Elements of convex analysis and convex optimization part 1 - Convex sets



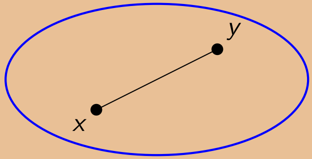


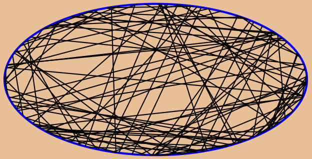
March 22, 2024

Introduction

Convexity is easy to define, to visualize and to get an intuition about.

A set is convex if for every two points x and y in the set, the straight line interval $[x, y]$ is also in the set. Thus the main building block of convexity theory is a straight line interval.





Convexity is more intuitive than linear algebra. In linear algebra, the interval is replaced by the whole straight line.
On the other hand, the structure of convexity is richer than the one in linear algebra. It is already evident in the fact that all points on the line are alike whereas the interval has two points, x and y, which clearly stand out.

Introduction

Indeed, convexity has an immensely rich structure and numerous applications.

Another important observation (that we used before in the multivariable calculus section) is that almost every idea in convexity can be explained by a 2D picture.

There is a reason for that (apart from the obvious one that all drawings live in a 2D world):

"One possible explanation is that since the definition of a convex set involves only three points (the two points x and y and a point z contained in the interval) and every three points lie in some plane, whenever we invoke a convexity argument in our reasoning, it can be properly pictured" (taken from [1]).

[1] Barvinok, A. (2002). *A Course in Convexity.* Providence, RI: American Mathematical Society.

Part I - Convex sets

Convex sets

Definition

A set $X \subseteq \mathbb{R}^n$ is called convex if for all $x, y \in X$ we have

$$\alpha x + (1-\alpha) y \in X , \quad \forall \alpha \in [0,1] .$$

Examples: (one convex, two nonconvex sets)

Examples

The following sets are convex:

- *The interval* $[0,1] \subset \mathbb{R}$ *(intervals are all possible convex sets in* $\mathbb{R}$*).*
- *A rectangle in* $\mathbb{R}^2$ *(cartesian product of two intervals).*
- $\emptyset$ *and* $\mathbb{R}^n$*.*
- *A line in* $\mathbb{R}^n$*.*
- *Hyperplanes and halfspaces.*
- *Norm balls.*
- *Ellipsoids.*
- *Polyhedrons and simplexes.*
- *Convex cones.*
- *A vector subspace.*

Convex sets: Line in $\mathbb{R}^n$

A line in $\mathbb{R}^n$ is a set of the form

$$L(p,v) = \{p + tv \in \mathbb{R}^n : t \in \mathbb{R}\},$$

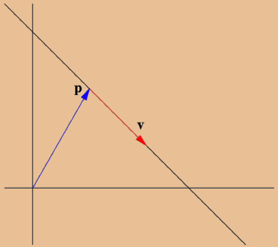
where $p \in \mathbb{R}^n$ is the offset form the origin, and $v \in \mathbb{R}^n$, $v \neq 0$.

To show it is convex, take $x, y \in L(p,v)$. Then we can write

$$x = p + t_1 v$$
$$y = p + t_2 v$$

We now show that $\alpha x + (1-\alpha)y \in L$ for any $\alpha \in [0,1]$. Indeed,

$$\alpha x + (1-\alpha)y = \alpha(p + t_1 v) + (1-\alpha)(p + t_2 v) = p + (\alpha t_1 + (1-\alpha)t_2)v \quad \in L.$$



Convex sets: Hyperplane

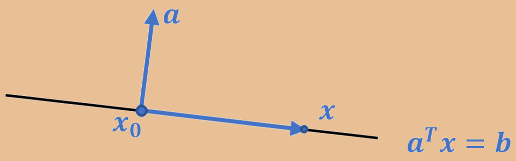
Given a normal vector $a \in \mathbb{R}^n$, $a \neq 0$, and a point $x_0 \in \mathbb{R}^n$, a hyperplane is the set of points

$$\mathcal{H}(a, x_0) = \{x \in \mathbb{R}^n : a^\top (x - x_0) = 0\} .$$

Denoting $a^\top x_0$ with b we can also write the hyperplane as

$$\mathcal{H}(a, b) = \{x \in \mathbb{R}^n : a^\top x = b\} ,$$

where $a \neq 0 \in \mathbb{R}^n, b \in \mathbb{R}$.



Convex sets: Hyperplane

In other words

$$a^\top x = b \iff a \perp (x - x_0)\,,$$

where $x_0 = \frac{b}{a^\top a} a$.

Analytical interpretation of the hyperplane: solution set of a nontrivial linear equation.

To show that a hyperplane is convex, take $z = \alpha x + (1{-}\alpha) y$ for $x, y \in \mathcal{H}(a, b)$, and $\alpha \in [0, 1]$. Then $a^\top z = b$ because $a^\top(\alpha x + (1{-}\alpha)y) = \alpha a^\top x + (1{-}\alpha)a^\top y = \alpha b + (1{-}\alpha)b = b$.

Convex sets: Hyperplane

Another way to visually represent a hyperplane:

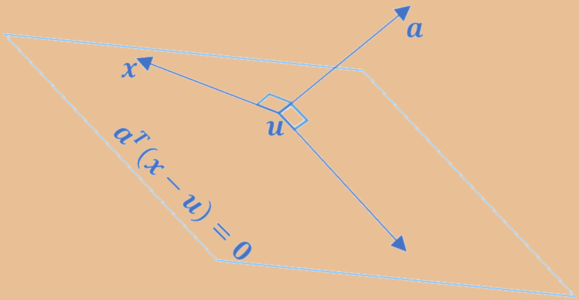


Figure: The hyperplane $\mathcal{H}(a, u) = \{x \in \mathbb{R}^n : a^\top(x-u) = 0\}$

Convex sets: Halfspace

Halfspace is either of the two parts into which a hyperplane divides the Euclidean space.

- a hyperplane divides $\mathbb{R}^n$ into two halfspaces.
- the boundary of a halfspace is a hyperplane.

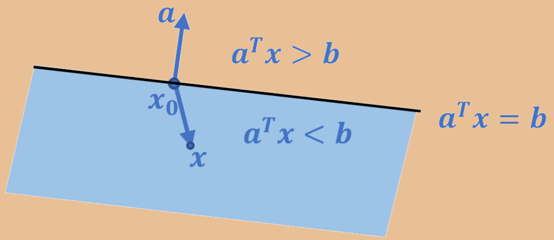
Given a normal vector $a \in \mathbb{R}^n$, $a \neq 0$, and a point $x_0 \in \mathbb{R}^n$ we have the halfspace

$$\mathcal{H}_-(a, x_0) = \{x \in \mathbb{R}^n : a^\top (x - x_0) \leq 0\},$$

or

$$\mathcal{H}_-(a, b) = \{x \in \mathbb{R}^n : a^\top x \leq b\},$$

where $a \neq 0 \in \mathbb{R}^n$, $b \in \mathbb{R}$.



Convex sets: Polyhedron

Solution set of a finite number of linear inequalities and equalities (can be bounded or unbounded)

$$\mathcal{P} = \{x \in \mathbb{R}^n : Ax \leq b, \;\; Cx = d\}.$$

where $\leq$ denotes componentwise inequality between vectors.

Intersection of finite number of halfspaces and hyperplanes (defined by the normal vectors a_i).

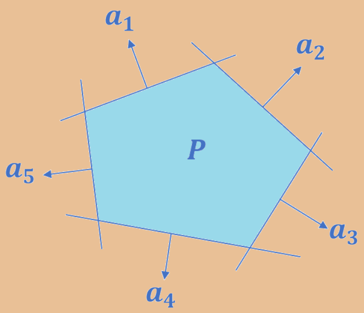
$$\begin{aligned}\mathcal{P} &= \{x : a_j^\top x \leq b_j, \;\; j{=}1,\ldots,m, \; c_j^\top x = d_j, \;\; j{=}1,\ldots,p\} \\ &= \{x : Ax \leq b, \;\; Cx = d\}\end{aligned}$$

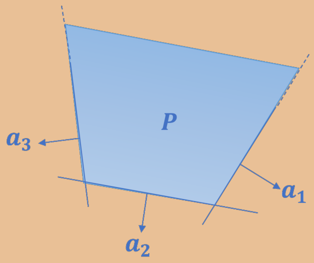
where

$$A = \begin{bmatrix} a_1^\top \\ \vdots \\ a_m^\top \end{bmatrix}, \;\; C = \begin{bmatrix} c_1^\top \\ \vdots \\ c_m^\top \end{bmatrix}, \quad a_j, c_j \in \mathbb{R}^n, \;\; b_j, d_j \in \mathbb{R}.$$

Using this latter fact, we can <u>show it is convex</u> using the convexity of hyperplanes and halfspaces.

Convex sets: Polyhedron





Without losing generality we can assume a polyhedron is a set

$$\mathcal{P} = \{x \in \mathbb{R}^n : Ax \leq b\}.$$

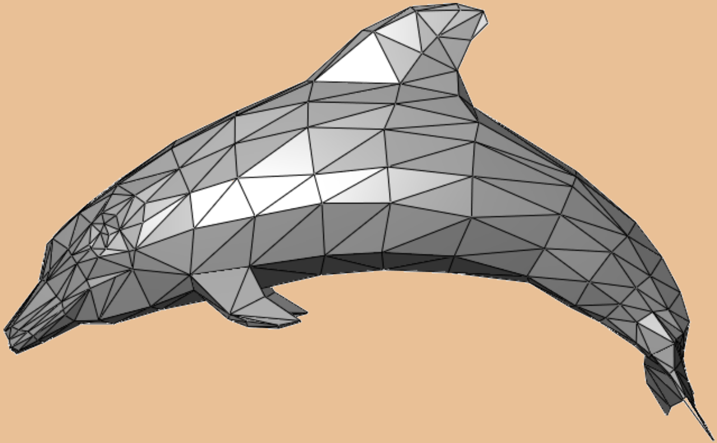
- Special cases of polyhedra: subspaces, hyperplanes, lines, halfspaces.
- Unit simplex: important polyhedron:

$$\Delta_n := \{w \in \mathbb{R}^n : w_i \geq 0, \, \mathbf{1}^\top w = 1\},$$

where $\mathbf{1}$ is a vector of ones so $\mathbf{1}^\top w = \sum_{i=1}^n w_i$.

Convex sets: Polyhedron

Polyhedra is essential in modern computational geometry, computer graphics, and CAD. Some applications: reconstruction of polyhedral surfaces or surface meshes from scattered data points, geodesics on polyhedral surfaces, visibility and illumination in polyhedral scenes, etc.



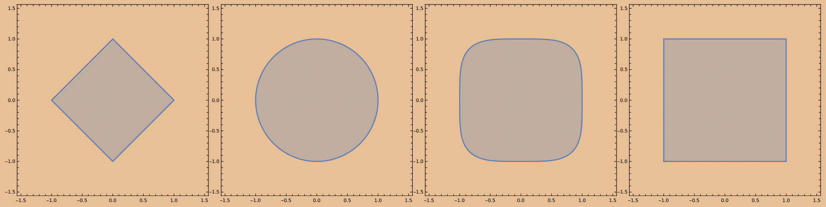
Convex sets: Norm ball

Norm ball with center $x_c \in \mathbb{R}^n$ and radius $r > 0$:

$$B[x_c, r] = \{x \in \mathbb{R}^n : \|x - x_c\| \leq r\},$$

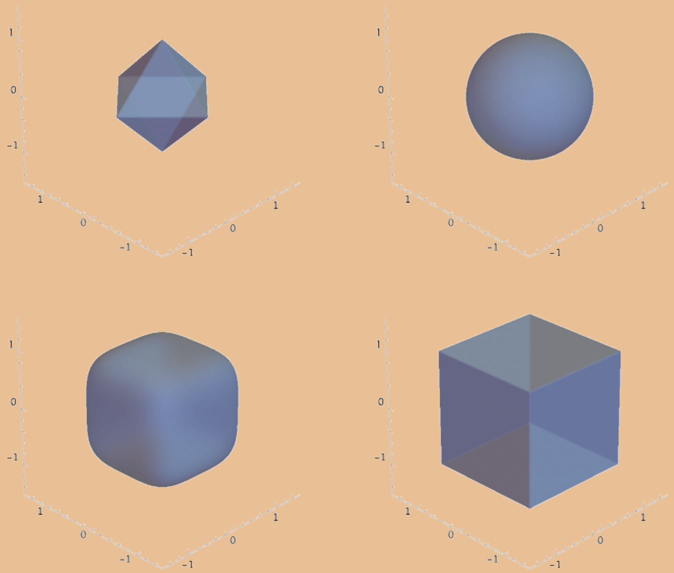
where $\|\cdot\|$ is a norm.

1-norm, 2-norm, 4-norm and infinite norm in $\mathbb{R}^2$:



Convex sets: Norm ball

1-norm, 2-norm, 4-norm and infinite norm in $\mathbb{R}^3$:



Convex sets: Ellipsoid

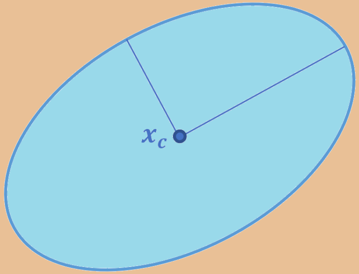
A generalization of the 2-norm ball

$$B[x_c, r] = \{x \in \mathbb{R}^n : \|x - x_c\|_2 \leq r\}$$

is the **Ellipsoid:**

$$\mathcal{E} = \{x \in \mathbb{R}^n : (x - x_c)^\top P^{-1}(x - x_c) \leq 1\}$$

with P^{-1} a pd matrix (and therefore P is pd).



When $P = r^2 I$, the ellipsoid becomes a ball.

Side note: Principal axes

Eigendecomposition: Since P is pd, $P = \mathcal{Q}\Lambda\mathcal{Q}^\top = \sum_{i=1}^n \lambda_i q_i q_i^\top$.

- $\mathcal{Q}$ is orthogonal ($\mathcal{Q}^\top = \mathcal{Q}^{-1}$) with columns v_i.
- Λ is diagonal with diagonal elements $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$.
- eigenvectors q_i of P form the principal axes of $\mathcal{E}$.
- the width of $\mathcal{E}$ along the principal axis determined by q_i is $2\sqrt{\lambda_i}$.
- We have $P^{-1} = \mathcal{Q}\Lambda^{-1}\mathcal{Q}^\top$.

Change of variables: $y = \mathcal{Q}^\top(x - x_c)$, $x = x_c + \mathcal{Q}y$.

- after the change of variables the ellipsoid is described by

$$y^\top \Lambda^{-1} y = y_1^2/\lambda_1 + \cdots + y_n^2/\lambda_n \leq 1$$

  that is an ellipsoid centered at the origin, and aligned with the coordinate axes.

Side note: Principal components

PCA relies on the eigendecomposition of a data covariance (or correlation) matrix of a standardized data set to identify the principal directions along which the data varies the most.

For a dataset with n variables, the covariance matrix C is an $n \times n$ matrix where each element C_{ij} represents the covariance between variable i and variable j.

By decomposing C, the eves represent the directions of the axes along which the data varies, while the evas give the magnitude of these variances.

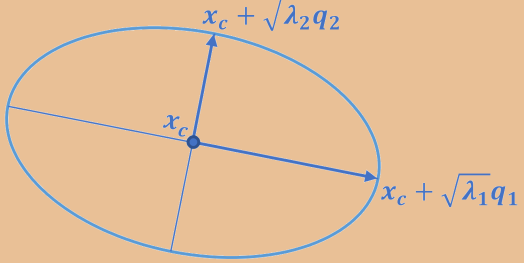
The eves of the covariance matrix correspond to the principal components (PCs) of the dataset, with the first principal component being the direction along which the data varies the most, the second principal component being the direction of the next highest variance orthogonal to the first, and so on.

The evas indicate the variance captured by each principal component. Sorting the eigenvectors by their corresponding eigenvalues in descending order prioritizes the principal components that capture the most variance.

PCA reduces the dimensionality of the data by projecting it onto the first few principal components.

Side note: Principal axes

Example (in $\mathbb{R}^2$)



This corresponds with the representation: $\{x_c + Au : \|u\|_2 \leq 1\}$ with A pd matrix.

Note: an ellipsoid can also be written as

$$\{x \in \mathbb{R}^n : x^\top Qx + 2b^\top x + c \leq 0\}.$$

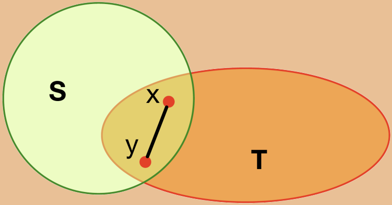
Task: Prove by definition that an Ellipsoid is a convex set.

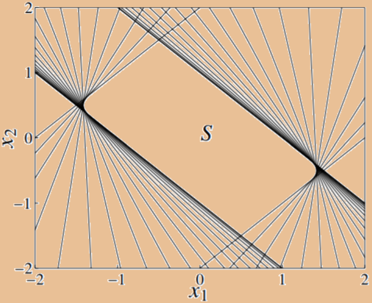
Operations that preserve set convexity

Useful to prove that a given set is convex, or to construct complex sets from simple ones.

Proposition (Intersection of convex sets is convex)

Let I be an arbitrary index set. Let $C_i \subseteq \mathbb{R}^n$, $i \in \mathcal{I}$ a family of convex sets, then the set $\bigcap_i C_i$ is convex.





Note: The union in general is not convex.

Right figure: Boyd and Vandenberghe. Convex Optimization. (2004) Cambridge University Press.

Intersection

Example

A polyhedron is the intersection of halfspaces and hyperplanes (which are convex), and so is convex:

$$\begin{aligned}\mathcal{P} &= \{x : a_i^\top x \le b_i,\ i{=}1,\dots,m,\ \ c_j^\top x = d_j,\ j{=}1,\dots,p\} \\ &= \left(\bigcap_{i=1}^{m}\{x : a_i^\top x \le b_i\}\right) \cap \left(\bigcap_{j=1}^{p}\{x : c_j^\top x = d_j\}\right).\end{aligned}$$

Here

- $\{x : a_i^\top x \le b_i\}$ *is a halfspace;*
- $\{x : c_j^\top x = d_j\}$ *is a hyperplane.*

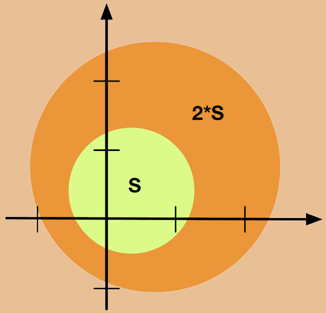
Operations that preserve set convexity

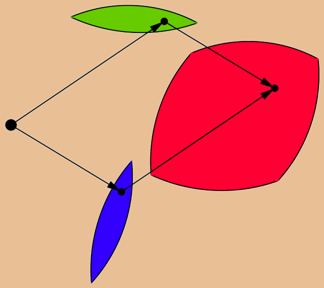
Let $\alpha \in \mathbb{R}$, $X, Y \subseteq \mathbb{R}^n$ then we define

$$\alpha X := \{\alpha x : x \in X\}$$

and (Minkowski sum)

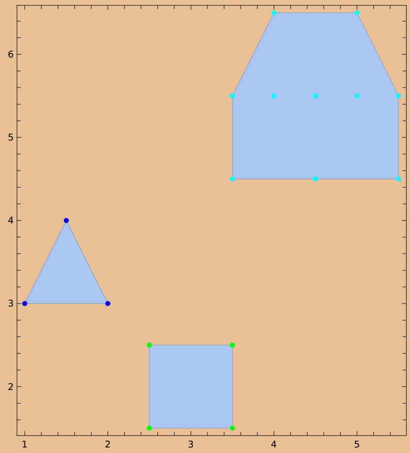
$$X + Y := \{x+y : x \in X, \ \ y \in Y\} \tag{1}$$

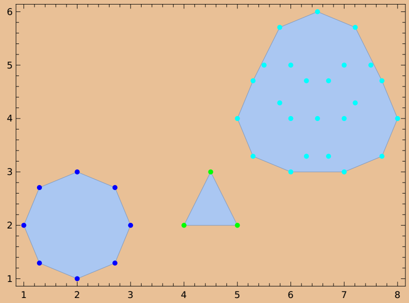




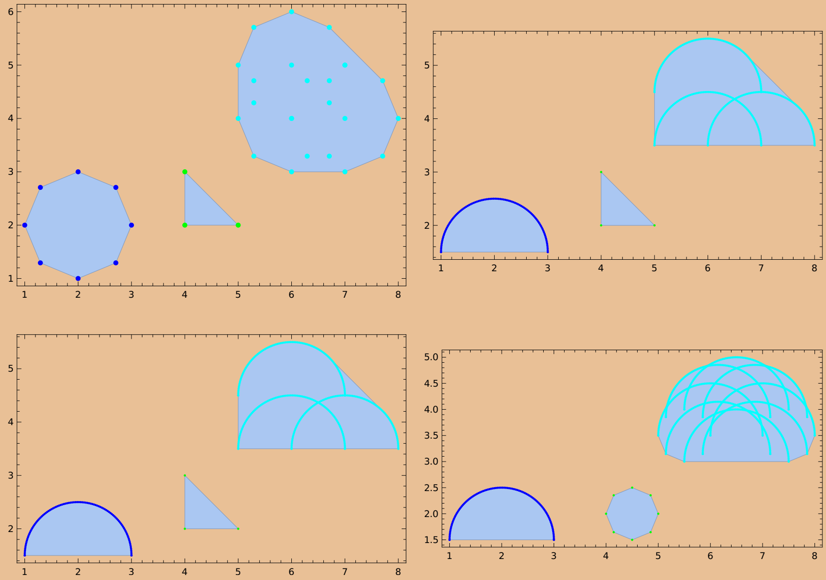
Minkowski sum

Examples:





Minkowski sum



Operations that preserve set convexity

Proposition (Linear combination of convex sets is convex)

Let $C_1, C_2, ..., C_k$ *convex sets in* $\mathbb{R}^n$ *and* $\alpha_1, ..., \alpha_k \in \mathbb{R}$, *then the set* $\alpha_1 C_1 + \alpha_2 C_2 + ... + \alpha_k C_k$ *is convex.*

In particular, if $C \subseteq \mathbb{R}^n$ convex and $b \in \mathbb{R}^n$ then the translation

$$C + b = \{x + b : x \in C\} \tag{2}$$

is convex.

Proposition (Cartesian product of convex sets is convex)

Let $C_1, C_2, ..., C_k$ *convex sets in* $\mathbb{R}^n$, *then the set*

$$C_1 \times C_2 \times ... \times C_k = \{(x_1, x_2, ..., x_k) : x_i \in C_i\}$$

is convex.

Operations that preserve set convexity

Proposition (Linear transformation is convex)

Let $C \subseteq \mathbb{R}^n$ convex, and $A \in \mathbb{R}^{m \times n}$. Then

$$A(C) = \{Ax \in \mathbb{R}^m : x \in C\}$$

is convex.

Let $D \subseteq \mathbb{R}^m$ convex, and $A \in \mathbb{R}^{m \times n}$. Then

$$A^{-1}(D) = \{x \in \mathbb{R}^n : Ax \in D\}$$

is convex.

Proof of i): Let $y_1, y_2 \in A(C)$ and $\lambda \in [0, 1]$. Then, there exists $x_1, x_2 \in C$ such that:

$$y_1 = Ax_1\,, \quad \text{and} \quad y_2 = Ax_2\,.$$

Then, we have:

$$\lambda y_1 + (1-\lambda) y_2 = \lambda Ax_1 + (1-\lambda)Ax_2 = A(\lambda x_1 + (1-\lambda)x_2)\,.$$

Since C is convex, we have:

$$y_3 \equiv \lambda x_1 + (1-\lambda)x_2 \in C\,,$$

and thus $Ay_3 \in A(C)$. So $A(C)$ is convex.

Operations that preserve set convexity

Intuitively, a linear transformation maps line segments: Consider a line segment between two points in C. Under the linear transformation, each point on this line segment is mapped to a new location. Importantly, because linear transformations are linear, the line segment connecting two points in C is transformed into another line segment connecting the corresponding points in $A(C)$.

Since all line segments within C are transformed into line segments within $A(C)$, the set $A(C)$ maintains the property of convexity.

Online resource to interact with linear transformations: `https://yizhe-ang.github.io/matrix-explorable//`

Example - Linear transformation of convex sets

Example 1: i) Consider $n = m = 1$, i.e., a transformation from $\mathbb{R} \to \mathbb{R}$.

Let $C \subseteq \mathbb{R}$ be the interval $[0, 2]$, a convex set.

Let A be the 1×1 matrix $A = 3$ (a real number). Then, $Ax = 3x$.

Then the transformed set is $A(C) = \{3x : x \in [0, 2]\} = [0, 6]$. The transformed set is an interval and thus convex.

ii) Let D be the interval $[0, 9]$, a convex set.

Let $A = 3$.

Then inverse image set $A^{-1}(D)$: is the set of x such that $3x \in [0, 9]$, so we have $A^{-1}(D) = [0, 3]$. This set is again convex.

Example 2: Let B be a 2-norm ball with radius $r = 1$, centered at the origin: $B = \{x \in \mathbb{R}^2 : \|x\| = 1\}$.

Transformed set $A(B)$: When you apply the matrix A to each point in the set B, you obtain a new set $A(B)$ which is an ellipse. To see why it's an ellipse, consider that the equation of a 2-norm ball of radius 1 is $x^\top x = 1$. When you apply the transformation, you get an equation of the form $(Ax)^\top Ax = 1$, which is the equation of an ellipse as long as $A^\top A$ is pd.

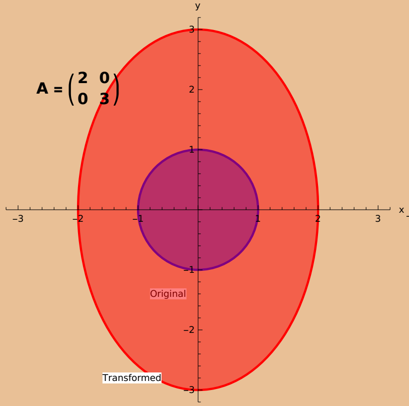
Note the major and minor axes of the ellipse will be determined by the eigenvalues and eigenvectors of the matrix A.

Example:

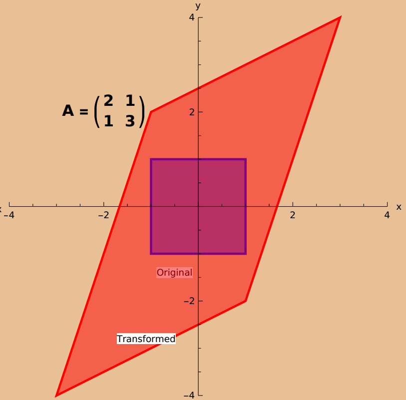
$$A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} .$$

Then $A(B) = \{Ax : x \in B\}$, which can be written $A(B) = \{y \in \mathbb{R}^2 : y^\top A^{-1\top} A^{-1} y = 1\}$ if A is invertible.

Examples

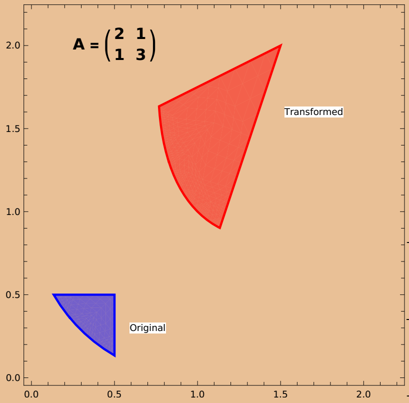


A = (2 0; 0 3)
Original
Transformed
x
y

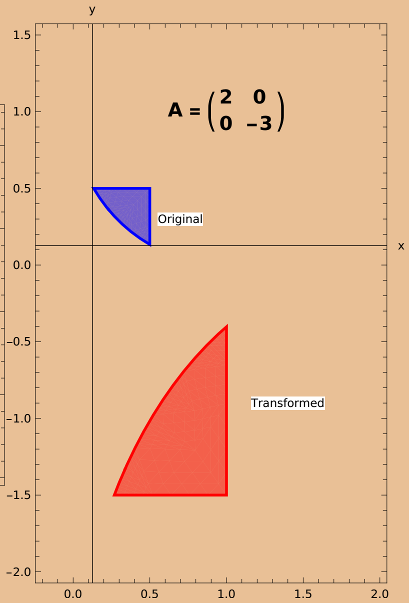


A = (2 1; 1 3)
Original
Transformed
x
y

Examples

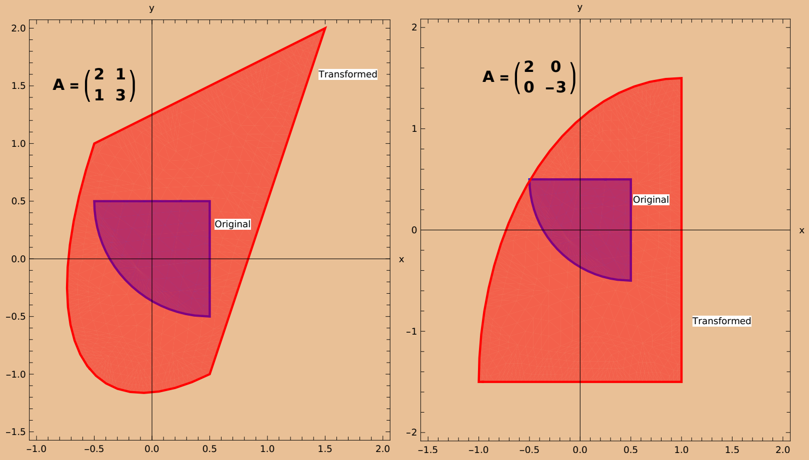


$A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$
Transformed
Original

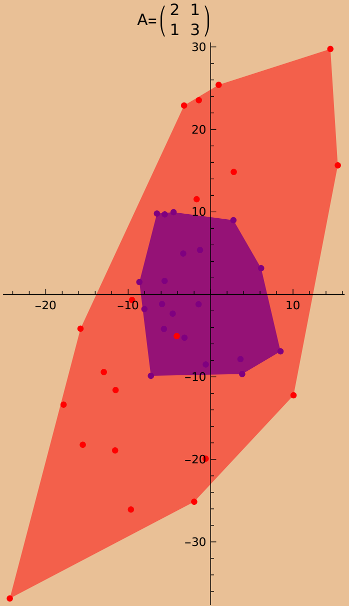


y
x
$A = \begin{pmatrix} 2 & 0 \\ 0 & -3 \end{pmatrix}$
Original
Transformed

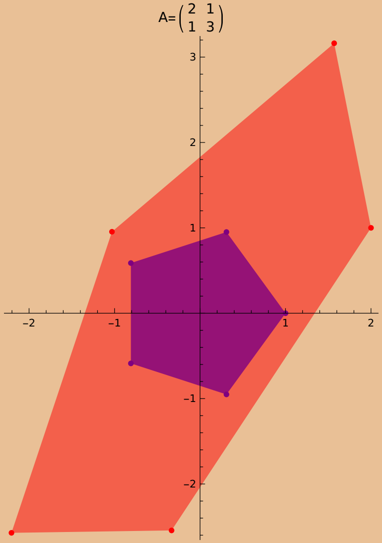
Examples


$A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$
Transformed
Original
$A = \begin{pmatrix} 2 & 0 \\ 0 & -3 \end{pmatrix}$
Original
Transformed
x
y

Examples



$A=\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$
30
20
10
−10
−20
−30
−20
−10
10



$A=\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$
3
2
1
−1
−2
−2
−1
1
2

Operations that preserve set convexity

Proposition (Affine transformation is convex)

The previous proposition, together with (2) *allows to state that, if* $C \subseteq \mathbb{R}^n$ *convex, and* $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ *then*

$$A(C+b) = \{Ax+b \in \mathbb{R}^m : x \in C\}$$

is convex. Also

$$A^{-1}(D-b) = \{x \in \mathbb{R}^n : Ax+b \in D\}$$

is convex.

Examples

Example: i) Let $C = [0, 2]$. Let $A = 3$ and $b = 1$.

Then the transformed set $A(C+b) = [0, 6] + 1 = [1, 7]$, which is convex.

ii) Let the convex set $D = [0, 9]$, and $A = 3$. In this case, the inverse image can be calculated: $A^{-1} = 1/3$.

Then, the inverse image is the set $A^{-1}(D - b)$: We have $A^{-1}(D-b) = A^{-1}([0, 9] - 1) = [-1, 8]/3 = [-1/3, 8/3]$, which is also convex.

Example 2: Let

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} .$$

Then $A(B + b) = \{Ax + b : x \in B\}$, which can be written $A(B) = \{y \in \mathbb{R}^2 : (y - b)^\top A^{-1\top} A^{-1}(y - b) = 1\}$ if A is invertible.

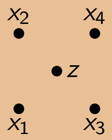
The idea is to construct convex sets out of nonconvex sets. First the following definition.

Definition

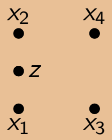
A point x is called a convex combination of the points $x_1, x_2, ..., x_m$ if there exist $\alpha \in \Delta_m$ such that

$$z = \alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_m x_m \,.$$

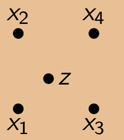
Examples:



$(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$



$(\frac{1}{2}, \frac{1}{2}, 0, 0)$



$(\frac{4}{10}, \frac{2}{10}, \frac{2}{10}, \frac{2}{10})$

What if we take all possible convex combinations of the points x_1, x_2, x_3, x_4?

Compare this with linear combinations.

Convex hull

Definition

The convex hull of a set $X \subseteq \mathbb{R}^n$, denoted $\text{Conv}(X)$, is the set of all possible convex combinations of points in X:

$$\text{Conv}(X) = \left\{ \sum_{i=1}^{k} \alpha_i x_i : \;\; x_1, x_2, ..., x_k \in X, \;\; \alpha \in \Delta_k, \;\; k \in \mathbb{N} \right\}.$$

Note: The convex hull of a set X is also the intersection of all convex sets containing X.

Convex hull

Examples: Sets in $\mathbb{R}^2$:

Convex hulls:

Intuitively, $\mathrm{Conv}\,(X)$ the smallest convex set that contains X.

Convex hull

Note that convex combination can be generalized to include infinite sums, integrals, and probability distributions.

Examples

- *Convex hull of three points in the plane is the triangle having these points as vertices.*
- *Particular case of polyhedron:*
  $P = \{x \in \mathbb{R}^n : a_i^\top x \leq b_i,\ i{=}1, \dots, m\}, \quad a_i \in \mathbb{R}^n,\ b_i \in \mathbb{R}.$
- *Convex hull of a finite set of points is a polyhedron (rightmost figure).*

Observation: Compare the convex hull with the span of a set S of vectors (also called linear hull):

$$\mathrm{Span}\,(S) = \left\{ \sum_{i=1}^{k} \alpha_i x_i : \ \ x_1, x_2, \dots, x_k \in S,\ \ \alpha \in \mathbb{R}^k,\ \ k \in \mathbb{N} \right\}.$$

See `https://www.wolframcloud.com/env/pguerra0/Regions.nb`

Convex hull

Proposition

Let $X \subseteq \mathbb{R}^n$. If $X \subseteq C$ for some convex set C, then $\mathrm{Conv}(X) \subseteq C$.

Recall a convex set is defined by the property that any convex combination of two points from the set is also in the set.

The next result shows that a convex combination of *any number* of points from a convex set is in the set.

Theorem

Let $C \subseteq \mathbb{R}^n$ be a convex set and let $x_1, x_2, ..., x_m \in C$. Then for any $\alpha \in \Delta_m$ we have $\sum_{i=1}^{m} \alpha_i x_i \in C$.

Proof by induction on m.

Caratheodory's Theorem

The following theorem shows how many points are required in a convex combination to describe any element in the convex hull. If a point $x \in \mathbb{R}^n$ lies in the convex hull of a set X, then x can be written as the convex combination of at most $n+1$ points in X.

Theorem

Let $X \subseteq \mathbb{R}^n$ and $x \in \operatorname{Conv}(X)$. Then there exist $x_1, x_2, ..., x_{n+1} \in X$ such that $x \in \operatorname{Conv}(\{x_1, x_2, ..., x_{n+1}\})$. In other words there exist $\alpha \in \Delta_{n+1}$ such that

$$x = \sum_{i=1}^{n+1} \alpha_i x_i \, .$$

The proof of this theorem is given below. It is interesting because gives an algorithm to construct such a representation of x using at most $n+1$ elements of X.

Note: We will see that, only some special points need to be considered, if the set X is compact.

Example 1:
Let

$$X = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}.$$

The convex hull of this set is the unit square. The theorem says you can write any element of the convex hull of X using at most 3 vectors, since $n = 2$.

Now take a point in the lower triangle of the hull, that is $z_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$, for $y_0 < x_0$, $0 < x_0 < 1$ and $0 < y_0 < 1$.

Then, z_0 can be written as a convex combination of

$$\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

What about the point $\begin{pmatrix} 1/2 \\ 0 \end{pmatrix}$? How many vectors? What about $\begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$?

Example 2: Let $n = 2$. Consider

$$x_1 = \begin{pmatrix}1\\1\end{pmatrix}, \;\; x_2 = \begin{pmatrix}1\\2\end{pmatrix}, \;\; x_3 = \begin{pmatrix}2\\1\end{pmatrix}, \;\; x_4 = \begin{pmatrix}2\\2\end{pmatrix}.$$

Let $x \in \mathrm{Conv}\left(\{x_1, x_2, x_3, x_4\}\right)$ be

$$x = \frac{1}{8}x_1 + \frac{1}{4}x_2 + \frac{1}{2}x_3 + \frac{1}{8}x_4 . \tag{*}$$

By Carat. theorem, x can be expressed as a convex combination of *at most* 3 vectors since $n = 2$. To find such coefficients, we first write (the goal is to eliminate one vector let's say x_1)

$$x_2 - x_1 = \begin{pmatrix}0\\1\end{pmatrix}, \;\; x_3 - x_1 = \begin{pmatrix}1\\0\end{pmatrix}, \;\; x_4 - x_1 = \begin{pmatrix}1\\1\end{pmatrix}.$$

These vectors are linearly dependent (l.d.) since the dimension of the space is 2.

Caratheodory's Theorem

We find a linear combination equal zero:

$$(x_2 - x_1) + (x_3 - x_1) - (x_4 - x_1) = 0$$

or

$$-x_1 + x_2 + x_3 - x_4 = 0\,.$$

Multiply by $\gamma \geq 0$ and sum (*) to get

$$x = (\frac{1}{8} - \gamma)x_1 + (\frac{1}{4} + \gamma)x_2 + (\frac{1}{2} + \gamma)x_3 + (\frac{1}{8} - \gamma)x_4\,.$$

The new coefficients sum 1. In order for the coefficient to form a convex combination we also need

$$(\frac{1}{8} - \gamma) \geq 0,\ \ (\frac{1}{4} + \gamma) \geq 0,\ \ (\frac{1}{2} + \gamma) \geq 0,\ \ (\frac{1}{8} - \gamma) \geq 0\,.$$

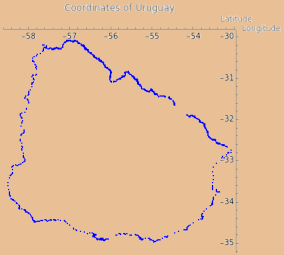
Solving the inequalities we get that $0 \leq \gamma \leq 1/8$. Choose $\gamma = 1/8$, to zero some coefficients and we get

$$x = \frac{3}{8}x_2 + \frac{5}{8}x_3\,.$$

Note we obtained a representation with 2 vectors.

Caratheodory's Theorem

Convex hull of Uruguay is a good approximation of Uruguay. Carat. theorem tells that any point in Conv (Uruguay) can be represented as a convex combination of at most 3 points from its frontier.



Mathematica: `https://www.wolframcloud.com/env/pguerra0/Uruguay.nb`

Caratheodory's Theorem Proof

Proof: Suppose x is the convex combination $x = \lambda_1 x_1 + \cdots + \lambda_k x_k$ of k points in $\mathbb{R}^n$, where $k > n+1$.

Let $y_i = (x_i, 1)^\top \in \mathbb{R}^{n+1}$ denote the vector x_i augmented by an extra component equal to 1.

The $k > n+1$ vectors $y_1, \ldots, y_k$ in $\mathbb{R}^{n+1}$ are linearly dependent, so there exist scalars $\alpha_1, \ldots, \alpha_k$, not all 0, such that

$$\alpha_1 y_1 + \cdots + \alpha_k y_k = 0 \in \mathbb{R}^{n+1},$$

that is

$$\alpha_1 x_1 + \cdots + \alpha_k x_k = 0 \in \mathbb{R}^n \quad \text{and} \quad \alpha_1 + \cdots + \alpha_k = 0.$$

Obviously, at least one α_j must be positive.

Let $r = \max\{-\lambda_j/\alpha_j : \alpha_j > 0\}$. Then $r \leq 0$ and $\lambda_i + r\alpha_i \geq 0$ for all $i = 1, 2, \ldots, k$. This latter inequality is satisfied with equality for at least one index i.

Caratheodory's Theorem Proof

Hence,

$$x = x + r\,0 = \lambda_1 x_1 + \cdots + \lambda_k x_k + r\,(\alpha_1 x_1 + \cdots + \alpha_k x_k)$$
$$= (\lambda_1 + r\alpha_1)\,x_1 + \cdots + (\lambda_k + r\alpha_k)\,x_k$$

where $\lambda_i + r\alpha_i \geq 0$ and $(\lambda_1 + r\alpha_1) + \cdots + (\lambda_k + r\alpha_k) = 1$ for $i = 1, 2, \ldots, k$.

Because $\lambda_i + r\alpha_i = 0$ for at least one i, x must be a convex combination of at most $k-1$ of the points $x_1, \ldots, x_k$ in $\mathbb{R}^n$.

This process of eliminating points x_i one at a time can be repeated until x is expressed as a convex combination of at most $n + 1$ points. $\square$

Note: Similarities with basis in $\mathbb{R}^n$: Both ideas deal with representing points in $\mathbb{R}^n$ using a limited number of other points. The difference is that basis stay fixed for any point you want to represent, whereas in Caratheodory's theorem the set of points depends on the point you want to represent.

Caratheodory's Theorem Proof

Implementation of the proof: Below is one iteration of the elimination process.

```
x1 = [0 0]'; x2 = [1 0]'; x3 = [0 1]'; x4 = [1 1]';
pts = [x1 x2 x3 x4];
la = [1/8 3/8 2/8 2/8]'; % la = [1/4 1/4 1/4 1/4]';

xo = pts*la  % original point

Y = [pts; ones(1,size(pts,2))];

% Find alpha such that Y*alpha = 0
alpha = null(Y);

% Find the maximum ratio -lambda_j/alpha_j for alpha_j > 0
ratios = -la./alpha;
ratios(alpha <= 0) = -inf; % Ignore non-positive alphas
r = max(ratios);

la = la + r*alpha;

% Find indices to eliminate
idx = find(la < 1e-10); % Threshold to account for numerical errors

pts(:, idx) = []; la(idx) = []; % Eliminate

pts*la
```

Approximate Caratheodory's theorem

Even more interesting for representing points. If we allow some error, we can represent any point in $\mathrm{Conv}(X)$ with a prescribed number of points ("budget" k).

Theorem

Let $X \subseteq \mathbb{R}^n$. Then for any $x \in \mathrm{Conv}(X)$ there exist $x_1, x_2, ..., x_k \in X$, $k \in \mathbb{N}$ such that

$$\left\| x - \frac{1}{k}\sum_{i=1}^{k} x_i \right\|_2 \le \frac{diam(X)}{\sqrt{2k}} .$$

The diameter of X is defined as $\mathrm{diam}(X) = \sup\{\|s-t\|_2 : s, t \in X\}$.

This approximation is particularly useful in scenarios where dealing with a large number of dimensions n, since it is independent of n.

What is surprising is that the coefficients of the convex combinations can be chosen all equal.

Note: If we want error $\epsilon = \frac{D}{\sqrt{2k}}$, then we need $k = \frac{D^2}{2\epsilon^2}$.

ref. ACT.m

Supporting hyperplane

Essential concept in optimization theory.

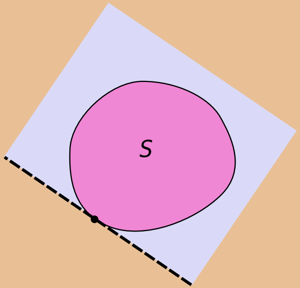
Definition (Supporting hyperplane)

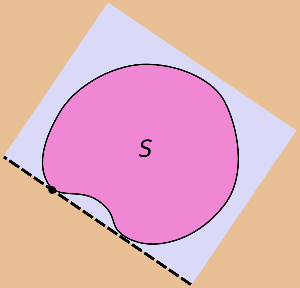
Let $S \subseteq \mathbb{R}^n$ be closed (not necessarily convex). Let x_0 a boundary point. A hyperplane that satisfies the property

$$y^\top x \leq y^\top x_0\,, \quad \forall x \in S\,,$$

for a non-zero vector $y \in \mathbb{R}^n$ is called a supporting hyperplane.

This means all elements of S falls in the halfspace $y^\top x \leq y^\top x_0$.





Supporting hyperplane

The supporting hyperplane may not exist for some points on the boundary (left) or it may not be unique (right):

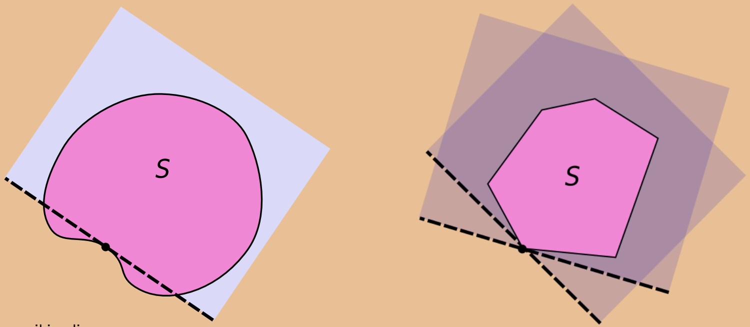
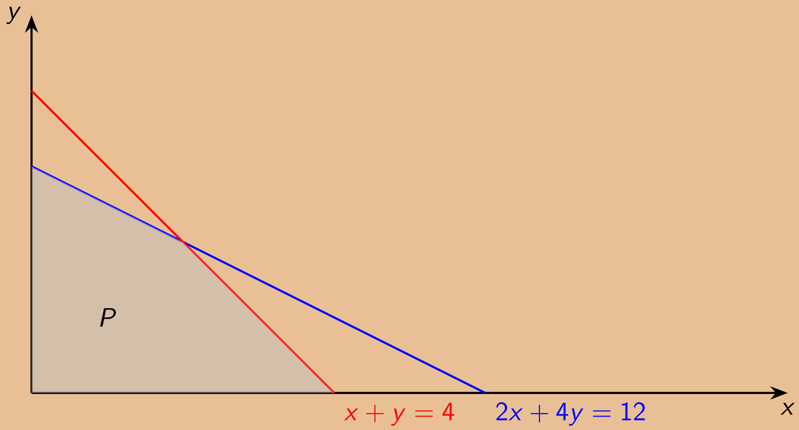


Figure: wikipedia.

Supporting hyperplane - Example

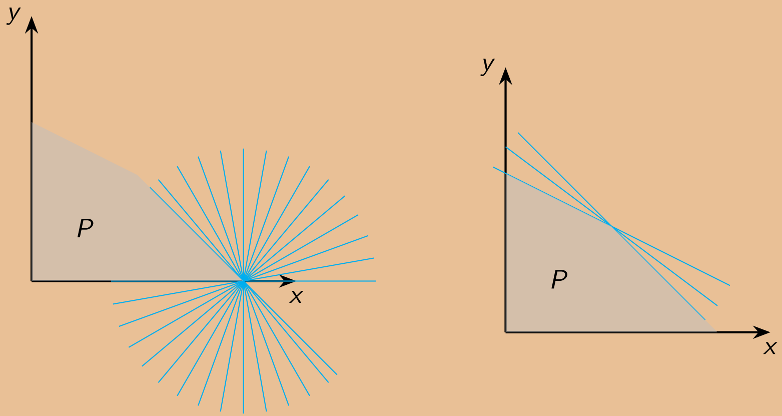
Let

$$P = \{(x, y) : 2x + 4y \leq 12\} \cap \{(x, y) : x + y \leq 4\} \cap \{(x, y) : x \geq 0\,, y \geq 0\}\,.$$



Supporting hyperplane - Example

In the figure we see some supporting hyperplanes at two of the 4 vertices of the polyhedron P:



Supporting hyperplane of a convex set

Theorem (Supporting hyperplane of a convex set)

Let $S \subseteq \mathbb{R}^n$ be a non-empty closed convex set and x_0 in the boundary of S. Then, there exists a supporting hyperplane containing x_0.

Conversely, a closed set with nonempty interior such that every point in the boundary has a supporting hyperplane, then it is convex.

Note: This theorem can be used to characterize convex sets.

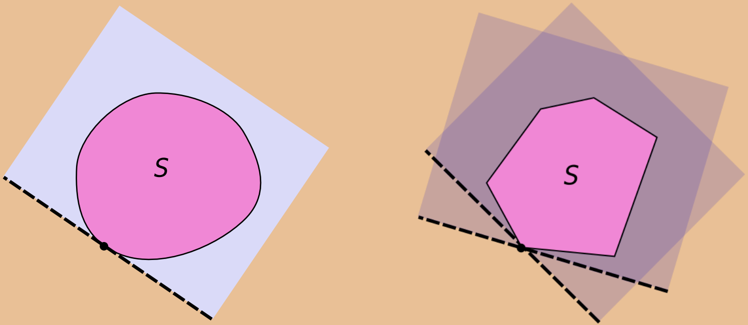


Figure: wikipedia.

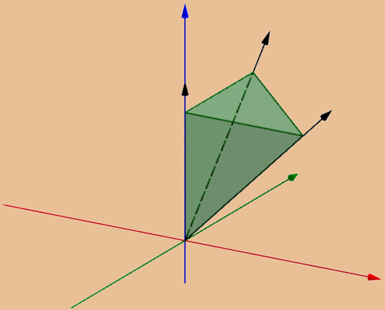
Cones

Definition

A set $K \subseteq \mathbb{R}^n$ is called a cone if for every $x \in K$ one has $\alpha x \in K$ for all $\alpha \geq 0$.

A convex cone is a cone that is a convex set. Convex cone is a close analog of subspace.

Convex cones in $\mathbb{R}$? $(-\infty, 0]$, $[0, +\infty)$, $\{0\}$, $\emptyset$ and $\mathbb{R}$.



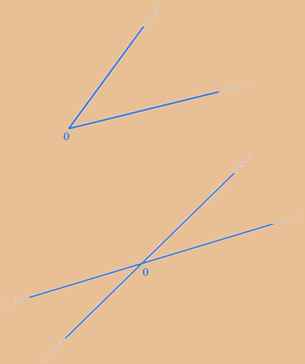


Figure: wikipedia.

Left: convex cone. Right: non-convex cones.

Convex Cones

Proposition (Characterization of a convex cone)

A set K is a convex cone iff the following holds:

a) *If $x, y \in K$ then $x + y \in K$.*

b) *If $x \in K$, and $\alpha \geq 0$ then $\alpha x \in K$.*

Example (Homogeneous linear inequalities)

The solutions of a homogeneous system of linear inequalities form a convex cone:

$$K = \{x \in \mathbb{R}^n : Ax \geq 0, \ \text{for } A \in \mathbb{R}^{m \times n}\}.$$

This cone is a polyhedron.

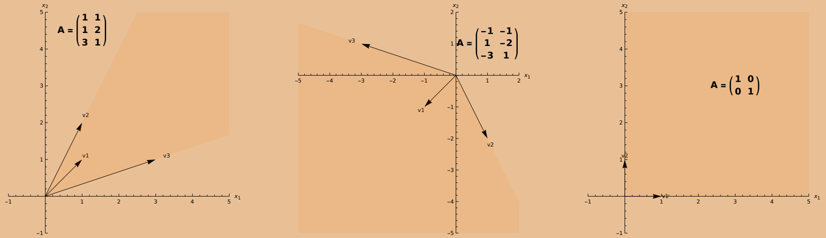
We showed before it is a convex set. It is also a cone since

$$x \in K \implies Ax \geq 0 \implies A(\alpha x) \geq 0 \implies \alpha x \in K, \quad \text{for } \alpha \geq 0.$$

Some examples of

$$K = \left\{x \in \mathbb{R}^n : Ax \geq 0, \text{ for } A \in \mathbb{R}^{m \times n}\right\}.$$

Note: When $A = I$ $(m = n)$ we have that K is the nonnegative orthant.



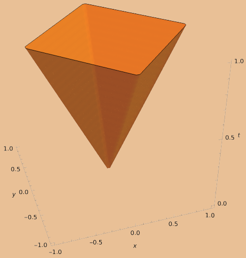
Example of convex cone: Norm cone

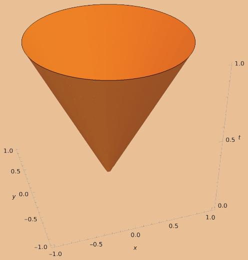
Norm cone: (compare this with norm balls)

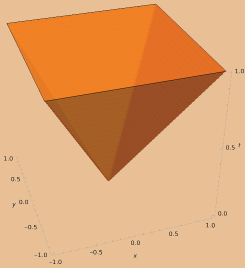
$$K_n = \left\{ \begin{pmatrix} x \\ t \end{pmatrix} \in \mathbb{R}^{n+1} : \|x\| \leq t,\ t \in \mathbb{R} \right\}.$$

Proof: a) Take $(x,t)^\top, (y,s)^\top \in K_n$ and show $(x+y, t+s)^\top \in K_n$. That is $\|x\| \leq t$, $\|y\| \leq s$ implies $\|x+y\| \leq t+s$. b) Take $(x,t)^\top \in K_n$ and show $\alpha(x,t)^\top \in K_n$, for $\alpha \geq 0$.

For $x \in \mathbb{R}^2$ we have:







Example of convex cone: Nonnegative polynomials

Let

$$K_n = \{x \in \mathbb{R}^n : x_1 t^{n-1} + x_2 t^{n-2} + ... + x_{n-1} t + x_n \geq 0,\ \ \forall t \in \mathbb{R}\}.$$

Task: Verify this is a convex cone.

Consider two special cases. For $n = 2$ then

$$K_2 = \{(x_1, x_2)^\top : x_1 t + x_2 \geq 0\} = \{(0, x_2)^\top : x_2 \geq 0\},$$

we have the nonnegative part of x_2-axis.

For $n = 3$ then

$$K_3 = \{(x_1, x_2, x_3)^\top : x_1 t^2 + x_2 t + x_3 \geq 0\}.$$

We need to determine under which condition a quadratic polynomial is nonnegative on $\mathbb{R}$. First observe that $x_1 > 0$ (if $x_1 = 0$ then we are in the linear case) for the "horns" of the quadratic to point upwards.

Example of convex cone: Positive semidefinite cone

Define the set of symmetric $n \times n$ matrices:

$$S^n := \{A \in \mathbb{R}^{n \times n} : A = A^\top\}.$$

It is a vector space with dimension $n(n+1)/2$ and a convex cone.

Positive semidefinite cone:

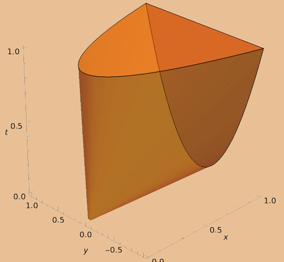
$$S^n_+ := \{A \in S^n : A \text{ is psd }\}.$$

In other words

$$A \in S^n_+ \iff x^\top A x \geq 0, \quad \forall x \in \mathbb{R}^n.$$

Example: $X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in S^2_+$

$\iff x \geq 0, \;\; z \geq 0, \;\; xz \geq y^2.$



Example of convex cone: Positive semidefinite cone

The positive semidefinite cone is the intersection of an infinite number of halfspaces, and so is convex

$$S_+^n = \{A \in S^n : A \text{ is psd}\} = \bigcap_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \{A \in S^n : x^\top A x \geq 0\}\,.$$

Note

- $x^\top A x$ with $x \neq 0$ is a linear function of A;
- $\{A \in S^n : x^\top A x \geq 0\}$ is a halfspace in S^n.

What about the pd cone? $S_{++}^n := \{A \in S^n : A \text{ is pd}\}$?

Conic Hull

Similarly to the notion of a convex combination, we have the conic combination.

Definition (Conic combination)

Let $x_1, x_2, ..., x_m \in \mathbb{R}^n$. A conic combination is $\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_m x_m$ for scalars $\alpha_1 \geq 0, \alpha_2 \geq 0, ..., \alpha_m \geq 0$.

Note unlike the convex combination, the conic combination coefficients are not restricted to sum up to 1.

Proposition (Another characterization of convex cone)

The set K is a convex cone iff for any points $x_1, x_2, ..., x_m \in K$, $m \in \mathbb{N}$ and scalars $\alpha_1 \geq 0, \alpha_2 \geq 0, ..., \alpha_m \geq 0$ we have $\sum_{i=1}^{m} \alpha_i x_i \in K$.

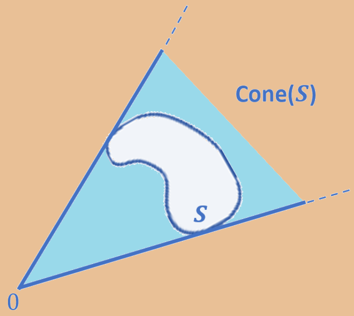
Conic Hull

Definition

Let $S \subseteq \mathbb{R}^n$ then the conic hull

$$\text{Cone}\,(S) := \left\{ \sum_{i=1}^{m} \alpha_i x_i : x_1, x_2, ..., x_m \in S,\;\; \alpha \in \mathbb{R}_+^m,\;\; m \in \mathbb{N} \right\}.$$

Similar to the convex hull, the conic hull of a set S is the smallest convex cone containing S.



Conic representation theorem

In the same way we can represent vectors in the convex hull of a subset of $\mathbb{R}^n$ by using at most $n+1$ vectors (Caratheodory theorem) we can represent a vector in the conic hull using at most n vectors.

Theorem (Conic representation)

Let $S \subseteq \mathbb{R}^n$, and $x \in \mathrm{Cone}\,(S)$. Then there exist k l.i. vectors $x_1, x_2, \ldots, x_k \in S$ such that $x \in \mathrm{Cone}\,(\{x_1, x_2, \ldots, x_k\})$. That is there exists $\alpha \in \mathbb{R}^k_+$ such that

$$x = \sum_{i=1}^{k} \alpha_i x_i \,,$$

with $k \leq n$.

Conic representation theorem

The conic theorem has an important application in convex sets of the form:

$$P = \{x \in \mathbb{R}^n : Ax = b, \ \ x \geq 0\},$$

were $A \in \mathbb{R}^{m \times n}$, for $n \geq m$ and $\operatorname{rank}(A) = m$. We assume P is nonempty. This is the constraint set of a standard linear program [2].

The conic theorem guarantees that P contains at least one point with at most m nonzero elements. These points are called basic feasible solutions:

Definition (Basic feasible solution)

Let $P = \{x \in \mathbb{R}^n : Ax = b, \ \ x \geq 0\}$*, where* $A \in \mathbb{R}^{m \times n}$*. Suppose that the rows of* A *are* $\ell.\imath$*. Then* $\bar{x}$ *is a basic feasible solution (bfs) of* P *if* $A\bar{x} = b$ *and the columns of* A *corresponding to the indices of the positive values of* $\bar{x}$ *are* $\ell.\imath$*.*

Since the columns of A belong to $\mathbb{R}^m$, it follows that a bfs has at most m nonzero elements.

[2] Assuming $b \geq 0$, which we can without losing generality.

Conic representation: basic solutions of a linear program

Example: Consider the polytope

$$\begin{aligned} x_1 + x_2 + x_3 &= 6, \\ x_2 + x_3 &= 3, \\ x_1, x_2, x_3 &\geq 0. \end{aligned}$$

An example of a basic feasible solution (bfs) is $(x_1, x_2, x_3) = (3, 3, 0)$ since

a it satisfies all the constraints

b and the columns corresponding to the positive elements, that is

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

(columns 1 and 2), are $\ell.\imath.$

Conic representation: basic solutions of a linear program

Proposition

Let $P = \{x \in \mathbb{R}^n : Ax = b,\ \ x \geq 0\}$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. If $P \neq \emptyset$, then it contains at least one bfs.

Proof: Since $P \neq \emptyset$, it follows that $b \in \text{Cone}\left(\{a_1, a_2, \ldots, a_n\}\right)$, where a_i denotes the i-th column of A. By the conic representation theorem, we have that b can be represented as a conic combination of k $\ell.\imath.$ vectors from $\{a_1, a_2, \ldots, a_n\}$; that is, there exist indices $i_1 < i_2 < \cdots < i_k$ and k numbers $y_{i_1}, y_{i_2}, .., y_{i_k} \geq 0$ such that

$$b = \sum_{j=1}^{k} y_{i_j} a_{i_j}\ ,$$

and $a_{i_1}, a_{i_2}, \ldots, a_{i_k}$ are $\ell.\imath.$. Denote $x = \sum_{j=1}^{k} y_{i_j} e_{i_j}$. Then $x \geq 0$ and

$$Ax = \sum_{j=1}^{k} y_{i_j} A e_{i_j} = \sum_{j=1}^{k} y_{i_j} a_{i_j} = b.$$

Therefore, x is contained in P and satisfies that the columns of A corresponding to the indices of the positive components of x are $\ell.\imath.$, meaning that P contains a bfs.

Side note: Representations of linear constraints

Let $A \in \mathbb{R}^{m \times n}$, for $n \geq m$ and $\operatorname{rank} A = m$. We denote the null space of A by

$$\mathcal{N}(A) = \{p \in \mathbb{R}^n : Ap = 0\} \; .$$

Note the null space of a matrix is the set of vectors orthogonal to the rows of the matrix. Any linear combination of two vectors in $\mathcal{N}(A)$ is also in $\mathcal{N}(A)$, and thus the null space is a subspace of $\mathbb{R}^n$.

It can be shown that the dimension of $\mathcal{N}(A)$ is $n - \operatorname{rank}(A)$, in this case $n - m$.

Important note: The null space represents the set of feasible directions for the constraints $Ax = b$. This means the following: If the system has at least one solution, then there exists x_0 such that $Ax_0 = b$.

Side note: Representations of linear constraints

Any vector in the null space of A can be added to x_0 to yield another solution to $Ax = b$:

$$A(x_0 + v) = Ax_0 + Av = b + 0 = b \quad \text{for any } v \in \mathcal{N}(A)\,.$$

Key idea: Vectors in the null space of A represent "directions" in which we can move from one solution to another without violating the constraint $Ax = b$.

Another important subspace is the row space of A, that is

$$\operatorname{row}(A) \equiv \mathcal{R}\left(A^\top\right) = \left\{q \in \mathbb{R}^n : q = A^\top \lambda\,, \quad \text{for some } \lambda \in \mathbb{R}^m\right\}\,.$$

The dimension of $\mathcal{R}(A^\top)$ is equal to $\operatorname{rank} A^\top = \operatorname{rank} A$.

There is an important relationship between $\mathcal{N}(A)$ and $\mathcal{R}\left(A^T\right)$: any vector in one subspace is orthogonal to any vector in the other.

Side note: Representations of linear constraints

That is, for any vector $p \in \mathcal{N}(A)$ and $q \in \mathcal{R}\left(A^\top\right)$ we have

$$q^\top p = \lambda^\top A p = 0\,,$$

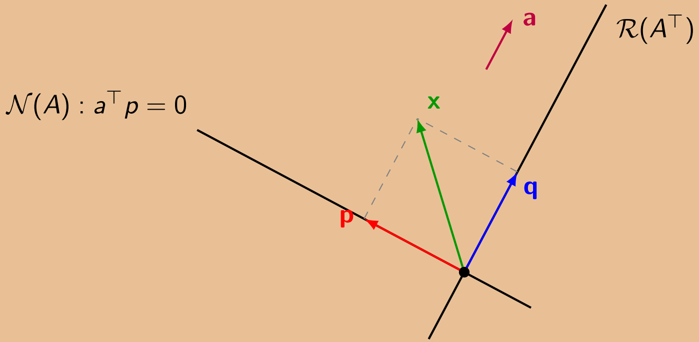
noting that $q \in \mathcal{R}\left(A^\top\right)$ can be expressed as $q = A^\top \lambda$ for some $\lambda \in \mathbb{R}^m$.

Moreover, any n-dimensional vector x can be written uniquely as the sum of a null-space and a range-space component:

$$x = p + q,$$

where $p \in \mathcal{N}(A)$ and $q \in \mathcal{R}\left(A^\top\right)$.

Side note: Representations of linear constraints



The null and range spaces for $A = \left(a^\top\right)$, where a is a two-dimensional nonzero vector.

Side note: Representations of linear constraints

How can we represent vectors in the null space of A?

Definition (Null-space matrix)

Z is a null-space matrix for A if any vector in $\mathcal{N}(A)$ can be expressed as a linear combination of the columns of Z. [3]
If Z is an $n \times r$ null-space matrix, the null space can be represented as

$$\mathcal{N}(A) = \{p \in \mathbb{R}^n : p = Zv\,, \quad \textit{for } v \in \mathbb{R}^r\}\ ,$$

thus $\mathcal{N}(A) = \mathcal{R}(Z)$.

Definition (Basis matrix)

If $r = n - m$, the columns of Z are $\ell.\imath.$, and Z is called a basis matrix. In this case, $r = dim(\mathcal{N}(A))$.

[3] The representation of a null-space matrix is not unique: If A has full row rank m, any matrix Z of dimension $n \times r$ and rank $n-m$ that satisfies $AZ = 0$ is a null-space matrix. The column dimension r must be at least $n-m$.

Side note: Representations of linear constraints

The importance of Z is that, if x_0 is any point satisfying $Ax = b$, then all other feasible points can be written as

$$x = x_0 + Zv$$

for some vector v.

Example: Consider

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

The null space of A is the set of all vectors p such that

$$Ap = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} p_1 - p_2 \\ p_3 + p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix};$$

Side note: Representations of linear constraints

that is, the vector must satisfy $p_1 = p_2$ and $p_3 = -p_4$. Thus any null-space vector must have the form

$$p = \begin{pmatrix} v_1 \\ v_1 \\ v_2 \\ -v_2 \end{pmatrix}$$

for $v_1, v_2 \in \mathbb{R}$. A possible basis matrix for the null space of A is

$$Z = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}$$

and the null space can be expressed as

$$\mathcal{N}(A) = \left\{ p \in \mathbb{R}^4 : p = Zv \,, \quad \text{for some } v \in \mathbb{R}^2 \right\} .$$

Side note: Representations of linear constraints

The matrix

$$\bar{Z} = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

is also a null-space matrix for A, but it is not a basis matrix since its third column is a linear combination of the first two columns.

The null space of A can be expressed in terms of $\bar{Z}$ as

$$\mathcal{N}(A) = \left\{ p \in \mathbb{R}^4 : p = \bar{Z}\bar{v}\,, \quad \text{for } \bar{v} \in \mathbb{R}^3 \right\}\,.$$

Side note: Generating null-space matrices

We now present 2 commonly used methods for deriving a null-space matrix for A. We assume that A is an $m \times n$, $m \leq n$ and $\operatorname{rank} A = m$.

1. Variable reduction method, which produces a $n \times (n-m)$ basis matrix for $\mathcal{N}(A)$.
2. Orthogonal Projection Matrix, which produces a $n \times n$ null-space matrix.

Another commonly used method is QR factorization (not treated here).

Side note: Variable reduction method

This method is the one used by the simplex algorithm for linear programming. It is also used in nonlinear optimization.

Motivational example: Consider the linear system of equations:

$$\begin{aligned} p_1 + p_2 - p_3 &= 0 \\ -2p_2 + p_3 &= 0. \end{aligned}$$

This system has the form $Ap = 0$. We want to generate all solutions to this system.

Solve for any two variables whose associated columns in A are $\ell.\imath.$ in terms of the third variable. For example, we can solve for p_1 and p_3 in terms of p_2 as follows:

$$\begin{aligned} p_1 &= p_2 \\ p_3 &= 2p_2. \end{aligned}$$

Side note: Variable reduction method

The set of all solutions to the system can be written as

$$p = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} p_2,$$

where p_2 is chosen arbitrarily. Thus $Z = (1, 1, 2)^\top$ is a basis for the null space of A.

Here variables p_1 and p_3 dependent variables (depend from p_2). They are also called basic variables. The variable p_2 can take any value is an independent variable, or a non-basic.

Side note: Variable reduction method

In general, consider the $m \times n$ system $Ap = 0$.

- Select any set of m variables whose corresponding columns are $\ell.\imath.$ (basic variables).
- Denote by B the $m \times m$ matrix defined by these columns.
- The remaining variables will be the nonbasic variables and denote the $m \times (n - m)$ matrix of their respective columns by N.
- The general solution to the system $Ap = 0$ is obtained by expressing the basic variables in terms of the nonbasic variables, where the nonbasic variables can take on any arbitrary value.

Side note: Variable reduction method

For ease of notation assume the first m variables are the basic variables. Thus

$$Ap = \begin{pmatrix} B & N \end{pmatrix} \begin{pmatrix} p_B \\ p_N \end{pmatrix} = Bp_B + Np_N = 0.$$

Premultiplying the last equation by B^{-1} we get

$$p_B = -B^{-1}Np_N.$$

Thus the set of solutions to the system $Ap = 0$ is

$$p = \begin{pmatrix} p_B \\ p_N \end{pmatrix} = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} p_N,$$

and the $n \times (n-m)$ matrix

$$Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix}$$

Side note: Variable reduction method

is a basis for the null space of A.

Now consider the system $Ax = b$. One feasible solution is

$$\bar{x} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix}.$$

If x is any point that satisfies $Ax = b$, then x can be written in the form

$$x = \bar{x} + p = \bar{x} + Zp_N = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix} + \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} p_N.$$

Note: If the basis matrix B is chosen differently, then the representation of the feasible points changes, but the set of feasible points does not.

Note 2: If the basic variables are not the first ones, then the rows in Z must be reordered to correspond to the ordering of the basic and nonbasic variables.

Side note: Variable reduction method

Example: Variable Reduction. Consider the system of constraints $Ax = b$ with

$$A = \begin{pmatrix} 1 & -2 & 1 & 3 \\ 0 & 1 & 1 & 4 \end{pmatrix} \text{ and } b = \begin{pmatrix} 5 \\ 6 \end{pmatrix}.$$

Let B consist of the first two columns of A, and let N consist of the last two columns:

$$B = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad N = \begin{pmatrix} 1 & 3 \\ 1 & 4 \end{pmatrix}.$$

Then

$$\bar{x} = \begin{pmatrix} B^{-1}b \\ 0 \end{pmatrix} = \begin{pmatrix} 17 \\ 6 \\ 0 \\ 0 \end{pmatrix}$$

Side note: Variable reduction method

and

$$Z = \begin{pmatrix} -B^{-1}N \\ I \end{pmatrix} = \begin{pmatrix} -3 & -11 \\ -1 & -4 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is easy to verify that $A\bar{x} = b$ and $AZ = 0$. Every point satisfying $Ap = 0$ is of the form

$$Zp_N = \begin{pmatrix} -3 & -11 \\ -1 & -4 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} -3p_3 - 11p_4 \\ -p_3 - 4p_4 \\ p_3 \\ p_4 \end{pmatrix}.$$

Every point x that satisfies $Ax = b$ is of the form

$$x = \bar{x} + \begin{pmatrix} -3p_3 - 11p_4 \\ -p_3 - 4p_4 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 17 - 3p_3 - 11p_4 \\ 6 - p_3 - 4p_4 \\ p_3 \\ p_4 \end{pmatrix}.$$

Side note: Variable reduction method

If instead B is chosen as columns 4 and 3 of A (in that order), and N as columns 2 and 1, then

$$B = \begin{pmatrix} 3 & 1 \\ 4 & 1 \end{pmatrix} \text{ and } N = \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix}.$$

Care must be taken in defining $\bar{x}$ and Z to ensure that their components are positioned correctly. In this case

$$B^{-1}b = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } \bar{x} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 1 \end{pmatrix}.$$

Notice that the components of $B^{-1}b$ are at positions 4 and 3 in $\bar{x}$, corresponding to the columns of A that were used to define B.

Side note: Variable reduction method

Similarly

$$-B^{-1}N = \begin{pmatrix} -3 & 1 \\ 11 & -4 \end{pmatrix} \quad \text{and } Z = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 11 & -4 \\ -3 & 1 \end{pmatrix}.$$

The rows of $-B^{-1}N$ are placed in rows 4 and 3 of Z, and the rows of I are placed in rows 2 and 1. As before, $A\bar{x} = b$ and $AZ = 0$. Every point satisfying $Ap = 0$ is of the form

$$Zp_N = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 11 & -4 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} p_2 \\ p_1 \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ 11p_2 - 4p_1 \\ -3p_2 + p_1 \end{pmatrix}.$$

Side note: Variable reduction method

Every point x that satisfies $Ax = b$ is of the form

$$x = \bar{x} + \begin{pmatrix} p_1 \\ p_2 \\ 11p_2 - 4p_1 \\ -3p_2 + p_1 \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ 2 + 11p_2 - 4p_1 \\ 1 - 3p_2 + p_1 \end{pmatrix}.$$

Note: In practice, the matrix Z is not required to be formed explicitly, and the inverse of B should not be computed. Using that fact, this approach has been enhanced to solve large systems. These enhancements exploit the sparsity that is often present in large problems, in order to reduce computational effort and increase accuracy.

Side note: Orthogonal projection matrix

Let $x \in \mathbb{R}^n$, and let $A \in \mathbb{R}^{m \times n}$ full row rank. Then x can be expressed as

$$x = p + q, \tag{*}$$

where $p \in \mathcal{N}(A)$, $q \in \mathcal{R}\left(A^\top\right)$ such that

$$Ap = 0\,, \quad \text{and} \quad q = A^\top \lambda\,, \quad \text{for } \lambda \in \mathbb{R}^m\,.$$

Plug q in (*) and multiply on the left by A:

$$Ax = 0 + AA^\top \lambda\,,$$

from which we obtain $\lambda = \left(AA^\top\right)^{-1} Ax$. Substituting for q gives the null-space component of x:

$$p = x - A^\top \left(AA^\top\right)^{-1} Ax = \left(I - A^\top \left(AA^\top\right)^{-1} A\right) x.$$

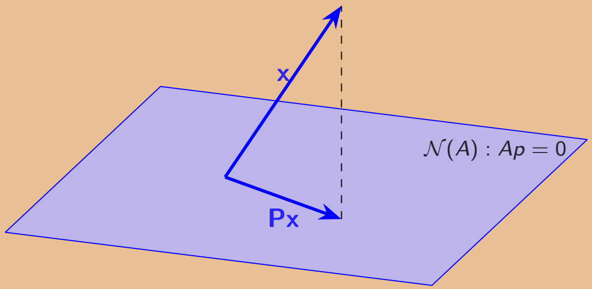
Side note: Orthogonal projection matrix

The $n \times n$ matrix

$$P = I - A^T \left(AA^T\right)^{-1} A$$

is called an orthogonal projection matrix into $\mathcal{N}(A)$.

The null-space component of the vector x can be found by premultiplying x by P; the resulting vector Px is also termed the orthogonal projection of x onto $\mathcal{N}(A)$.



Side note: Orthogonal projection matrix

The orthogonal projection matrix P is the unique matrix with the following properties:

- It is a null-space matrix for A;
- $P^2 = P$, which means repeated application of the orthogonal projection has no further effect;
- $P^\top = P$;

Note: The term "orthogonal projection" may be misleading because unless P is the identity matrix it is not orthogonal.

There are a number of ways to compute the projection matrix. Selection of the method depends in general on the application, the size of m and n, as well as the sparsity of A.

Side note: Orthogonal projection matrix

To demonstrate this point, suppose that A consists of a single row: $A = a^\top$, where $a \in \mathbb{R}^n$. Then

$$P = I - \frac{1}{a^\top a} a a^\top .$$

The matrix-vector product is computed as $Px = x - a\left(a^\top x\right) / \left(a^\top a\right)$.

Note: In general, the task of "inverting" $AA^\top$ becomes expensive and numerically unstable. In particular, the its condition number is the square of that of A. Often, this is done by the Cholesky factorization. A more stable approach is to use a QR factorization of $A^\top$.
For the case when A is large and sparse, the QR factorization may be too expensive, so special techniques that exploit sparsity structure of A have been developed.

Conic representation: basic solutions of a linear program

Consider the system

$$Ax = b,$$

where $A \in \mathbb{R}^{m \times n}$, for $n > m$ and $\operatorname{rank} A = m$.

To solve this system, we apply the reduction method and consider a subset of columns of size m of the matrix A.

For convenience, we often reorder the columns of A so that the m considered columns appear first.

Assume that those m columns are l.i.[4] Denote that matrix A_B. Then,

$$A = [A_B, A_N],$$

where A_N is an $m \times (n - m)$ matrix whose columns are the remaining columns of A.

[4] If the columns are not l.i., then that column set is not of interest since does not correspond to a bfs.

Basic solutions of a linear program

The matrix A_B is nonsingular, and thus we can solve the equation

$$A_B x_B + A_N x_N = b$$

for $x = [x_B^\top, x_N^\top]^\top$.

The solution is $x_B = A_B^{-1} b - A_B^{-1} A_N x_N$. Note this describes an infinite set of solutions since x_N can be arbitrarily chosen.

Note that, for the particular case $x_N = 0$, we have that

$$x = [x_B^\top, 0^\top]^\top$$

is a solution to $Ax = b$ (called basic with respect to the basis A_B).

If in addition $A_B^{-1} b \geq 0$ we have a basic feasible solution (bfs).

Simplex method: efficiently iterate among basic feasible solutions and choose the one that has smallest functional value.

Basic solutions of a linear program

Why bfs are important for optimization?

Theorem (Fundamental theorem of linear programming (FTLP))

Consider a linear program in standard form, that is

$$\underset{x \in P}{\text{minimize}} \;\; c^\top x$$

where $P = \{x \in \mathbb{R}^n : Ax = b, \;\; x \geq 0\}$ *such that* P *is bounded. Then, if* x^* *is an optimal solution to the problem, then* x^* *is a bfs.*

Observe that the FTLP essentially reduces the task of solving a linear programming problem to searching over a finite number of bfs's. The total number of basic solutions (not necessary feasible) is at most

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}, \;\; \text{(finite but large).}$$

(binomial coefficient). See: https:

//www.wolframcloud.com/objects/demonstrations/TheFundamentalTheoremOfLinearProgramming-source.nb

Basic solutions of a linear program

Example

Consider the equation $Ax = b$ with

$$A = \begin{pmatrix} 2 & 3 & -1 & -1 \\ 4 & 1 & 1 & -2 \end{pmatrix}, \;\; b = \begin{pmatrix} -1 \\ 9 \end{pmatrix}.$$

Let $A_B = \begin{pmatrix} 2 & 3 \\ 4 & 1 \end{pmatrix}$ with rank $= 2$. The inverse is $\frac{1}{10}\begin{pmatrix} -1 & 3 \\ 4 & -2 \end{pmatrix}$. Then

$$\frac{1}{10}\begin{pmatrix} -1 & 3 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} -1 \\ 9 \end{pmatrix} - \frac{1}{10}\begin{pmatrix} -1 & 3 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} -1 & -1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} s \\ t \end{pmatrix}$$

for any $s, t \in \mathbb{R}$ is a solution. In particular, if $s = t = 0$ we get that

$$\frac{1}{10}\begin{pmatrix} -1 & 3 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} -1 \\ 9 \end{pmatrix} = \begin{pmatrix} 14/5 \\ -11/5 \end{pmatrix}$$

is a basic solution.

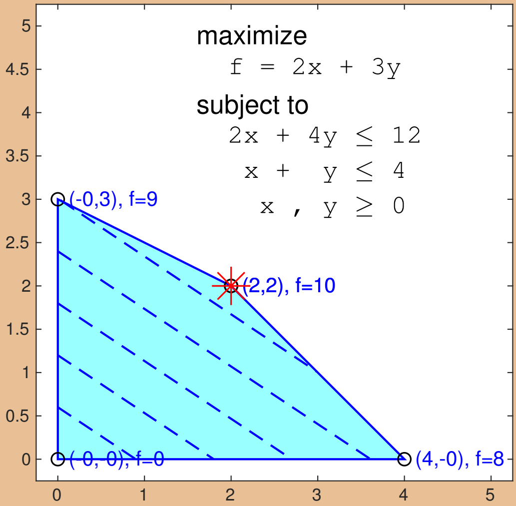
Listing all basic solutions

Output:

```
x =
      14/5         -11/5          0            0

x =
       4/3           0          11/3           0

x =
        0            2            7            0

x =
        0          -11/5          0          -28/5

x =
        0            0          11/3         -8/3
```

Task: Code a MATLAB or Python program that lists all possible basic solutions.

Illustration of a bounded linear optimization problem in $\mathbb{R}^2$



Listing all basic solutions

Output:

```
x =
       2          2          0          0
x =
       4          0          4          0
x =
       6          0          0         -2
x =
       0          4         -4          0
x =
       0          3          0          1
x =
       0          0         12          4
```

What about feasibility?

Bounded linear optimization problem in $\mathbb{R}^2$

The previous optimization problem reads:

$$\begin{aligned} \text{minimize} \quad & 2x + 3y \\ \text{s.t.} \quad & 2x + 4y \leq 12 \\ & x + y \leq 4 \\ & x, y \geq 0\,. \end{aligned}$$

Which is not in standard form. To achieve equalities, we use slack variables s_1 and s_2 to get:

$$\begin{aligned} \text{minimize} \quad & 2x + 3y \\ \text{s.t.} \quad & 2x + 4y + s_1 = 12 \\ & x + y + s_2 = 4 \\ & x, y, s_1, s_2 \geq 0\,. \end{aligned}$$

<u>Why both problems are equivalent?</u> See next slide.

Why equivalent?

Example

Suppose that we are given the inequality constraint

$$x_1 \leq 7.$$

We convert this to an equality constraint by introducing a slack variable $y \geq 0$ *to obtain*

$$\begin{aligned} x_1 + y &= 7 \\ y &\geq 0. \end{aligned}$$

Consider

$$C_1 = \{x_1 : x_1 \leq 7\}$$

and

$$C_2 = \left\{(x_1, y)^\top : x_1 + y = 7, \;\; y \geq 0\right\}.$$

Q: Are both sets equivalent?

Why equivalent?

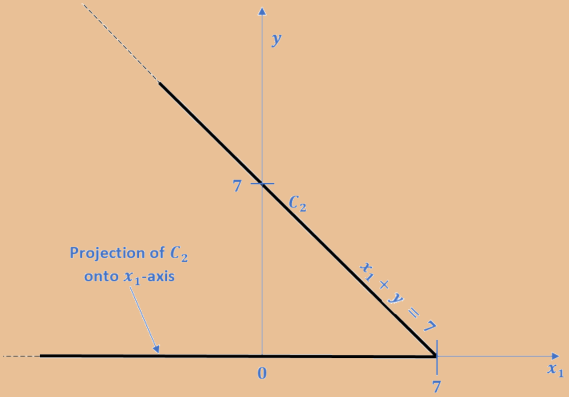


Figure: Projection of the set C_2 onto the x_1-axis

The set C_2 consists of all points on the line to the left and above the point of intersection of the line with the x_1-axis. This set, being a subset of $\mathbb{R}^2$, is not equal to the set C_1 (a subset of $\mathbb{R}$).

However, we can project the set C_2 onto the x_1-axis by taking $(x_1, 0)^\top$.

Establishing the convexity of a given set

Summary of methods for establishing the convexity of a given set C:

- apply definition.
- show that C is constructed by starting with simple known convex sets (hyperplanes, halfspaces, norm balls, ...) and using operations that preserve convexity.

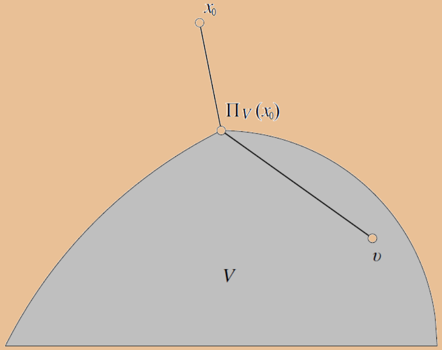
Note: Brute force to establish a set is not convex: take at many random points of the set and check convexity (also cover the boundaries).

Projection on a convex set

Useful operation because many optimization problems can be restated as a projection (recall LS's), and also for constrained optimization.

Definition

For a non-empty convex closed set $V \subseteq \mathbb{R}^n$ and a point $x_0 \in \mathbb{R}^n$, the point in V that is closest to x_0 is called the projection of x_0 on V and is denoted by $\Pi_V(x_0)$.



Proposition

When distance is measured with the 2-norm (Euclidean), then exactly one point in V is the closest to x_0. This is the case proved in the next slide, but the proof is identical when using an inner product $\langle \cdot, \cdot \rangle$.

Figure: A. Ruszczynski. Nonlinear Optimization. (2011) Princeton University Press.

Projection on a convex set

Proposition

Let $V \subseteq \mathbb{R}^n$ be a non-empty closed convex set and $x_0 \in \mathbb{R}^n$. Then, $z = \Pi_V(x_0)$ iff $z \in V$ and

$$(x_0 - z)^\top (v - z) \leq 0\,, \quad \forall v \in V.$$

Proof. ($\Longrightarrow$):
Let $z = \Pi_V(x_0)$. By convexity any point in $x \in V$ can be written

$$x = \alpha v + (1-\alpha) z\,, \quad 0 \leq \alpha \leq 1\,,$$

where $v \in V$. We have

$$\begin{aligned} \|x - x_0\|^2 &= \|\alpha v + (1-\alpha) z - x_0\|^2 \\ &= \|z - x_0\|^2 + 2\alpha (z - x_0)^\top (v - z) + \alpha^2 \|v - z\|^2\,. \end{aligned}$$

The last expression as a function of $\alpha \geq 0$ is bounded below by $\|z - x_0\|^2$ only if the coefficient of the linear term is positive. That is, when $(z - x_0)^\top (v - z) \geq 0$ or equivalently, $(x_0 - z)^\top (v - z) \leq 0$, for $v \in V$.

($\Longleftarrow$): Task.

Projection on a convex set

Proposition (Orthogonal projection)

If the set V is a linear subspace, for any $v \in V$ we have

$$w = 2\Pi_V(x_0) - v \in V$$

as well. Therefore the inequalities

$$(v - \Pi_V(x_0))^\top (x - \Pi_V(x_0)) \leq 0,$$

$$(w - \Pi_V(x_0))^\top (x - \Pi_V(x_0)) = (\Pi_V(x_0) - v)^\top (x_0 - \Pi_V(x_0)) \leq 0,$$

imply that

$$(v - \Pi_V(x_0))^\top (x_0 - \Pi_V(x_0)) = 0.$$

Consequently, we have the usual orthogonal projection $x_0 - \Pi_V(x_0) \perp V$.

Example: Least squares.

The projection operator is nonexpansive

This result will be essential later because it says that, each time we apply a projection in a step algorithm, we are not getting farther from the optimum.

Theorem

If $V \subseteq \mathbb{R}^n$ is a closed convex set, then for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ we have

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

Extreme points

Extreme points fully characterizes a closed bounded convex set.

Definition

A point x of a convex set X is called an extreme point of X if no other points $y, z \in X$, $y \neq z$ exist, such that $x = \alpha y + (1-\alpha)z$, for $\alpha \in [0, 1]$.

That is, x does not lie between any other two points y and z.

Let us denote the set of extreme points of X by $\hat{X}$.

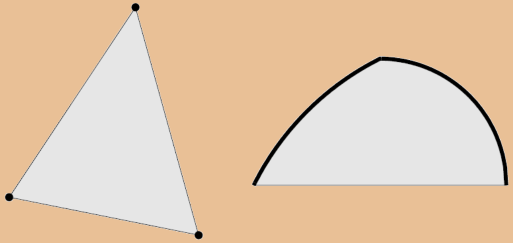


Figure: A. Ruszczynski. Nonlinear Optimization, Princeton University Press 2011.

Theorem (Krein-Milman)

A convex, closed and bounded set $X \subseteq \mathbb{R}^n$ is equal to the convex hull formed by the set $\hat{X}$. That is $X = \mathrm{Conv}\,(\hat{X})$

Extreme points

Examples

1. *The extreme points of*

$$B[0,1] = \{y \in \mathbb{R}^2 : y_1^2 + y_2^2 \leq 1\}$$

are all points on its boundary, that is, on the corresponding circle:

$$\partial B[0,1] = \{y \in \mathbb{R}^2 : y_1^2 + y_2^2 = 1\}\,.$$

2. *The extreme points of the two dimensional set*

$$\{y \in \mathbb{R}^2 : y_1 \geq 1,\;\; y_2 \geq 1,\;\; y_1 + y_2 \leq 3\}$$

are the points (1,1), (1,2), and (2,1). Notice that not all boundary points are extreme points.

Extreme points and linear programs

Example (Extreme points of a polyhedral set)

Let $A \in \mathbb{R}^{m \times n}$ *and let*

$$P = \{x \in \mathbb{R}^n : Ax = b, \;\; x \geq 0\}.$$

A point x *is an extreme point of* P *iff the columns of* A *that correspond to positive components of* x *are* $\ell.\imath.$.

Formally,

Theorem (Equivalence between extreme points and bfs)

$\bar{x}$ *is a bfs of* P *iff it is an extreme point of* P.

Extreme points and linear programs

The goal of linear programming is to minimize (or maximize) a linear objective function $c^\top x$ subject to linear constraints as P.

Theorem

If the set $P = \{x \in \mathbb{R}^n : Ax = b,\ \ x \geq 0\}$ is bounded, it is the convex hull of the set of bfs of the system $Ax = b$.

We have the following geometric interpretation. Let H be a supporting hyperplane of P defined by

$$c^\top x = 0\,.$$

Let $\tilde{H}$ be a hyperplane parallel to H and positioned as shown in Figure 3.

Extreme points and linear programs

The equation of $\tilde{H}$ has the form

$$c^\top x = \beta\,,$$

and $\forall x \in P$, we have $c^\top x \leq \beta$. That is, positioned in such a way that the vector c points in the direction of the halfspace that does not contain P.

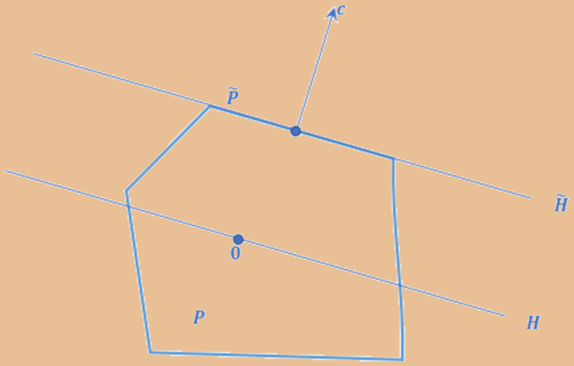


Figure 3: Infinite optimal points of a linear function on the polytope P

Extreme points and linear programs

Let y be a point of $\tilde{P}$, and let x be a point of $P \backslash \tilde{P}$, that is, x is a point of P that does not belong to $\tilde{P}$ (see Figure 4). Then,

$$c^\top x < \beta = c^\top y, \quad \text{which implies} \quad c^\top x < c^\top y.$$

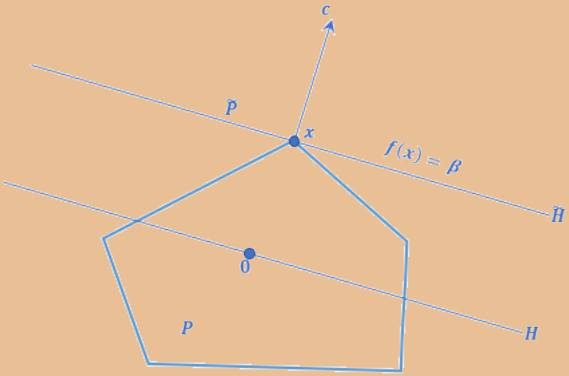


Figure 4: Unique maximum point of a linear function on the polytope P

Exercises

1. If S is an arbitrary set in $\mathbb{R}^n$, prove that the set $\text{Conv}(S)$ is convex. Hint: Let $x = \lambda_1 u_1 + \cdots + \lambda_p u_p$ and $y = \mu_1 v_1 + \cdots + \mu_q v_q$ be arbitrary points in $\text{Conv}(S)$ with $u_1, \ldots, u_p$ and $v_1, \ldots, v_q$ all in S. Let $\lambda \in [0, 1]$ and prove by a direct argument that $\lambda x + (1 - \lambda)y$ is a convex combination of the points $u_1, \ldots, u_p, v_1, \ldots, v_q$.

2. Show that an extreme point of a convex set must be a boundary point of the set. Hint: Show that an interior point cannot be an extreme point.

3. Let

$$A = \{x \in \mathbb{R}^5 : \max\left(|-x_1 + 2x_2 + x_3 - x_4|, x_1 + x_3^2 - x_5\right) \le 2\};$$
$$B = \{x \in \mathbb{R}^5 : -2x_1 + x_2 + x_3 = 4,\ \ -x_1 + x_2 - x_4 + x_5 = 1,\ \ x_i \ge 0\}.$$

   (a) Is the set $A \cap B$ convex?
   (b) Is the point $(0, 1, 3, 0, 0)$ a bfs for the set B?

Solution (last one):

(a) It is sufficient to show that each set is convex since intersection of convex sets is convex. The set B is of the form $\{x \in \mathbb{R}^5 : Ax = b, x \geq 0\}$ and thus a polyhedral convex set. The set A is the level set of the function

$$g(x) = \max(|-x_1 + 2x_2 + x_3 - x_4|, x_1 + x_3^2 - x_5)$$

Level sets of convex functions are convex, so it is suffucient to show that g is convex. The function g is a maximum of two functions: the first, $|-x_1 + 2x_2 + x_3 - x_4|$, is a composition of a convex function $|\cdot|$ and a linear mapping, thus, it is convex. Another way to view it, is as a maximum of two linear functions:

$$|-x_1 + 2x_2 + x_3 - x_4| = \max(-x_1 + 2x_2 + x_3 - x_4, x_1 - 2x_2 - x_3 + x_4).$$

The second function $x_1 + x_3^2 - x_5$ is differentiable; its Hessian has only one non-zero element equal to 2 on its diagonal. This means that the Hessian is psd and the second function is convex as well.

(b) This point satisfies all constraints, so it is feasible. Columns $(1,1)^\top$ and $(1,0)^\top$, corresponding to the positive components of the point, are linearly independent. Thus, this point is a feasible basic solution.

Recession cone of a set

Important to characterize a closed, unbounded convex set.

Definition

Let $X \subseteq \mathbb{R}^n$ be a set. The recession cone of X is the set

$$X_\infty = \{d \in \mathbb{R}^n : x + \alpha d \in X, \ \forall x \in X, \ \forall \alpha \geq 0\}.$$

It is the set of all directions d such that, when starting from $x \in X$ and going indefinitely in the direction d we never cross the boundary of X.

If the set X is convex, then we can write the recession cone as

$$X_\infty = \{d \in \mathbb{R}^n : \forall x \in X, \ x + d \in X\}.$$

Proposition

For a convex set X, the recession cone X_∞ is a convex cone.

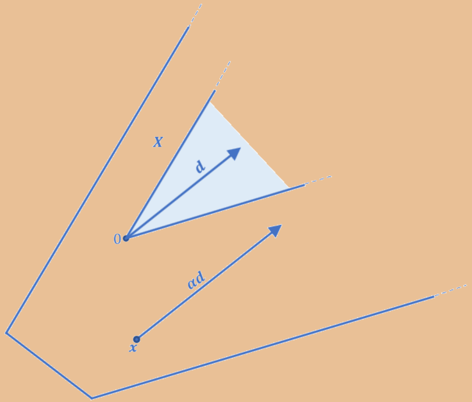
Recession cone of a convex set

Proposition

A closed convex set X is bounded iff $X_\infty = \{0\}$.

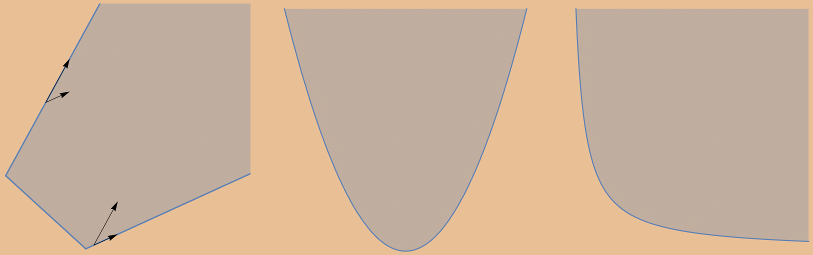
Theorem

A closed convex set X, with non-empty $\hat{X}$ can be represented as $X = \mathrm{Conv}\,(\hat{X}) + X_\infty$.



Recession cone of a convex set

Examples:



Left: $\{(x, y) \in \mathbb{R}^2 : y \geq -x + 1,\ y \leq 2x + 3,\ y \geq x/2 - 3\}$.
$X_\infty = \text{Cone}\left(\left\{(1, 2)^\top, (1, 1/2)^\top\right\}\right)$

Center: $\{(x, y) \in \mathbb{R}^2 : y \geq x^2\}$.
$X_\infty = \text{Cone}\left(\left\{(0, 1)^\top\right\}\right) = \{(0, y) \in \mathbb{R}^2 : y \geq 0\}$.

Right: $\{(x, y) \in \mathbb{R}^2 : x > 0,\ y \geq 1/x\}$.
$X_\infty = \text{Cone}\left(\left\{(0, 1)^\top, (1, 0)^\top\right\}\right) = \mathbb{R}^2_{\geq 0} = \{(x, y) \in \mathbb{R}^2 : x \geq 0,\ y \geq 0\}$.

Examples - Recession cone

Proposition

Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then $P_\infty = \{x \in \mathbb{R}^n : Ax \geq 0\}$.

We apply the proposition to the first example: Let $A = \begin{pmatrix} 1 & 1 \\ 2 & -1 \\ -1/2 & 1 \end{pmatrix}$.

Solving the system $Ax \geq 0$ we get the solutions

1. $(x_1, x_2) = (0, 0)$
2. $x_1 > 0$ and $\frac{x_1}{2} \leq x_2 \leq 2x_1$.

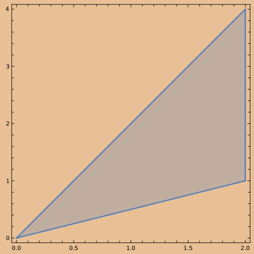
Examples - Recession cone

The second solution describes two rays. The first inequality $\frac{x_1}{2} \leq x_2$ is a line with slope 0.5. The second inequality $x_2 \leq 2x_1$ is a line with a slope 2.

The director vectors of those lines are:

$$v_1 = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}, \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Therefore, any point in the recession cone can be represented as a nonnegative linear combination of v_1 and v_2.



Recession cone of a standard linear program constraint set

Proposition

Let

$$P = \{x \in \mathbb{R}^n : Ax = b, \ \ x \geq 0\}.$$

Then

$$P_\infty = \{d \in \mathbb{R}^n : Ad = 0, \ \ d \geq 0\}.$$

Proof sketch: Find all d such that $A(x + \lambda d) = b$ and $x + \lambda d \geq 0$, for all $\lambda \geq 0$.

Expanding $A(x + \lambda d) = b$ gives $Ax + \lambda Ad = b$. Since $Ax = b$ for all $x \in X$, this simplifies to $\lambda Ad = 0$. As this must hold for all $\lambda \geq 0$, it implies $Ad = 0$.

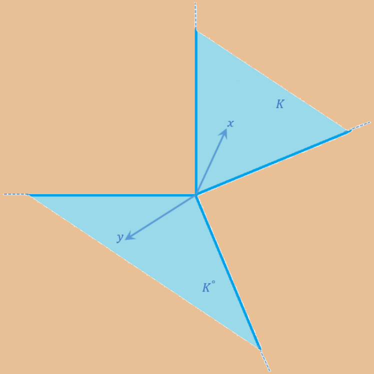
The condition $x + \lambda d \geq 0$ for all $\lambda \geq 0$ implies that $d \geq 0$ since, for large λ, any negative component of d would make $x + \lambda d$ negative.

Definition (Polar cone)

Let K be a cone in $\mathbb{R}^n$. Then the polar cone of K is the set

$$K^\circ = \{y \in \mathbb{R}^n : y^\top x \leq 0\,, \quad \forall x \in K\}\,.$$

This is the set of all vectors y that form an angle of 90 degrees or more with respect to every vector in K.



Proposition

Let K be a convex cone in $\mathbb{R}^n$:

- *The polar cone K° is convex and closed.*
- *$K^\circ = (\bar{K})^\circ$, where $\bar{K}$ is the closure of K*
- *If K is closed, then $(K^\circ)^\circ = K$.*

Example (Non-negative orthant)

The cone $K = \{x \in \mathbb{R}^n : x_i \geq 0, \;\; i = 1, .., n\}$ is convex and closed. Its polar is $K^\circ = \{x \in \mathbb{R}^n : x_i \leq 0, \;\; i = 1, .., n\}$.

Example (Homogeneous linear inequalities)

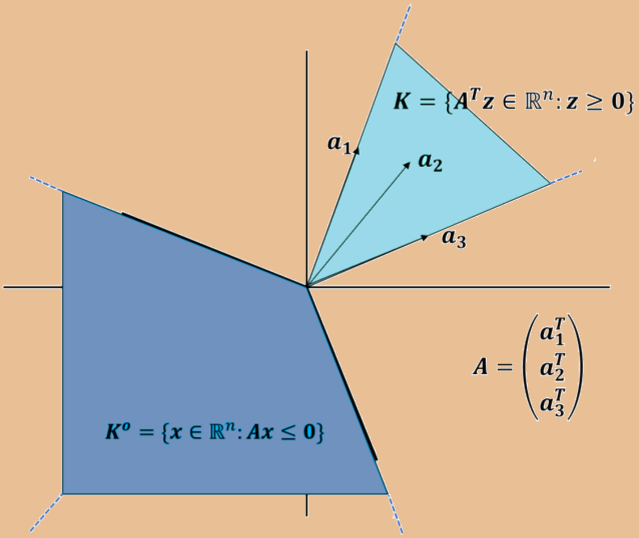
The polar cone of the set

$$K = \{y \in \mathbb{R}^n : y = A^\top z \,, \;\; z \in \mathbb{R}^m \,, \;\; z \geq 0\}$$

is the convex cone

$$K^\circ = \{x \in \mathbb{R}^n : Ax \leq 0 \,, \;\; \text{for } A \in \mathbb{R}^{m \times n}\} \,.$$

Example polar cone



Homogeneous linear inequalities

We have that

$$K = \{y \in \mathbb{R}^n : y = A^\top z\,,\ \ z \in \mathbb{R}^m\,,\ \ z \geq 0\}\,.$$

To derive K° using the definition of polar cone we note that to belong to K°, a vector x must satisfy $y^\top x \leq 0$ for all y in K. Substituting $y = A^\top z$, we get $(A^\top z)^\top x \leq 0$ for all $z \geq 0$.

The expression $(A^\top z)^\top x$ simplifies to $z^\top (Ax)$ because $(AB)^\top = B^\top A^\top$. Thus, we have $z^\top (Ax) \leq 0$ for all $z \geq 0$.

The inequality $z^\top (Ax) \leq 0$ must hold for all $z \geq 0$. This implies that Ax must be less than or equal to 0 (i.e., $Ax \leq 0$).

Why? Because if any component of Ax were positive, we could choose z with a positive value in the corresponding component and 0 elsewhere to make $z^\top (Ax) > 0$, violating the condition. Hence, the only way to ensure $z^\top (Ax) \leq 0$ for all $z \geq 0$ is if every component of Ax is non-positive.

Use of polar cones in LP

We have an interesting application of polar cones in LP that motivates a geometric optimality condition in the constraint case that we will study later. Let the following LP:

$$\begin{aligned} \text{maximize } & c^\top x \\ \text{s.t. } & Ax \leq b \end{aligned}$$

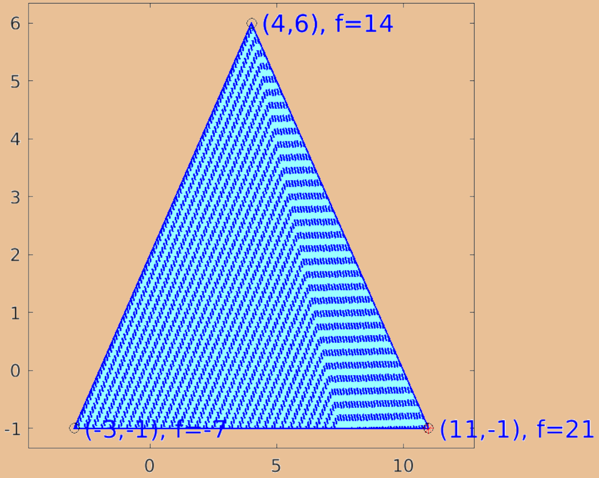
where $A \in \mathbb{R}^{m\times n}$ such that $Ax \leq b$ describes a nonempty bounded polyhedron (polytope). Here we do not require that $x \geq 0$.

Theorem

Let $\bar{x}$ a feasible solution. Then $\bar{x}$ is optimal iff the gradient of the objective function (namely c) is contained in the cone generated by the gradients of the active constraints at $\bar{x}$.

Example: $A = \begin{pmatrix} -1 & 1 \\ 1 & 1 \\ 0 & -1 \end{pmatrix}$, $b = (2, 10, 1)^\top$, $c = (2, 1)^\top$.

Use of polar cones in LP



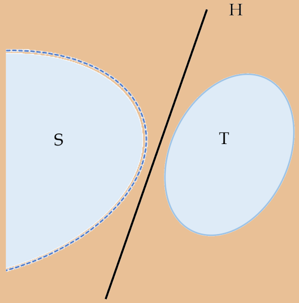
Here $c \in \mathrm{Cone}(\{a_2, a_3\})^\circ$, $(c = 2a_2 + a_3)$, therefore the rightmost vertice is the optimal one. (Verify it!).

Separation theorems

Recall if $a \neq 0$ in $\mathbb{R}^n$ and b is a real number, then the set

$$H(a, b) = \{x : a^\top x = b\}$$

is a hyperplane in $\mathbb{R}^n$, where a is a normal vector. Also recall a hyperplane separates $\mathbb{R}^n$ into two closed half-spaces.



If S and T are subsets of $\mathbb{R}^n$, then H is said to separate S and T if S is contained in one of the closed half-spaces determined by H and T is contained in the other.

In other words, S and T can be separated by a hyperplane if there exist a vector $a \neq 0$ and a scalar b such that

$$a^\top x \leq b \leq a^\top y\,, \quad \text{for all } x \in S \text{ and all } y \in T\,.$$

If both inequalities are strict, then H strictly separates S and T.

The first separation theorem deals with the case where S is closed and convex, and T consists of only one point, $T = \{y\}$.

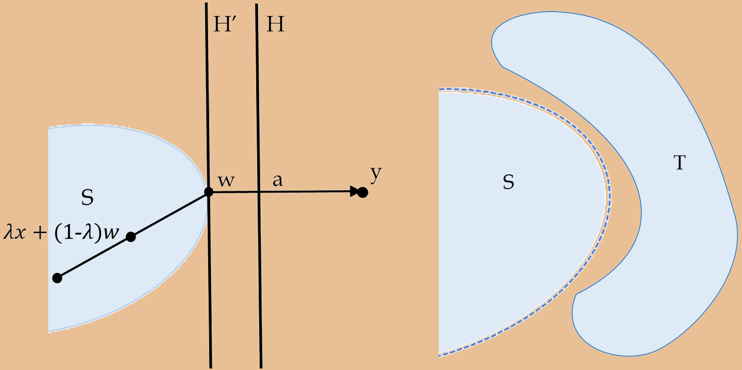
Theorem (First separation - closed convex and point)

Let S be a closed, convex set in $\mathbb{R}^n$, and let y be a point in $\mathbb{R}^n$ that does not belong to S. Then there exists a nonzero vector a in $\mathbb{R}^n$ and a number α such that

$$a^\top x < b < a^\top y\,, \quad \textit{for all } x \textit{ in } S\,. \tag{*}$$

For every such b the hyperplane $\{x : a^\top x = b\}$ strictly separates S and y.

Separation theorems



Left: Geometric idea of the theorem in $\mathbb{R}^2$: Drop the perpendicular line from y to the nearest point w of the set S. Let H' be the hyperplane through w with the vector $a = y - w$ as a normal. Then H' will separate y and S because S is convex.

Right: Sets cannot be separated by a hyperplane.

Separation theorems

To obtain the desired hyperplane H we need to determine w. For this, the following problem must be solved:

$$w \in \arg\min_{x \in S} \|y - x\| \,.$$

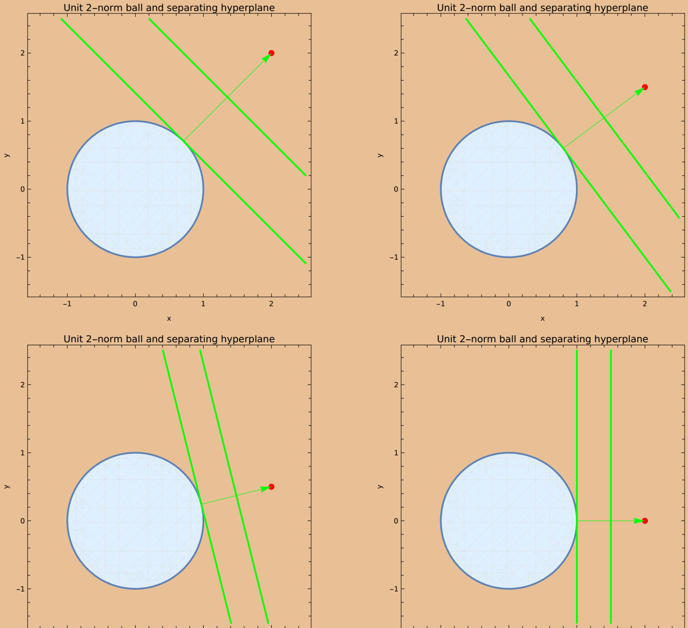
For simple sets S, formulas for w are known:

- Let $S = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$, then $w = \frac{y}{\|y\|}$ since $y \notin S$.
- Let $S = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$, then

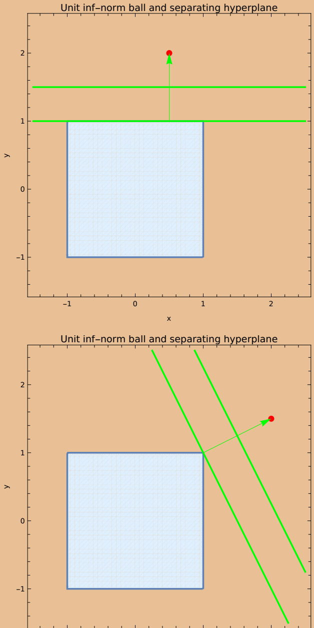
$$w_i = \begin{cases} 1 & \text{if } y_i \geq 1 \\ x_i & \text{if } -1 < x_i < 1 \\ -1 & \text{if } x_i \leq -1 \,, \end{cases}$$

  for $i = 1, .., n$ since $y \notin S$.

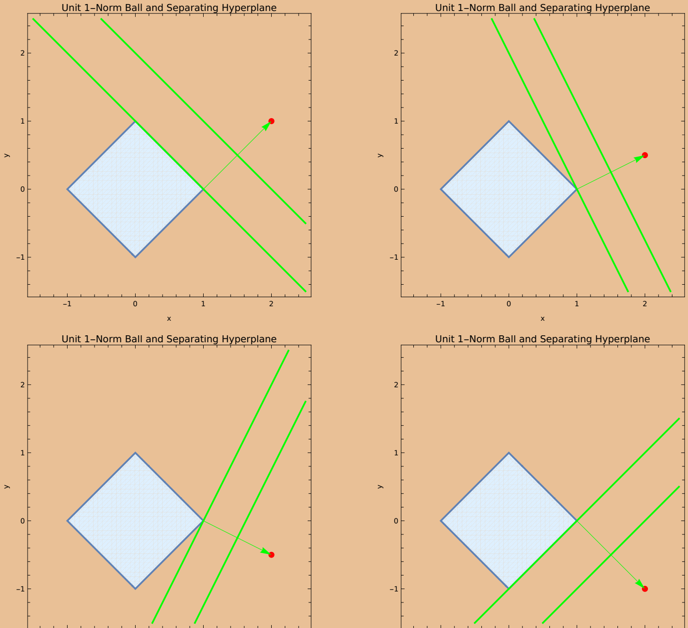
Separation theorems



Separation theorems



Separation theorems



First separation theorem proof

Proof: Because S is a closed set, among all the points of S there is one $w = (w_1, \ldots, w_n)$ that is closest to y.

Let $a = y - w$, the vector from w to y. (See previous figures). Since $w \in S$ and $y \notin S$, it follows that $a \neq 0$. Note that $a^\top(y - w) = a^\top a > 0$, and so $a^\top w < a^\top y$. Suppose we prove that

$$a^\top x \leq a^\top w \quad \text{for all } x \text{ in } S\,. \tag{3}$$

Then (*) will hold for every number b in the open interval $(a^\top w, a^\top y)$. To prove (3), let x be any point in S, and let r denote $x - w$. Since S is convex, the point $w + \lambda b = \lambda x + (1 - \lambda)w$ belongs to S for each λ in $[0, 1]$. Now define $g(\lambda)$ as the square of the distance from $w + \lambda r$ to the point y. Thus

$$g(\lambda) = \|w + \lambda r - y\|^2 = \|a - \lambda r\|^2\,.$$

First separation theorem proof

Differentiating $g(\lambda) = (a_1 - \lambda r_1)^2 + \cdots + (a_n - \lambda r_n)^2$ w.r.t. λ gives

$$g'(\lambda) = -2\,(a_1 - \lambda r_1)\,r_1 - \cdots - 2\,(a_n - \lambda r_n)\,r_n, \quad \text{or} \quad g'(\lambda) = -2(a - \lambda r)^\top r$$

Also $g(0) = \|a\|^2 = \|y - w\|^2$, the square of the distance between y and w. It follows that $0 \leq g'(0) = -2a^\top r$. This proves (3) because $r = x - w$. $\square$

In the proof it was essential that y did not belong to S, and this gave the strict inequality in (*). If S is an arbitrary convex set (not necessarily closed), and if y is not an interior point of S, then it seems plausible that y can still be separated from S by a hyperplane.

If y is a boundary point of S, such a hyperplane is called a supporting hyperplane to S at y. It passes through y and has the property that, for a suitable normal $a = (a_1, \ldots, a_n) \neq 0$ to it, $a^\top x \leq a^\top y$ for all $x = (x_1, \ldots, x_n)$ in S (the vector a points away from S).

Minkowski's separation theorem

The next theorem is important to prove the third one.

Theorem (Second theorem - Separating hyperplane)

Let S be a convex set in $\mathbb{R}^n$ and suppose y is not an interior point of S. Then there exists a nonzero vector a in $\mathbb{R}^n$ such that

$$a^\top x \leq a^\top y\,, \qquad \text{for every } x \in S\,.$$

Theorem (Minkowski's separation theorem)

Let S and T be two disjoint nonempty convex sets in $\mathbb{R}^n$. Then there exists a nonzero vector a in $\mathbb{R}^n$ and a scalar b such that

$$a^\top x \leq b \leq a^\top y\,, \qquad \text{for all } x \in S \text{ and all } y \in T\,.$$

Thus S and T are separated by the hyperplane $H = \left\{z \in \mathbb{R}^n : a^\top z = b\right\}$.

Proof of Minkowski's separation theorem

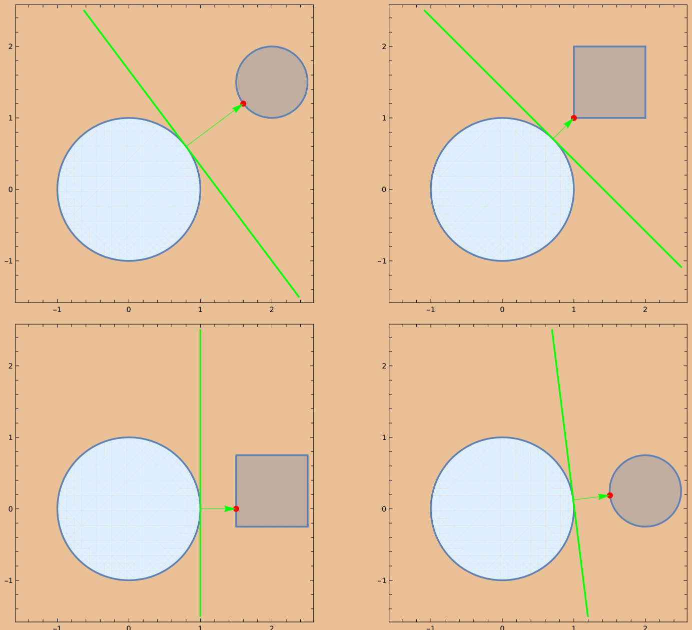
Proof: Let $W = S - T$ be the vector difference of the two convex sets S and T. Since S and T are disjoint, $0 \notin W$.
The set W is convex since it is a Minkowski sum of sets. Thus by second separating theorem, there exists an $a \neq 0$ such that $a^\top w \leq a^\top 0 = 0$ for all $w \in W$.

Let $x \in S$ and $y \in T$ be any two points of these sets. Then $w = x - y \in W$, so $a^\top (x - y) \leq 0$. Hence,

$$a^\top x \leq a^\top y\,, \quad \text{for all } x \in S \text{ and all } y \in T\,. \tag{4}$$

From (4) it follows, in particular, that the set $A = \{a^\top x : x \in S\}$ is bounded above by $a^\top y$ for any y in T. Hence, A has a supremum γ, say. Since γ is the least of all the upper bounds of A, it follows that $\gamma \leq a \cdot y$ for every y in T. Therefore, $a^\top x \leq \gamma \leq a^\top y$ for all x in S and all y in T. Thus S and T are separated by the hyperplane H. □

Minkowski's separation theorem



References

- Axler, S. (2014).
  Linear Algebra Done Right.
  Springer.
- Bazaraa, M. S. and Shetty, C. M. (1979).
  Nonlinear Programming: Theory and Algorithms.
  Wiley, 1 edition.
- Beck, A. (2014).
  Introduction to Nonlinear Optimization.
  MOS-SIAM Series on Optimization. SIAM.
- Bertsekas, D. P. (1999).
  Nonlinear Programming.
  Athena Scientific, 2 edition.
- Boyd, S. and Vandenberghe, L. (2004).
  Convex Optimization.
  Cambridge University Press.
- Chong, E. K. P. and Żak, S. H. (2013).
  An Introduction to Optimization.
  Wiley, 4 edition.
- Fletcher, R. (1987).
  Practical Methods of Optimization.
  John Wiley & Sons, New York, NY, USA, second edition.
- Nocedal, J. and Wright, S. (2006).
  Numerical Optimization.
  Springer.
- Ruszczynski, A. (2006).
  Nonlinear Optimization.
  Princeton University Press.