)

- the image of a convex set under f is convex

$$S \subseteq \mathbf{R}^n \text{ convex} \quad \Longrightarrow \quad f(S) = \{f(x) \mid x \in S\} \text{ convex}$$

- the inverse image $f^{-1}(C)$ of a convex set under f is convex

$$C \subseteq \mathbf{R}^m \text{ convex} \quad \Longrightarrow \quad f^{-1}(C) = \{x \in \mathbf{R}^n \mid f(x) \in C\} \text{ convex}$$

examples

- scaling, translation, projection
- solution set of linear matrix inequality $\{x \mid x_1 A_1 + \cdots + x_m A_m \preceq B\}$ (with $A_i, B \in \mathbf{S}^p$)
- hyperbolic cone $\{x \mid x^T P x \leq (c^T x)^2,\ c^T x \geq 0\}$ (with $P \in \mathbf{S}^n_+$)

1.3. Operations that Preserve Convexity

设 $A\in\mathbb{R}^{m\times n}$，证明如果 $S\subset\mathbb{R}^m$ 是一个凸集，那么 $A^{-1}(S)=\{x\in\mathbb{R}^n : Ax\in S\}$ 也是一个凸集。

证明：我们取 $x, y\in A^{-1}(S)$，那么会有 $Ax\in S, Ay\in S$。取 $t\in[0,1]$，那么会有 $A[tx+(1-t)y]=tAx+(1-t)Ay\in S$（因为 S 是凸的）。这就可以推出 $tx+(1-t)y\in A^{-1}(S)$，也就证明完毕了。

$A^{-1}(S)$ 一般称为原象集（pre-image）。

1.3. Operations that Preserve Convexity

给定$A_1,\ldots,A_k,B\in\mathbb{S}_+^n$，那么所有满足不等式$\sum_{i=1}^k x_iA_i\preceq B$的点构成的集合为凸集。

这个问题要想解释清楚，使用常规的凸集证明方式即可，但是有一种更加有趣的证明方法。

设$f:\mathbb{R}^k\to\mathbb{S}_+^n$，$f(x)=B-\sum_{i=1}^k x_iA_i$那么这个时候我们会发现

$$f^{-1}(\mathbb{S}^n)=\{x: f(x)\in\mathbb{S}^n\}=\left\{x: B-\sum_{i=1}^k x_iA_i\succeq 0\right\}$$

而这个就是我们关注的那个集合。这个集合当然是凸集，因为它是一个凸集在线性映射下的原象集。

1.3. Operations that Preserve Convexity

perspective function $P : \mathbf{R}^{n+1} \to \mathbf{R}^n$:

$$P(x,t) = x/t, \qquad \mathbf{dom}\, P = \{(x,t) \mid t > 0\}$$

images and inverse images of convex sets under perspective are convex

linear-fractional function $f : \mathbf{R}^n \to \mathbf{R}^m$:

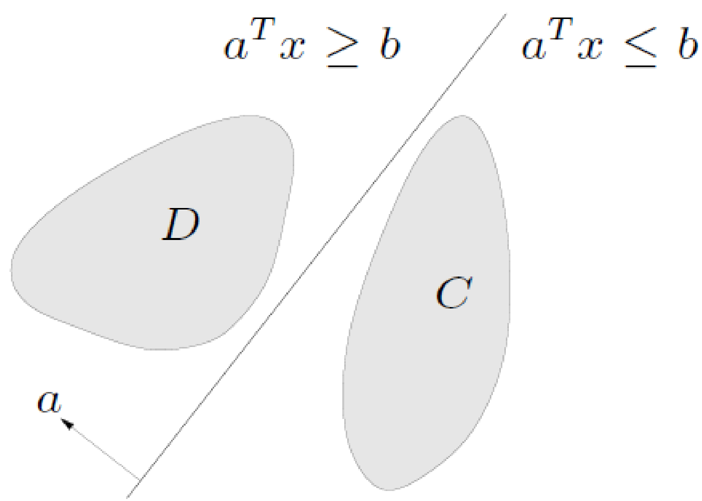
$$f(x) = \frac{Ax + b}{c^T x + d}, \qquad \mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$$

images and inverse images of convex sets under linear-fractional functions are convex

1.4. Separating Hyperplane Theorem

if C and D are disjoint convex sets, then there exists $a \neq 0$, b such that

$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$



the hyperplane $\{x \mid a^T x = b\}$ separates C and D

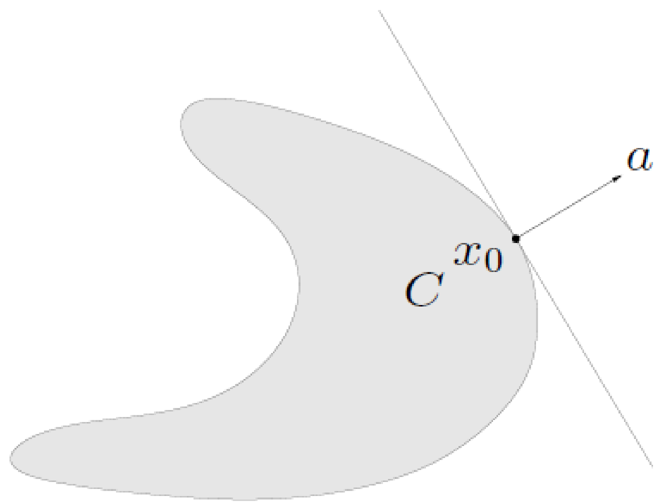
strict separation requires additional assumptions ($e.g.$, C is closed, D is a singleton)

1.4. Separating Hyperplane Theorem

supporting hyperplane to set C at boundary point x_0:

$$\{x \mid a^T x = a^T x_0\}$$

where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$



supporting hyperplane theorem: if C is convex, then there exists a supporting hyperplane at every boundary point of C

1.4. Separating Hyperplane Theorem

Theorem

If Ω be a closed convex set and $\boldsymbol{y}$ be a point outside of Ω, there must exist a hyperplane $\{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} = z\}$ that passes through (contains) $\boldsymbol{y}$ and meanwhile $\Omega \subset \{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} > z\}$.

Proof.

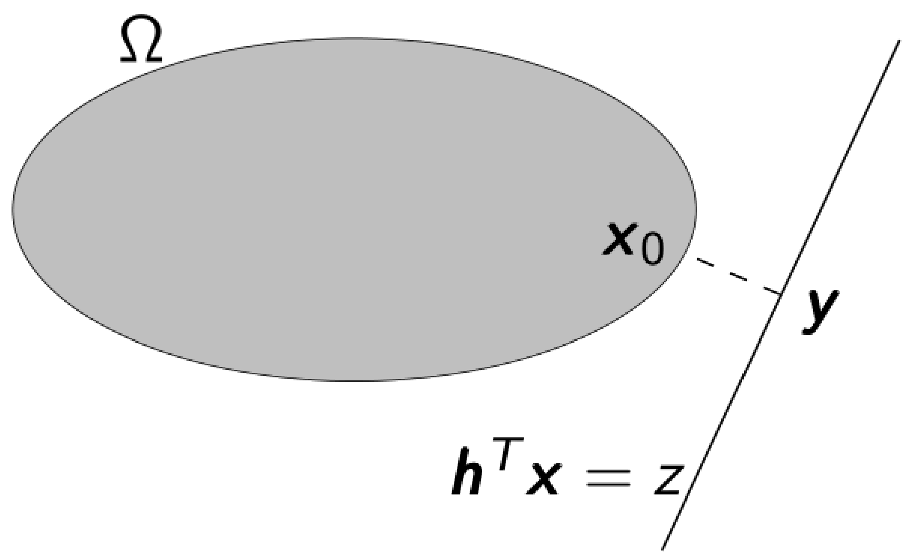


Figure: A hyperplane that does not intersect with set Ω.

1.4. Separating Hyperplane Theorem

Let us define $D = \inf_{x \in \Omega} |\boldsymbol{x} - \boldsymbol{y}|_2 > 0$ as the distance between Ω and $\boldsymbol{y}$. Since Ω is a closed set, there must exist a point $\boldsymbol{x}_0$ such that the continuous function $F(\boldsymbol{x}) = |\boldsymbol{x} - \boldsymbol{y}|_2$, $\boldsymbol{x} \in \Omega$ reaches its minimum.

Next, we use contradiction method to show that this point must be a boundary point of Ω, that is, $\boldsymbol{x}_0 \in \partial\Omega$. Suppose $\boldsymbol{x}_0 \in \Omega^o$, thus, there exists $2\epsilon > 0$ such that the open ball $B_{2\epsilon}(\boldsymbol{x}_0) \subset \Omega$. We can easily find a point $\boldsymbol{x}_1 = \boldsymbol{x}_0 + \epsilon \frac{\boldsymbol{y} - \boldsymbol{x}_0}{|\boldsymbol{y} - \boldsymbol{x}_0|_2} \in B_{2\epsilon}(\boldsymbol{x}_0) \subset \Omega$ and

$$\begin{aligned} F(\boldsymbol{x}_1, \boldsymbol{y}) &= \left|\boldsymbol{x}_0 + \epsilon \frac{\boldsymbol{y} - \boldsymbol{x}_0}{|\boldsymbol{y} - \boldsymbol{x}_0|_2} - \boldsymbol{y}\right|_2 \\ &= \left(1 - \frac{\epsilon}{|\boldsymbol{x}_0 - \boldsymbol{y}|_2}\right) |\boldsymbol{x}_0 - \boldsymbol{y}|_2 < |\boldsymbol{x}_0 - \boldsymbol{y}|_2 \end{aligned}$$

which contradicts with $f(\boldsymbol{x}_0, \boldsymbol{y}) = \inf_{\boldsymbol{x} \in \Omega} |\boldsymbol{x} - \boldsymbol{y}|_2$.

1.4. Separating Hyperplane Theorem

Define $\boldsymbol{h} = \boldsymbol{x}_0 - \boldsymbol{y}$, we select the hyperplane $\{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} = z\}$ with $\boldsymbol{h}^T\boldsymbol{y} = z$, and we claim to show that $\Omega \subset \{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} > z\}$.
For all $\boldsymbol{x} \in \Omega$ and $\lambda \in (0, 1)$, by the convexity of Ω, we have

$$\bar{\boldsymbol{x}} = \boldsymbol{x}_0 + \lambda(\boldsymbol{x} - \boldsymbol{x}_0) \in \Omega$$

Note that $F(\bar{\boldsymbol{x}}, \boldsymbol{y}) \geq F(\boldsymbol{x}_0, \boldsymbol{y})$, i.e.
$(\bar{\boldsymbol{x}} - \boldsymbol{y})^T(\bar{\boldsymbol{x}} - \boldsymbol{y}) - (\boldsymbol{x}_0 - \boldsymbol{y})^T(\boldsymbol{x}_0 - \boldsymbol{y}) \geq 0$. We have

$$2\lambda(\boldsymbol{x}_0 - \boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{x}_0) + \lambda^2(\boldsymbol{x} - \boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) \geq 0$$

By eliminating the common multiplier λ at the left side and substituting $\boldsymbol{x}_0 - \boldsymbol{y} = \boldsymbol{h}$, we have

$$\boldsymbol{h}^T(\boldsymbol{x} - \boldsymbol{x}_0) + \lambda(\boldsymbol{x} - \boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) \geq 0$$

1.4. Separating Hyperplane Theorem

Finally, we show that $\boldsymbol{h}^T(\boldsymbol{x}-\boldsymbol{x}_0) \geq 0$ by contradiction. Suppose $\boldsymbol{h}^T(\boldsymbol{x}-\boldsymbol{x}_0) < 0$, we can always find a small enough $\lambda \to 0$, such that $\boldsymbol{h}^T(\boldsymbol{x}-\boldsymbol{x}_0) + \lambda(\boldsymbol{x}-\boldsymbol{x}_0)^T(\boldsymbol{x}-\boldsymbol{x}_0) < 0$, which leads to a contradiction.
Then, we have

$$\boldsymbol{h}^T\boldsymbol{x} = \boldsymbol{h}^T(\boldsymbol{x}-\boldsymbol{x}_0) + \boldsymbol{h}^T(\boldsymbol{x}_0-\boldsymbol{y}) + \boldsymbol{h}^T\boldsymbol{y} \geq \boldsymbol{h}^T\boldsymbol{h} + z > z$$

Therefore $\Omega \subset \{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} > z\}$. This completes the proof.

1.4. Separating Hyperplane Theorem

Theorem

If Ω be a closed convex set and $\boldsymbol{y}$ be a point outside of Ω, there must exist a hyperplane $\{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} = z\}$ that separates $\boldsymbol{y}$ and Ω, i.e. $\Omega \subset \{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} > z\}$ and $\boldsymbol{h}^T\boldsymbol{y} < z$.

Proof.

Similar to the proof for Theorem 2, we can find a boundary point $\boldsymbol{x}_0$ such that $\inf_{\boldsymbol{x}\in\Omega} |\boldsymbol{x} - \boldsymbol{y}|_2$ reaches its minimum at $\boldsymbol{x}_0$. It is easy to further shown that $H = \{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} = z\}$, with $\boldsymbol{h} = \boldsymbol{x}_0 - \boldsymbol{y}$, $z = \boldsymbol{h}^T(\frac{1}{2}\boldsymbol{y} + \frac{1}{2}\boldsymbol{x}_0)$ is a hyperplane that separates $\boldsymbol{y}$ and Ω. $\square$

1.4. Separating Hyperplane Theorem

Definition

If $\boldsymbol{y}$ be a boundary point of a convex set Ω, we call $\Gamma = \{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} = z\}$ is called a *supporting hyperplane* of Ω at $\boldsymbol{y}$ if $\boldsymbol{y} \in \Gamma$ and $\Omega \subseteq \{\boldsymbol{x} \mid \boldsymbol{h}^T\boldsymbol{x} \geq z\}$.

Theorem

If $\boldsymbol{y}$ is a boundary point of a closed convex set C, there exists at least one supporting hyperplane at $\boldsymbol{y}$.

Proof.

Define $\boldsymbol{y}_0$ as point outside of Ω and $\boldsymbol{y}_n = \boldsymbol{y} + \frac{1}{2^n}(\boldsymbol{y}_0 - \boldsymbol{y})$. Then $\boldsymbol{y}_n \notin \Omega$ and $\lim_{n\to\infty} \boldsymbol{y}_n = \boldsymbol{y}$.

By Theorem 3, there exists a sequence of vectors $\boldsymbol{h}_n, n \in \mathbb{N}$ with $|\boldsymbol{h}_n| = 1$ such that

$$C \subseteq \{\boldsymbol{x} \mid \boldsymbol{h}_n^T\boldsymbol{x} > \boldsymbol{h}_n^T\boldsymbol{y}_n\}, n \in \mathbb{N}$$

1.4. Separating Hyperplane Theorem

Since $\boldsymbol{h}_n$ is bounded, there exists a convergent subsequence $\{\boldsymbol{h}_{n_k}, k \in \mathbb{N}\}$ in $\{\boldsymbol{h}_n, n \in \mathbb{N}\}$ and a vector $\boldsymbol{h}$ such that $\lim_{k\to\infty} \boldsymbol{h}_{n_k} = \boldsymbol{h}$. Therefore for all $\boldsymbol{x} \in \Omega$, we have

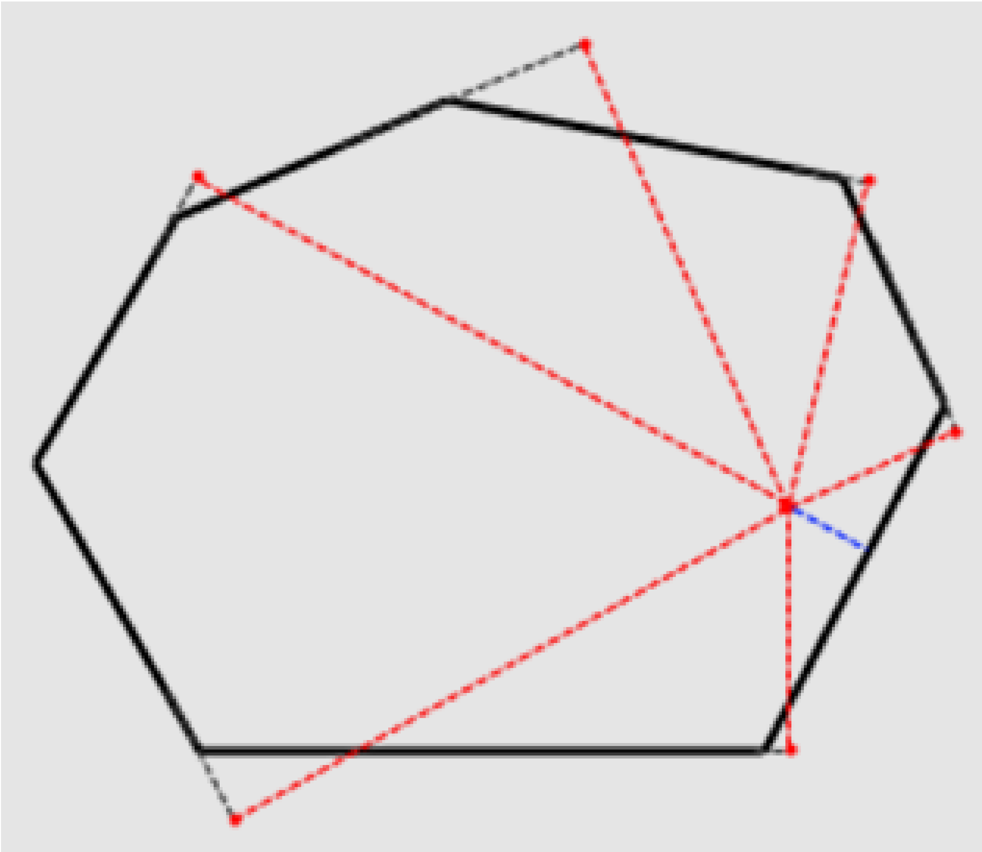
$$\boldsymbol{h}_{n_k}^T \boldsymbol{x} > \boldsymbol{h}_{n_k}^T \boldsymbol{y}_{n_k}$$

Let $k \to \infty$, we have $\boldsymbol{h}^T \boldsymbol{x} \geq \boldsymbol{h}^T \boldsymbol{y}$. Therefore there exists at least one supporting hyperplane at $\boldsymbol{y}$.

1.5. Some Quiz Problems

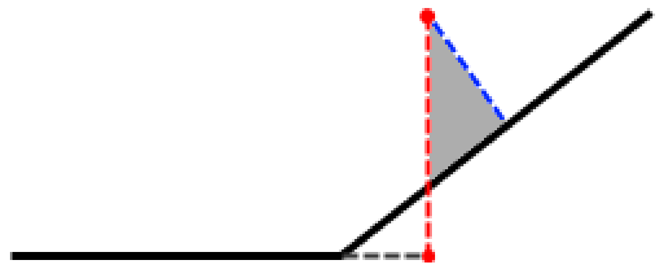
For any a convex $2n$ -sided polyhedron in a R^2 plane. Suppose we color any n vertices into red, the rest n vertices into blue. Please prove or disprove there must be a point on the plane so that the sum of its distance from all the red points is equal to the sum of its distance from all the blue points.

1.5. Some Quiz Problems



At least one perpendicular foot of a point inside a convex polygon to each edge of this convex polygon falls on an edge of this convex polygon.

1.5. Some Quiz Problems

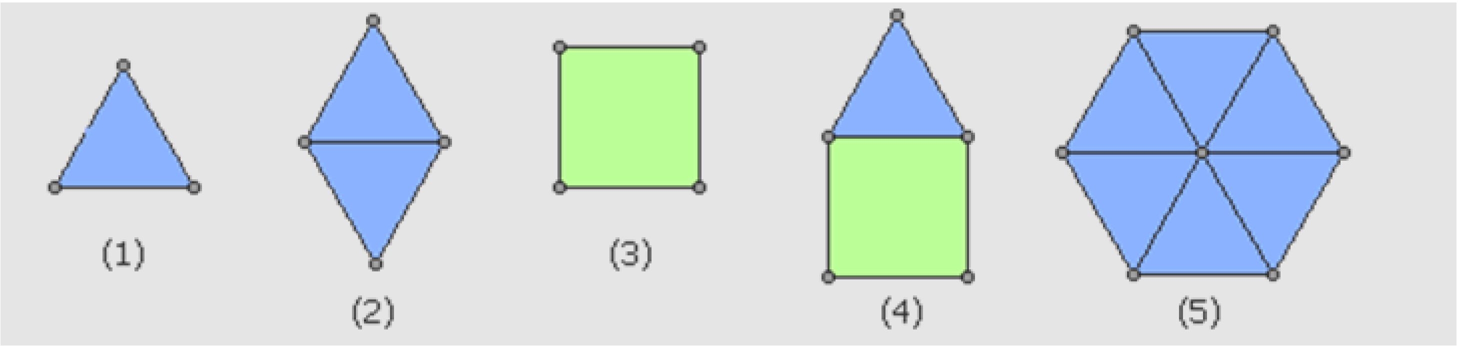


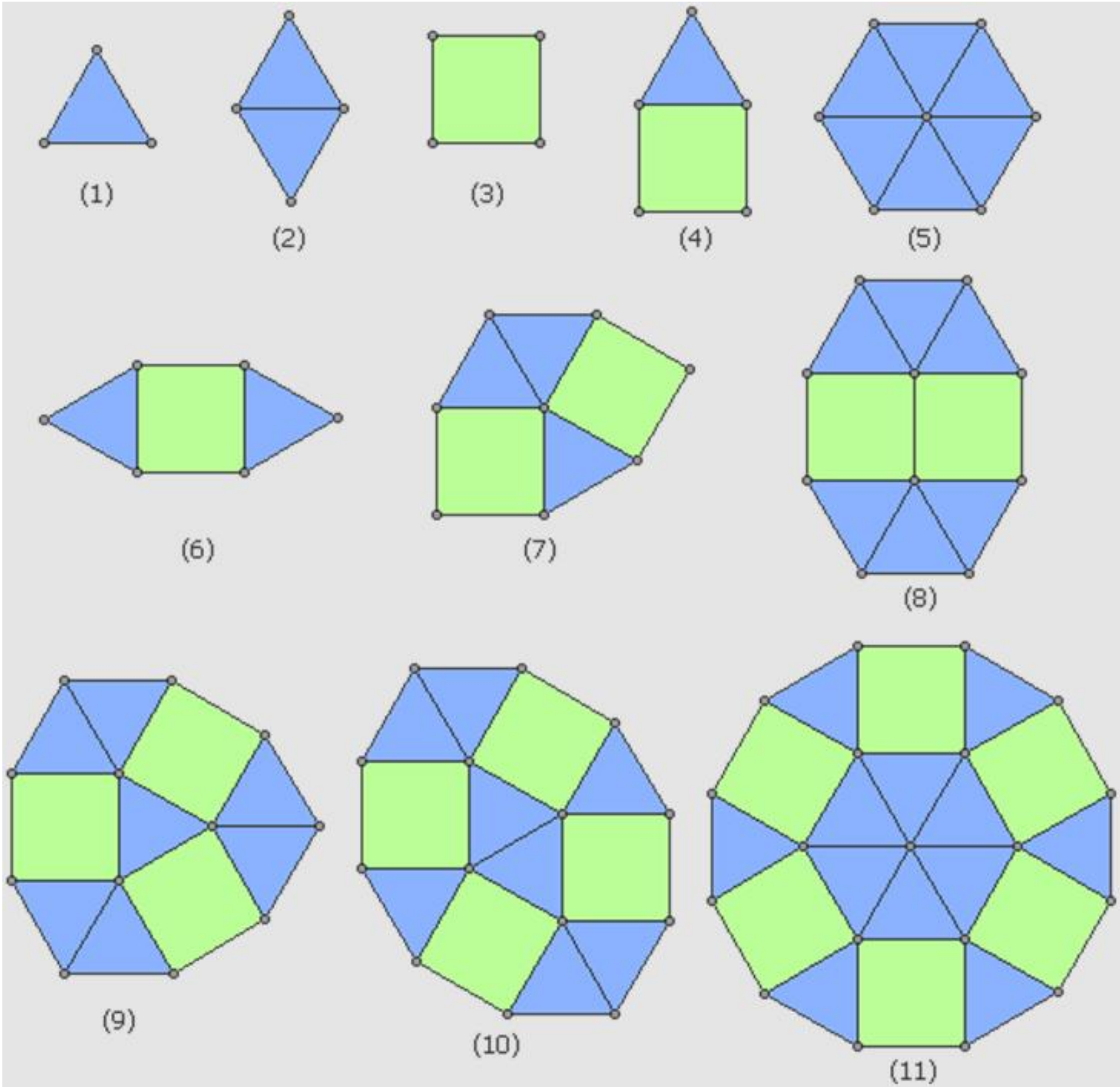
只需要证明离给定点最近的那条边的确满足投影点在边上的要求就可以了。假设上图中的红色虚线是给定点到所有边的垂线段中最短的一个，但垂足却在边的外面。我们立即发现，存在某个边对应的蓝色垂足（请问为什么？），由于灰色直角三角形中斜边大于直角边，蓝色垂线段显然要比红色线段更短（注意：蓝线的垂线不一定要求落在边上），这就与红色线段是所有垂线段中最短相矛盾。

这个命题对更高维的情形也都是成立的：对于给定凸多面体和它内部的一点，总能找到其中一个面使得，给定点在这个面上的投影恰好就落在这个面上。

1.5. Some Quiz Problems

Suppose you are given a collection of squares and equilateral triangles (of unit side). You are asked to form convex polygons by sticking the squares and triangles together with their sides aligned. These polygons must also have unit sides (so that for example sticking two squares together to form a domino does not count). How many distinct polygons can be formed in this way (including polygons formed by a single square or triangle)?





(1)
(2)
(3)
(4)
(5)
(6)
(7)
(8)
(9)
(10)
(11)

1.6. Concluding Remarks

了解基本概念

形成足够直觉

完成简单题目（需要综合分析，代数和拓扑的基础知识）

1.7. References

[1] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. http://www.stanford.edu/~boyd/cvxbook/ http://www.ee.ucla.edu/~vandenbe/cvxbook

[2] D. Bertsimas, J. N. Tsitsiklis, *Introduction to Linear Optimization*, Athena Scientific, 1997.

[3] http://domino.research.ibm.com/Comm/wwwr_ponder.nsf/Challenges/April2008.html