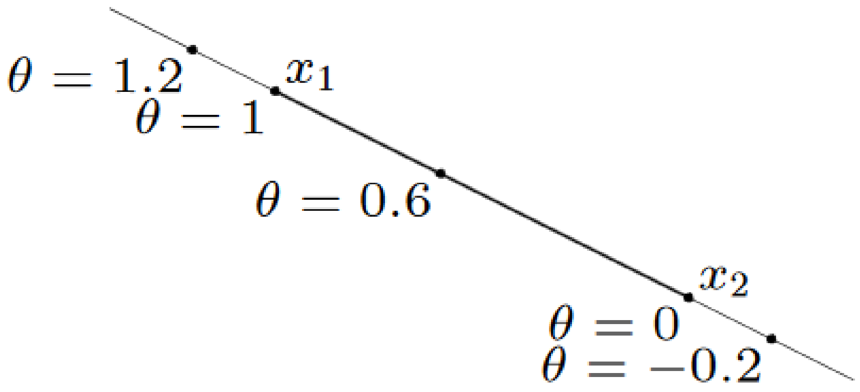
Lecture 03

Convex Sets

Affine set

line through x_1, x_2: all points

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

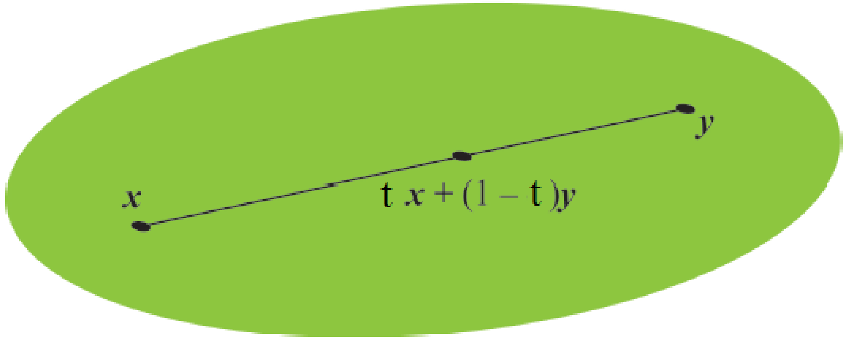


affine set: contains the line through any two distinct points in the set

example: solution set of linear equations $\{x \mid Ax = b\}$

(conversely, every affine set can be expressed as solution set of system of linear equations)

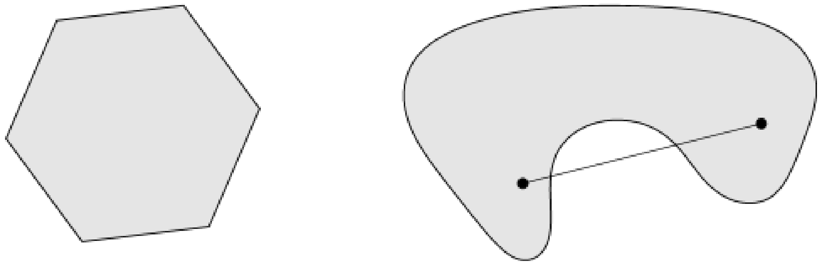
Convex sets

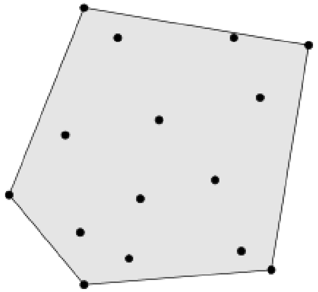


Convex set: $C \subseteq \mathbb{R}^n$ such that

$$x, y \in C \implies tx + (1-t)y \in C \text{ for all } 0 \le t \le 1$$

In words, line segment joining any two elements lies entirely in set

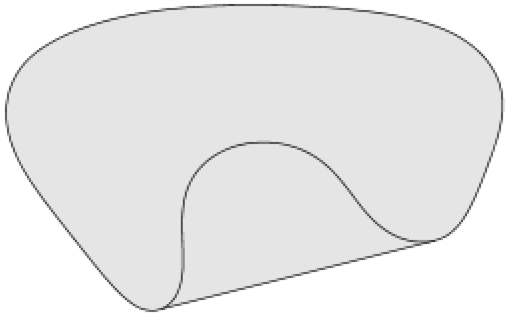




Convex combination of $x_1, \ldots x_k \in \mathbb{R}^n$: any linear combination

$$\theta_1 x_1 + \ldots + \theta_k x_k$$

with $\theta_i \ge 0$, $i = 1, \ldots k$, and $\sum_{i=1}^k \theta_i = 1$. Convex hull of a set C, $\mathrm{conv}(C)$, is all convex combinations of elements. Always convex



Examples of convex sets

- Trivial ones: empty set, point, line
- Norm ball: $\{x : \|x\| \le r\}$, for given norm $\|\cdot\|$, radius r
- Hyperplane: $\{x : a^T x = b\}$, for given a, b
- Halfspace: $\{x : a^T x \le b\}$
- Affine space: $\{x : Ax = b\}$, for given A, b

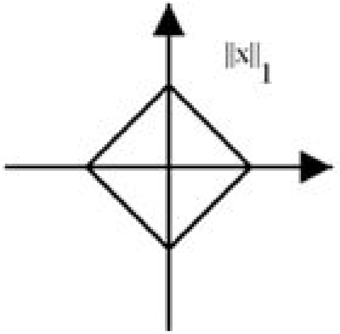
Lp norms and their unit balls

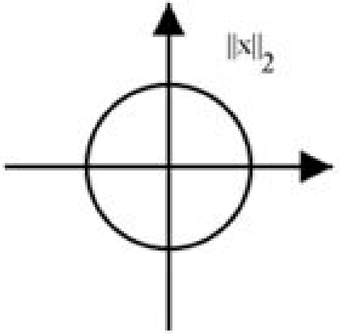
Recall the Lp norm for $z \in R^n$:

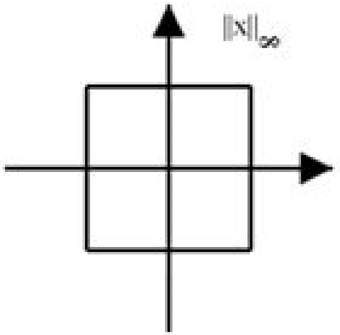
$$||z||_p = \left(\sum_{i=1}^{n} |z_i|^p \right)^{1/p} \quad , p \in [1, \infty)$$

$$||z||_\infty = \max_i |z_i|$$

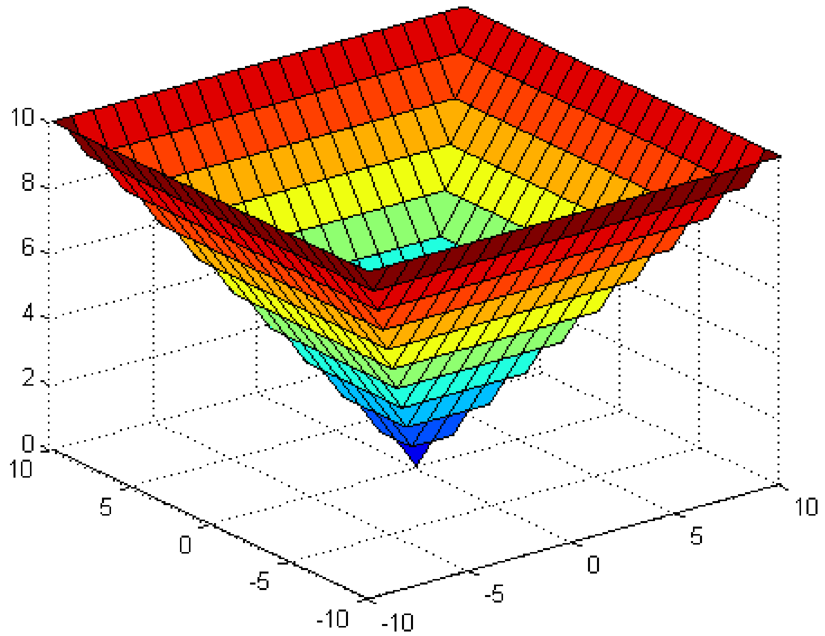
$$||z||_2^2 = \sum_i z_i^2 = z^T z$$



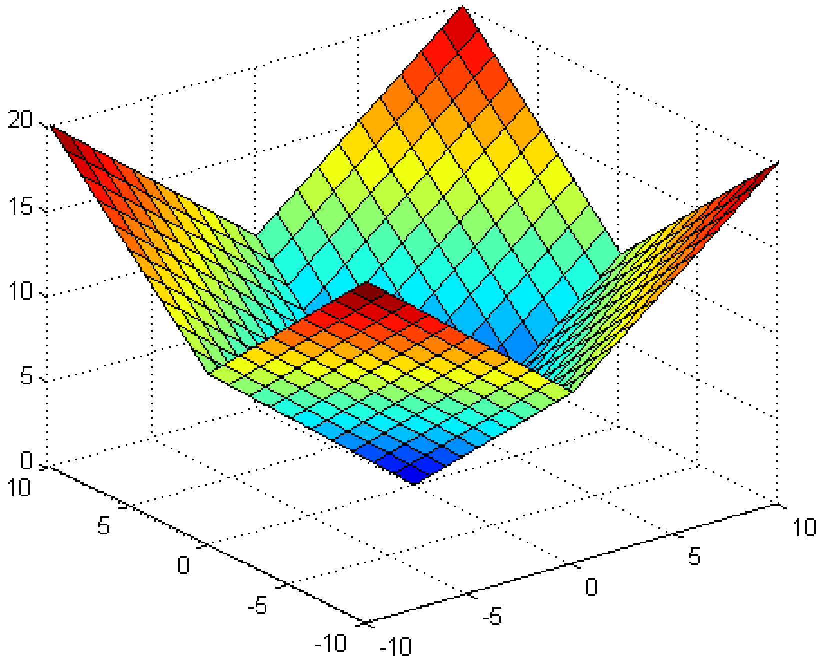


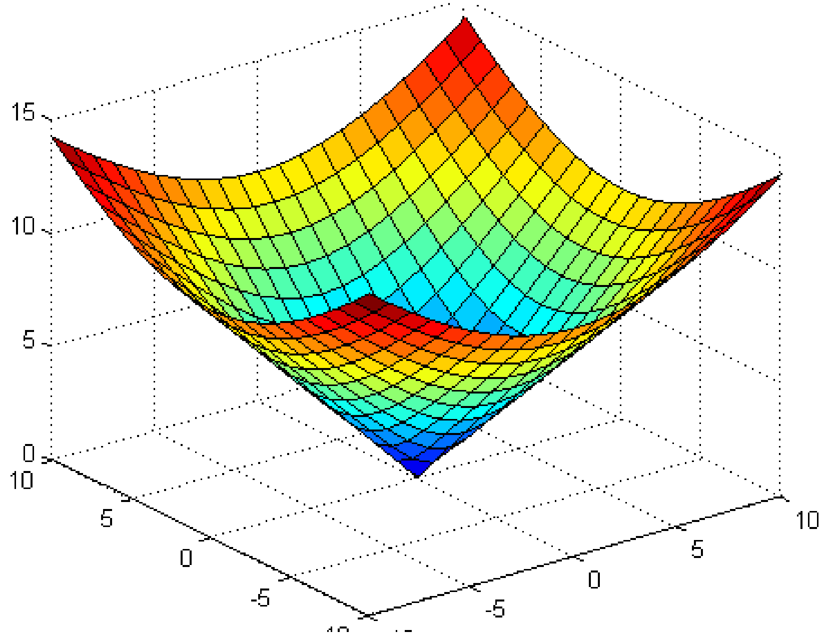


Max Norm (Infinity Norm)



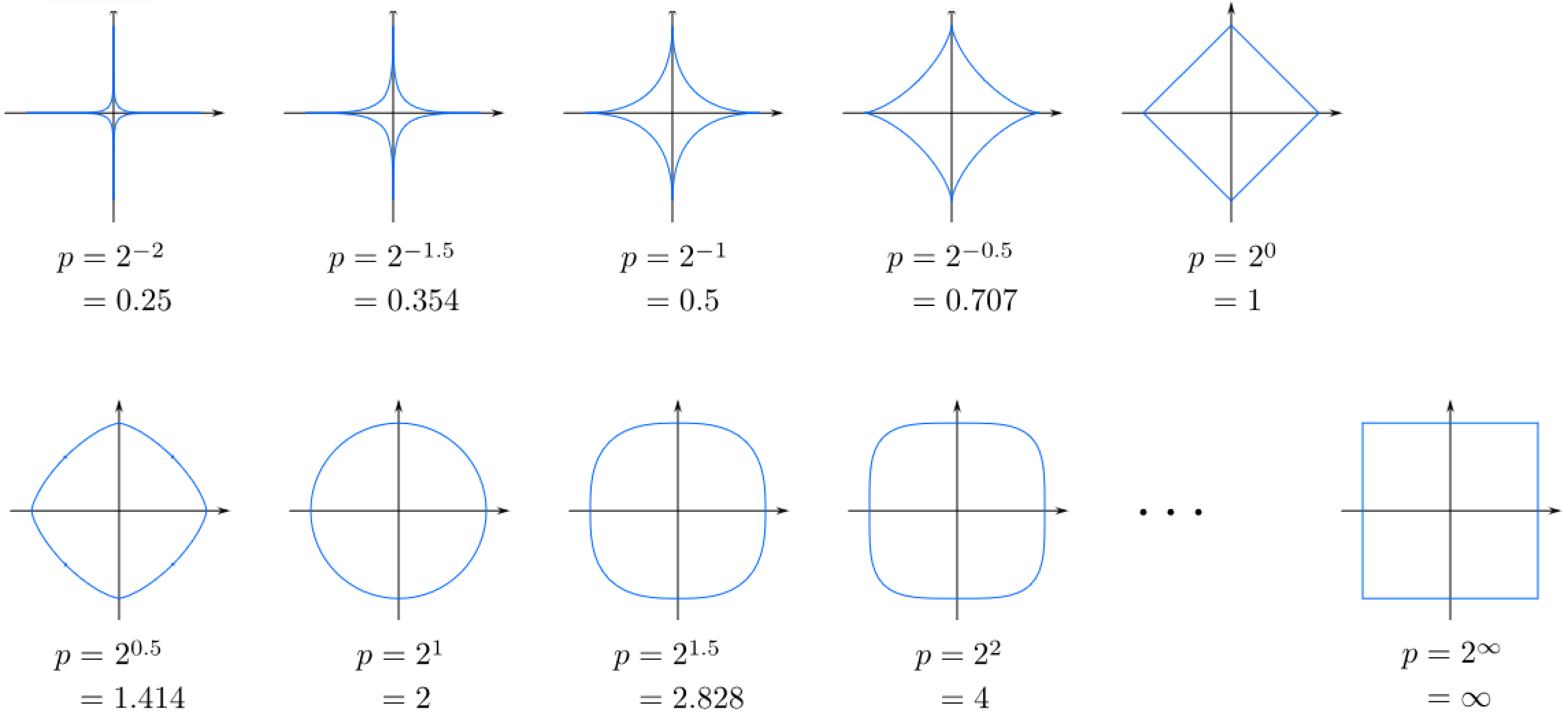
L1 Norm (Manhattan Distance)





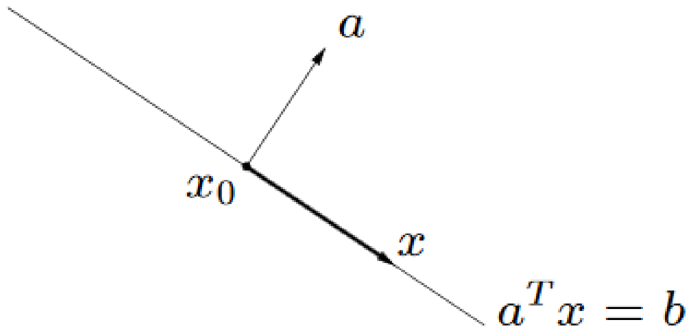
L2 Norm (Euclidean Norm)

Lp norms and their unit balls

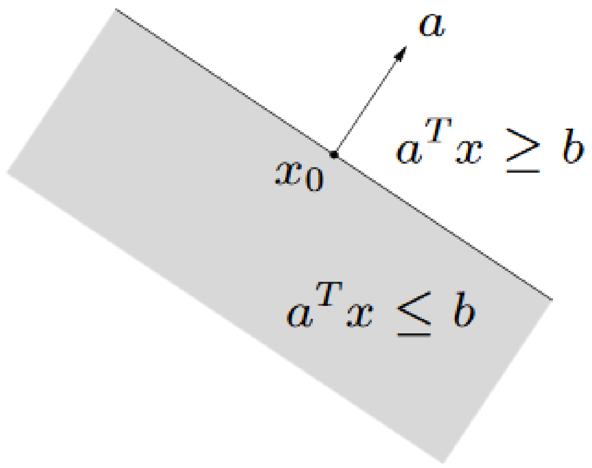


Hyperplanes and halfspaces

hyperplane: set of the form $\{x \mid a^T x = b\}$ $(a \neq 0)$

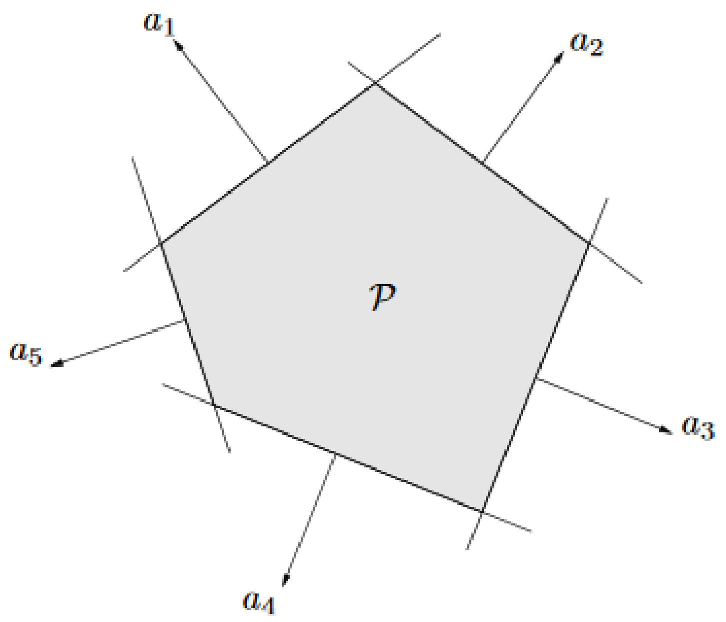


halfspace: set of the form $\{x \mid a^T x \leq b\}$ $(a \neq 0)$



- a is the normal vector
- hyperplanes are affine and convex; halfspaces are convex

- Polyhedron: $\{x : Ax \leq b\}$, where inequality $\leq$ is interpreted componentwise. Note: the set $\{x : Ax \leq b, Cx = d\}$ is also a polyhedron (why?)



- probability simplex :

$$\text{conv}\{e_1, \ldots e_n\} = \{w : w \geq 0,\ 1^T w = 1\}$$

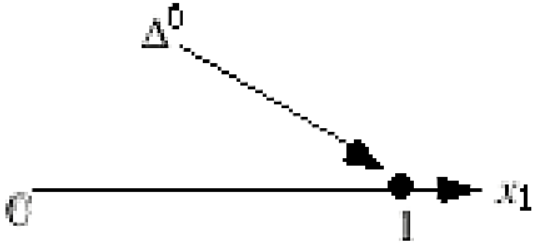
“Unit simplex” (probability simplex) is a convex set

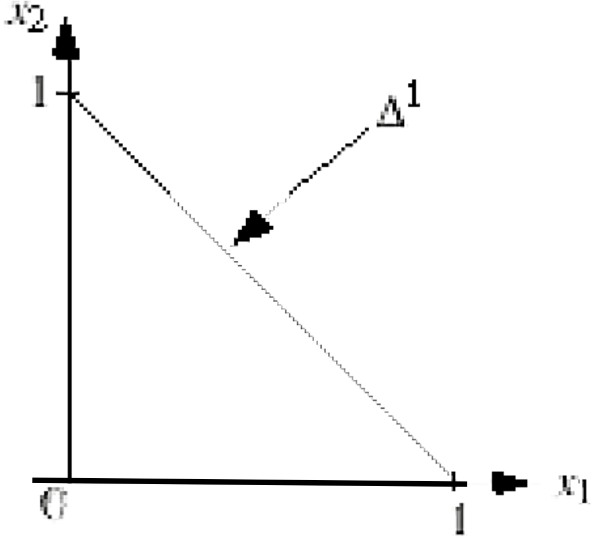
The $(k-1)$-dimensional unit simplex is the set of k-vectors whose components are all

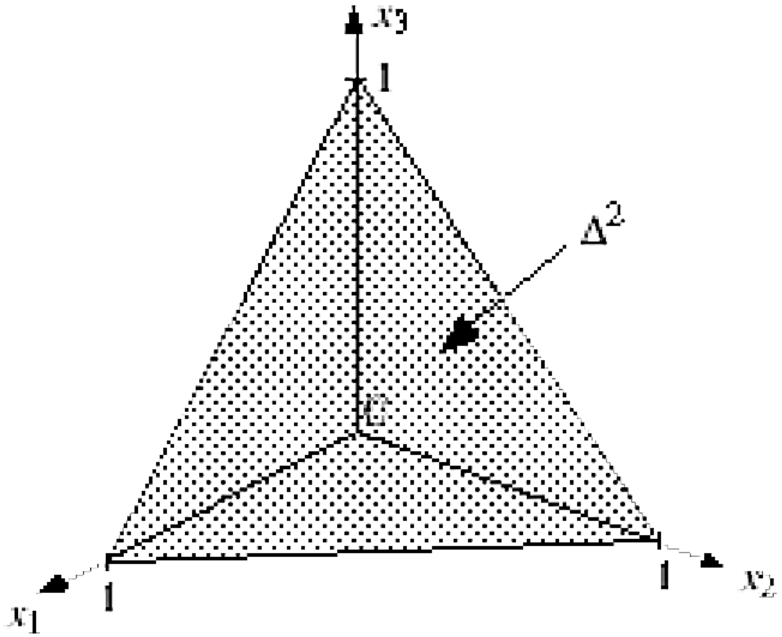
$$\Delta^{k-1} = \left\{ \boldsymbol{x} \in \mathbb{R}^k_+ : \sum_{j=1}^{k} x_j = 1. \right\}$$

$\mathbb{R}^k_+$ is the nonnegative orthant of $\mathbb{R}^k$

$$\{\boldsymbol{x} \in \mathbb{R}^k : \forall i \in \{1, \ldots, k\}, x_i \geq 0\}.$$







Cones

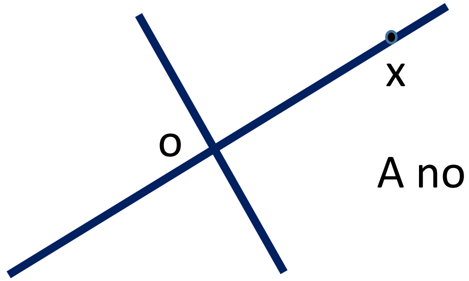
A set $C \subseteq \mathbb{R}^n$ is a *cone* when with every $x \in C$, the whole ray $\{\lambda x \mid \lambda \geq 0\}$ also belongs to the set C, i.e.,

$$\lambda x \in C \qquad \text{for all } x \in C \text{ and } \lambda \geq 0.$$

cones in general need not be convex. For example,

the set $\{x \in \mathbb{R}^2 \mid x_1 x_2 = 0\}$ is a cone that and it is nonconvex.

The **non-negative** orthant $R^n_+ = \{x \in \mathbb{R}^n \mid x \geq 0\}$ is a cone that is convex.



A non-convex cone

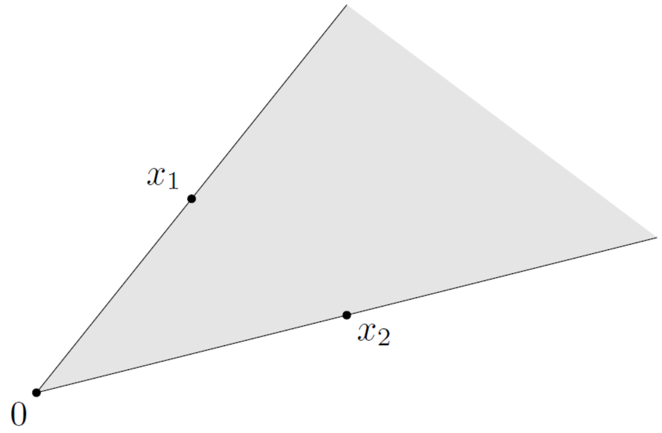
Convex Cone

A set C is a *convex cone* if it is convex and a cone, which means that

for any $x_1,\ x_2 \in C$ and $\theta_1,\ \theta_2 \geq 0$, we have

$$\theta_1 x_1 + \theta_2 x_2 \in C.$$

Points of this form can be described geometrically as forming the two-dimensional pie slice with apex 0 and edges passing through x_1 and x_2.



The pie slice shows all points of the form $\theta_1 x_1 + \theta_2 x_2$, where $\theta_1,\ \theta_2 \geq 0$.
The apex of the slice (which corresponds to $\theta_1 = \theta_2 = 0$) is at 0;
its edges (which correspond to $\theta_1 = 0$ or $\theta_2 = 0$) pass through the points x_1 and x_2.

Conic Combination

A point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$ with $\theta_1, \ldots, \theta_k \geq 0$ is called a *conic combination* (or a *nonnegative linear combination*) of $x_1, \ldots, x_k$.

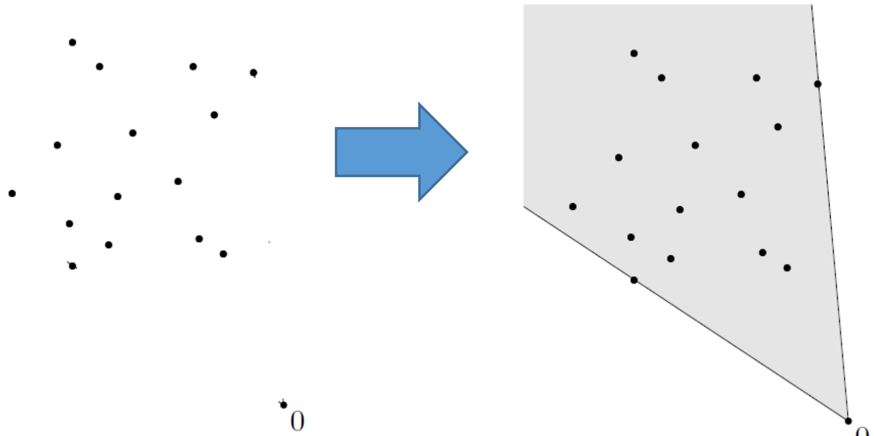
If x_i are in a convex cone C, then every conic combination of x_i is in C.

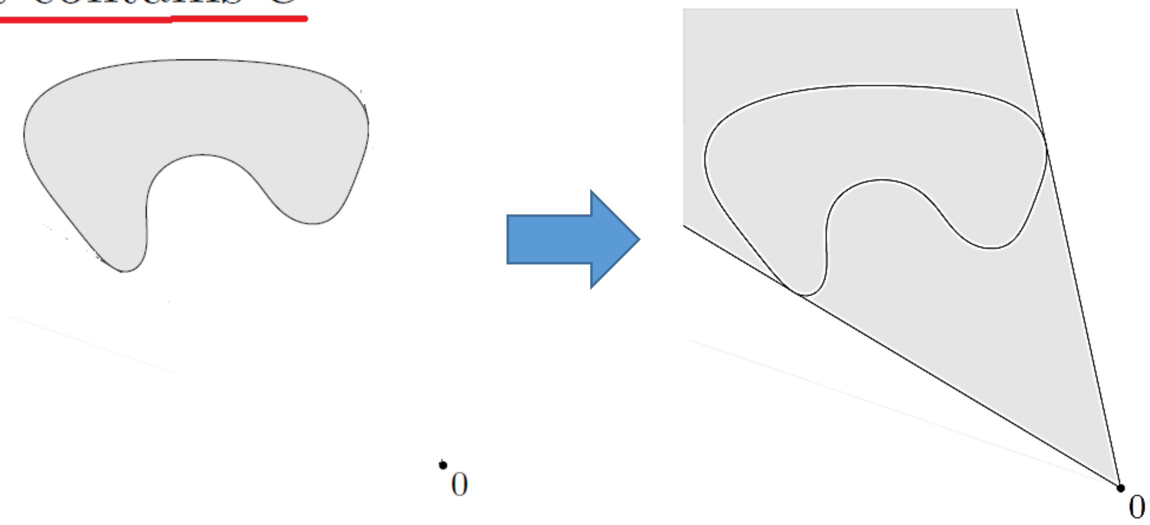
Conversely, a set C is a convex cone if and only if it contains all conic combinations of its elements.

The *conic hull* of a set C is the set of all conic combinations of points in C, *i.e.*,

$$\{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C,\ \theta_i \geq 0,\ i = 1, \ldots, k\},$$

which is also the smallest convex cone that contains C.





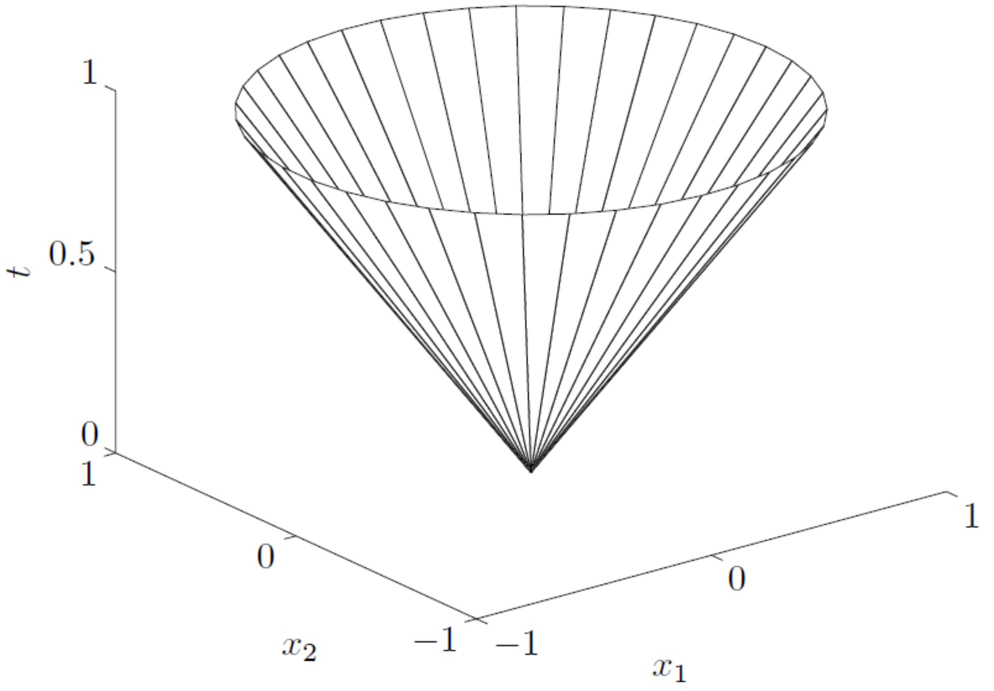
Example Convex Cone: Ice-Cream Cone

A *norm cone* is the set of the form

$$C = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} \mid \|x\| \leq t\},$$

where the norm $\|\cdot\|$ can be any norm in $\mathbb{R}^n$.

The norm cone for Euclidean norm is also known as *ice-cream cone* or *second-order cone.*



Boundary of second-order cone in $\mathbf{R}^3$, $\{(x_1, x_2, t) \mid (x_1^2+x_2^2)^{1/2} \leq t\}$.

Example Convex Cone: PSD Cone

The set $\mathbf{S}_+^n$ (The set of all positive semidefinite matrices) is a convex cone:

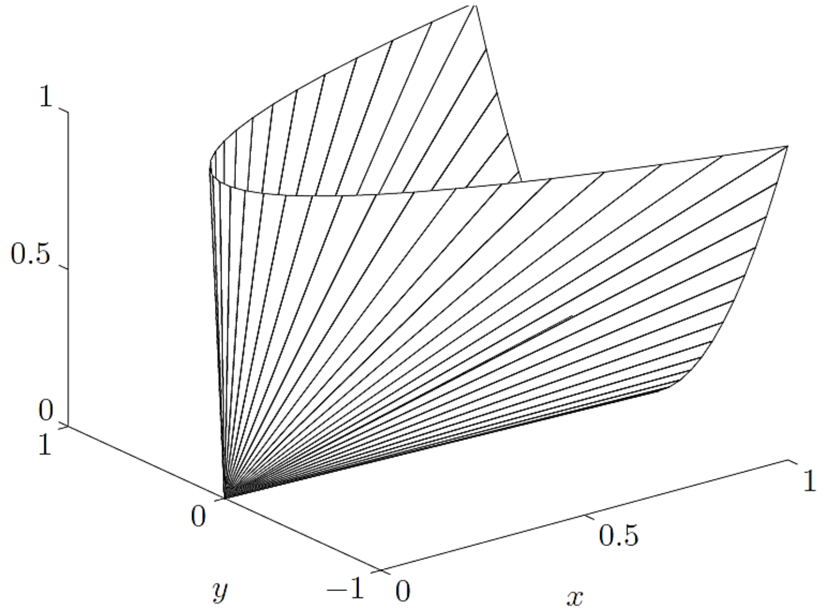
$$\text{if } \theta_1, \theta_2 \geq 0 \text{ and } A,\ B \in \mathbf{S}_+^n, \text{ then } \quad \theta_1 A + \theta_2 B \in \mathbf{S}_+^n.$$

Proof: This can be seen directly from the definition of positive semidefiniteness:

for any $x \in \mathbf{R}^n$, we have

$$x^T(\theta_1 A + \theta_2 B)x = \theta_1 x^T A x + \theta_2 x^T B x \geq 0,$$

if $A \succeq 0$, $B \succeq 0$ and $\theta_1,\ \theta_2 \geq 0$.



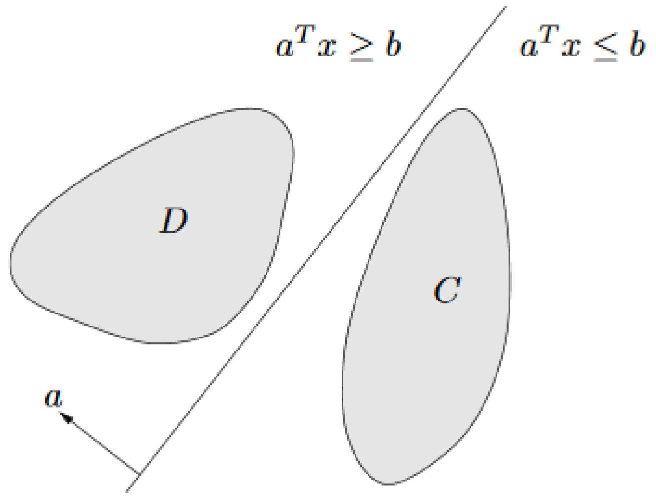
Boundary of positive semidefinite cone in $\mathbf{S}^2$.

Example *Positive semidefinite cone in* $\mathbf{S}^2$. We have

$$X = \begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbf{S}_+^2 \iff x \geq 0, \quad z \geq 0, \quad xz \geq y^2.$$

Key properties of convex sets

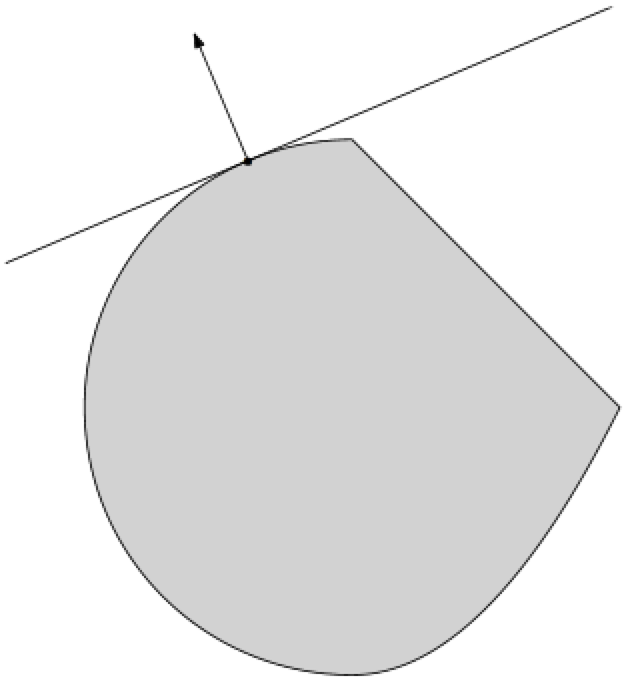
- Separating hyperplane theorem: two disjoint convex sets have a separating between hyperplane them



Formally: if C, D are nonempty convex sets with $C \cap D = \emptyset$, then there exists a, b such that

$$C \subseteq \{x : a^T x \leq b\}$$
$$D \subseteq \{x : a^T x \geq b\}$$

- Supporting hyperplane theorem: a boundary point of a convex set has a supporting hyperplane passing through it



Formally: if C is a nonempty convex set, and $x_0 \in \mathrm{bd}(C)$, then there exists a such that

$$C \subseteq \{x : a^T x \leq a^T x_0\}$$

Operations preserving convexity

- **Intersection**: Convexity is preserved under intersection: if S_1 and S_2 are convex, then $S_1 \cap S_2$ is convex.

  This property extends to the intersection of an infinite number of sets:

  if S_α is convex for every $\alpha \in \mathcal{A}$, then $\bigcap_{\alpha \in \mathcal{A}} S_\alpha$ is convex.

  As a simple example, a polyhedron is the intersection of halfspaces and hyperplanes (which are convex), and therefore is convex.

- **Scaling and translation**:

  If $S \subseteq \mathbf{R}^n$ is convex, $\alpha \in \mathbf{R}$, and $a \in \mathbf{R}^n$, then

  the sets αS and $S + a$ are convex, where

$$\alpha S = \{\alpha x \mid x \in S\}, \qquad S + a = \{x + a \mid x \in S\}.$$

Operations preserving convexity

- Affine images

Recall that a function $f : \mathbf{R}^n \to \mathbf{R}^m$ is *affine* if it is a sum of a linear function and a constant, *i.e.*, if it has the form

$$f(x) = Ax + b, \text{ where } A \in \mathbf{R}^{m \times n} \text{ and } b \in \mathbf{R}^m.$$

Suppose $S \subseteq \mathbf{R}^n$ is convex and $f : \mathbf{R}^n \to \mathbf{R}^m$ is an affine function.

Then the image of S under f,

$$f(S) = \{f(x) \mid x \in S\},$$

is convex.

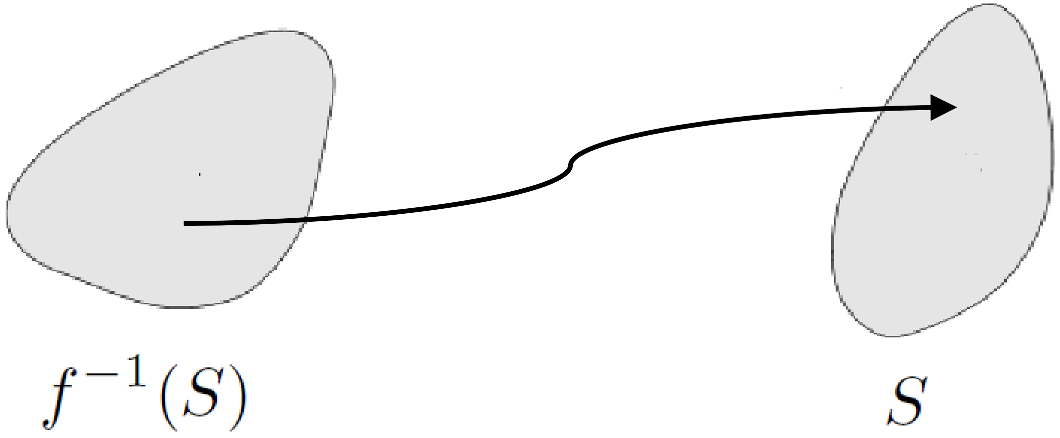
Operations preserving convexity

- Affine preimages:

if $f : \mathbf{R}^k \to \mathbf{R}^n$ is an affine function, the *inverse image* of S under f,

$$f^{-1}(S) = \{x \mid f(x) \in S\},$$

is convex.



Example: linear matrix inequality solution set

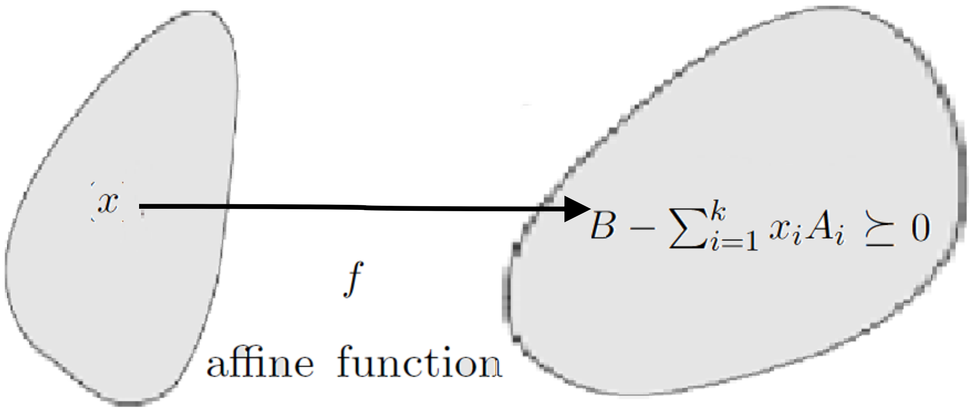
Given $A_1, \ldots A_k, B \in \mathbb{S}^n$, a linear matrix inequality is of the form

$$x_1 A_1 + x_2 A_2 + \ldots + x_k A_k \preceq B$$

for a variable $x \in \mathbb{R}^k$. Let's prove the set C of points x that satisfy the above inequality is convex

Proof:

let $f : R^k \to S^n_+, \; f(x) = B - \sum_{i=1}^k x_i A_i$.



Note that $C = f^{-1}(S^n_+)$, affine preimage of convex set.

Operations preserving convexity

- Projection: The *projection* of a convex set onto some of its coordinates is convex: if $S \subseteq \mathbf{R}^m \times \mathbf{R}^n$ is convex, then

$$T = \{x_1 \in \mathbf{R}^m \mid (x_1, x_2) \in S \text{ for some } x_2 \in \mathbf{R}^n\}$$

is convex.

- Sum: The *sum* of two sets is defined as

$$S_1 + S_2 = \{x + y \mid x \in S_1,\ y \in S_2\}.$$

If S_1 and S_2 are convex, then $S_1 + S_2$ is convex.

Proof: To see this, if S_1 and S_2 are convex, then so is the Cartesian product

$$S_1 \times S_2 = \{(x_1, x_2) \mid x_1 \in S_1,\ x_2 \in S_2\}.$$

The image of this set under the linear function $f(x_1, x_2) = x_1 + x_2$ is the sum $S_1 + S_2$.

More operations preserving convexity

- Perspective images and preimages: the perspective function is $P : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ (where $\mathbb{R}_{++}$ denotes positive reals),

$$P(x, z) = x/z$$

  for $z > 0$. If $C \subseteq \mathrm{dom}(P)$ is convex then so is $P(C)$, and if D is convex then so is $P^{-1}(D)$

- Linear-fractional images and preimages: the perspective map composed with an affine function,

$$f(x) = \frac{Ax + b}{c^T x + d}$$

  is called a linear-fractional function, defined on $c^T x + d > 0$. If $C \subseteq \mathrm{dom}(f)$ is convex then so if $f(C)$, and if D is convex then so is $f^{-1}(D)$

Example: conditional probability set

Let U, V be random variables over $\{1, \ldots n\}$ and $\{1, \ldots m\}$. Let $C \subseteq \mathbb{R}^{nm}$ be a set of joint distributions for U, V, i.e., each $p \in C$ defines joint probabilities

$$p_{ij} = \mathbb{P}(U = i, V = j)$$

Let $D \subseteq \mathbb{R}^{nm}$ contain corresponding conditional distributions, i.e., each $q \in D$ defines

$$q_{ij} = \mathbb{P}(U = i | V = j)$$

Assume C is convex. Let's prove that D is convex. Write

$$D = \left\{ q \in \mathbb{R}^{nm} : q_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}}, \text{ for some } p \in C \right\} = f(C)$$

where f is a linear-fractional function, hence D is convex

Appendix

Some notes from linear algebra

Affine Combination

We refer to a point of the form

$$\theta_1 x_1 + \cdots + \theta_k x_k, \qquad \text{where } \theta_1 + \cdots + \theta_k = 1,$$

as an *affine combination* of the points $x_1, \ldots, x_k$.

Using induction from the definition of affine set (*i.e.*, that it contains every affine combination of two points in it), it can be shown that

an affine set contains every affine combination of its points:

If C is an affine set, and

$x_1, \ldots, x_k \in C$, and

$\theta_1 + \cdots + \theta_k = 1$, then

the point $\theta_1 x_1 + \cdots + \theta_k x_k$ also belongs to C.

Linear Combination and Independence

The vectors $x_1, \ldots, x_m$ are said to be *linearly dependent* when the zero vector can be obtained as a nonzero linear combination of these vectors.

Formally, $x_1, \ldots, x_m$ are linearly dependent when there exists scalars $\alpha_1, \ldots, \alpha_m$ not all equal to zero and such that

$$\alpha_1 x_1 + \ldots + \alpha_m x_m = 0.$$

The vectors $x_1, \ldots, x_m$ are said to be *linearly independent* when they are not linearly dependent.

Formally, they are independent when the equality

$$\alpha_1 x_1 + \ldots + \alpha_m x_m = 0$$

holds *only for* $\alpha_1 = 0, \ldots, \alpha_m = 0.$

$x_0, \ldots, x_k$ are affinely independent means $x_1 - x_0, \ldots, x_k - x_0$ are linear independent.

By restricting the coefficients used in linear combinations, one can define the related concepts of **affine combination**, **conical combination**, and **convex combination**:

Type of combination	Restrictions on coefficients
Linear combination	no restrictions
Affine combination	$\sum a_i = 1$
Conical combination	$a_i \geq 0$
Convex combination	$a_i \geq 0$ and $\sum a_i = 1$

Quadratic Forms and Positive Semidefinite Matrices

Given a square matrix $A \in \mathbb{R}^{n\times n}$ and a vector $x \in \mathbb{R}^n$, the scalar value x^TAx is called a ***quadratic form***.

Written explicitly, we see that

$$x^TAx = \sum_{i=1}^{n} x_i(Ax)_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij}x_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_ix_j \ .$$

Note that,

$$x^TAx = (x^TAx)^T = x^TA^Tx$$

the transpose of a scalar is equal to itself

$$\Rightarrow 2x^TAx = x^TAx + x^TA^Tx = x^T(A + A^T)x$$

$$\Rightarrow x^TAx = x^T\frac{1}{2}(A + A^T)x$$

$(A + A^T)$ is always symmetric no matter what matrix A would be!

For this reason, we often implicitly assume that the matrices appearing in a quadratic form are symmetric.

Positive Semidefinite and Positive Definite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is ***positive definite*** (PD) if for all non-zero vectors $x \in \mathbb{R}^n$, $x^T A x > 0$.

  This is usually denoted $A \succ 0$ (or just $A > 0$), and

- A symmetric matrix $A \in \mathbb{S}^n$ is ***positive semidefinite*** (PSD) if for all vectors $x^T A x \geq 0$.

  This is written $A \succeq 0$ (or just $A \geq 0$), and

Positive Semidefinite and Positive Definite Matrices

We use the notation $\mathbf{S}^n$ to denote the set of symmetric $n \times n$ matrices,

$$\mathbf{S}^n = \{X \in \mathbf{R}^{n\times n} \mid X = X^T\},$$

We use the notation $\mathbf{S}^n_+$ to denote the set of symmetric positive semidefinite matrices:

$$\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\},$$

and the notation $\mathbf{S}^n_{++}$ to denote the set of symmetric positive definite matrices:

$$\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}.$$

(This notation is meant to be analogous to $\mathbf{R}_+$, which denotes the nonnegative reals, and $\mathbf{R}_{++}$, which denotes the positive reals.)

Negative Semidefinite, Negative Definite, and Indefinite Matrices

- Likewise, a symmetric matrix $A \in \mathbb{S}^n$ is ***negative definite*** (ND), denoted $A \prec 0$ (or just $A < 0$) if for all non-zero $x \in \mathbb{R}^n$, $x^T Ax < 0$.

- Similarly, a symmetric matrix $A \in \mathbb{S}^n$ is ***negative semidefinite*** (NSD), denoted $A \preceq 0$ (or just $A \leq 0$) if for all $x \in \mathbb{R}^n$, $x^T Ax \leq 0$.

- Finally, a symmetric matrix $A \in \mathbb{S}^n$ is ***indefinite***, if it is neither positive semidefinite nor negative semidefinite — i.e.,

  if there exists $x_1, x_2 \in \mathbb{R}^n$ such that $x_1^T Ax_1 > 0$ and $x_2^T Ax_2 < 0$.

It should be obvious that if A is positive definite, then

$-A$ is negative definite and vice versa.

Likewise, if A is positive semidefinite then

$-A$ is negative semidefinite and vice versa.

If A is indefinite, then so is $-A$.