CS 2601 Linear and Convex Optimization

3. Convex sets

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

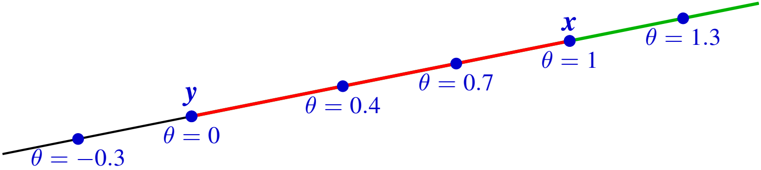
Outline

- Convex Sets
- Supporting and separating hyperplanes

Lines, line segments and rays

Given $x \neq y \in \mathbb{R}^n$, the line passing through x and y consists of points of the form

$$z = y + \theta(x - y) = \theta x + (1 - \theta)y, \quad \theta \in \mathbb{R}$$



The ray (half-line) with endpoint y and direction $x - y$ consists of points

$$\theta x + (1 - \theta)y, \quad \theta \geq 0$$

The line segment between x and y consists of points

$$\theta x + (1 - \theta)y, \quad 0 \leq \theta \leq 1$$

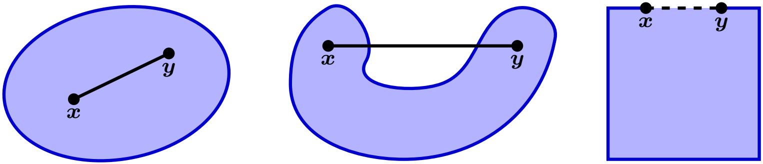
Note. Often use notation $\bar{\theta} = 1 - \theta$.

Convex sets

A set $C \subset \mathbb{R}^n$ is convex if the line segment between any two points $\boldsymbol{x}, \boldsymbol{y} \in C$ lies entirely in C, i.e.

$$\boldsymbol{x} \in C, \boldsymbol{y} \in C, \theta \in [0,1] \implies \theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y} \in C$$

Note. Only need to check the case that $\boldsymbol{x} \neq \boldsymbol{y}$ and $\theta \in (0,1)$.



convex nonconvex nonconvex

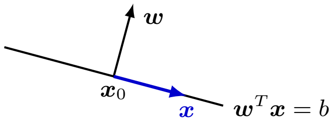
For $\theta \in [0,1]$, $\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}$ is called a convex combination of $\boldsymbol{x}$ and $\boldsymbol{y}$. In a more symmetric form, a convex combination is

$$\theta_1 \boldsymbol{x} + \theta_2 \boldsymbol{y} \quad \text{where } \theta_1 \geq 0, \theta_2 \geq 0, \theta_1 + \theta_2 = 1$$

Examples of convex sets

Example. Trivial examples of convex sets include empty set, $\mathbb{R}^n$, singletons (points), lines, line segments and rays.

Example. A hyperplane $P = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{w}^T\boldsymbol{x} = b\}$ is convex, where $\boldsymbol{w} \in \mathbb{R}^n$, $b \in \mathbb{R}$.

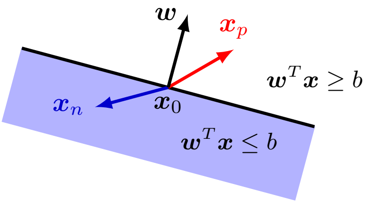


Proof. For $\boldsymbol{x}_1, \boldsymbol{x}_2 \in P$ and $\theta \in [0, 1]$,

$$\begin{aligned} \boldsymbol{w}^T(\theta\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2) &= \theta\boldsymbol{w}^T\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{w}^T\boldsymbol{x}_2 \\ &= \theta b + \bar{\theta} b = b \end{aligned}$$

Example: Halfspaces

A halfspace $H = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{w}^T\boldsymbol{x} \le b\}$ is convex.



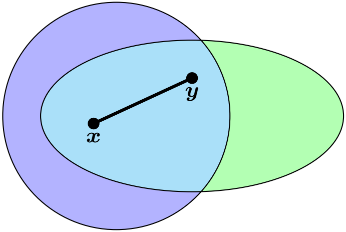
Note. $H = \{\boldsymbol{x} : f(\boldsymbol{x}) \le b\}$ is the so-called sublevel set of $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$.
Note $\nabla f(\boldsymbol{x}) = \boldsymbol{w}$, the outward normal to the boundary hyperplane.

Proof. For $\boldsymbol{x}_1, \boldsymbol{x}_2 \in H$ and $\theta \in [0, 1]$,

$$\begin{aligned} \boldsymbol{w}^T(\theta\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2) &= \theta\boldsymbol{w}^T\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{w}^T\boldsymbol{x}_2 \\ &\le \theta b + \bar{\theta} b = b \end{aligned}$$

Set intersection preserves convexity

Proposition. The intersection of an arbitrary collection of convex sets is convex.



Proof. Let $\{C_i : i \in I\}$ be an arbitrary collection of convex sets with index set I, and $C = \bigcap_{i \in I} C_i$ their intersection.

- Let $\boldsymbol{x}, \boldsymbol{y} \in C$, $\theta \in [0, 1]$
- $\boldsymbol{x}, \boldsymbol{y} \in C_i$ for any $i \in I$
- By convexity of C_i, $\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y} \in C_i$
- $\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y} \in C$

Example: Affine spaces

An affine space $S = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$, $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ is convex.

Note. An affine space is a shifted linear space, $S = \boldsymbol{x}_0 + S_0$, where $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}$, and $S_0 = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}$ is a linear space.

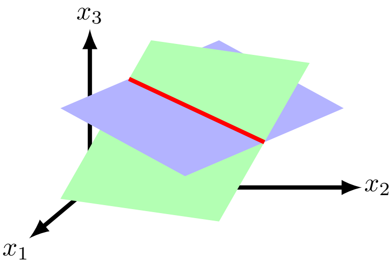
Can verify convexity by definition; here use the intersection property.

- let $\boldsymbol{A}^T = (\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_m)$,
  $\boldsymbol{b} = (b_1, \dots, b_m)^T$
- S is intersection of m hyperplanes

$$S = \bigcap_{i=1}^{m} P_i$$

where

$$P_i = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}_i^T \boldsymbol{x} = b_i\}$$



Note. An affine space S actually contains the line through any $\boldsymbol{x}, \boldsymbol{y} \in S$.

Example: Polyhedra

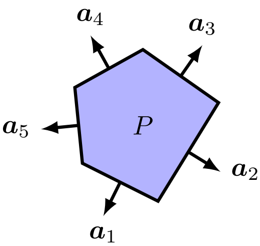
A polyhedron $P = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}\}$, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^m$ is convex, where vector inequality $\le$ is interpreted componentwise

- let $\boldsymbol{A}^T = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m)$, $\boldsymbol{b} = (b_1, \ldots, b_m)^T$
- P is intersection of m halfspaces

$$P = \bigcap_{i=1}^{m} H_i$$

where

$$H_i = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}_i^T \boldsymbol{x} \le b_i\}$$



Note. An affine space $S = \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$ is a polyhedron

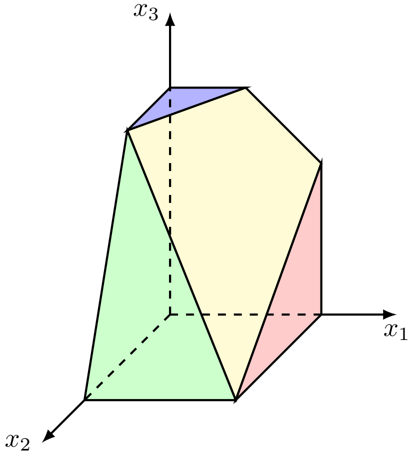
$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \iff (\boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b} \text{ and } -\boldsymbol{A}\boldsymbol{x} \le -\boldsymbol{b}) \iff \begin{pmatrix} \boldsymbol{A} \\ -\boldsymbol{A} \end{pmatrix} \boldsymbol{x} \le \begin{pmatrix} \boldsymbol{b} \\ -\boldsymbol{b} \end{pmatrix}$$

More generally, $\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}, \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}\}$ is a polyhedron.

Example: Polyhedra (cont'd)

The set of points satisfying the following inequalities is a polyhedron

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 4 \\ x_1 &\leq 2 \\ x_3 &\leq 3 \\ 3x_2 + x_3 &\leq 6 \\ \boldsymbol{x} &\geq \mathbf{0} \end{aligned}$$

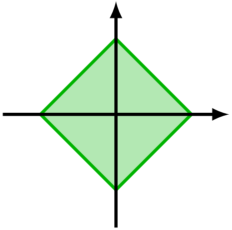


Example: Polyhedra (cont'd)

The 1-norm unit ball is a polyhedron,

$$B_1 = \{\boldsymbol{x} : \|\boldsymbol{x}\|_1 \le 1\}$$

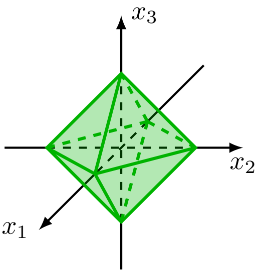
In 2d,



$$\begin{aligned} B_1^{(2)} = \{\boldsymbol{x} : x_1 + x_2 &\le 1, \\ x_1 - x_2 &\le 1, \\ -x_1 + x_2 &\le 1, \\ -x_1 - x_2 &\le 1\} \end{aligned}$$

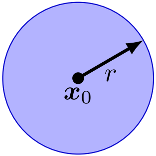
In 3d,

$$B_1^{(3)} = \{\boldsymbol{x} : \pm x_1 \pm x_2 \pm x_3 \le 1\}$$



Example: Norm balls

A closed ball $\bar{B}(\boldsymbol{x}_0, r) = \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x} - \boldsymbol{x}_0\| \le r\}$ is convex.



Proof. For $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \bar{B}(\boldsymbol{x}_0, r)$ and $\theta \in [0, 1]$,

$$\begin{aligned} \|(\theta \boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2) - \boldsymbol{x}_0\| &= \|\theta(\boldsymbol{x}_1 - \boldsymbol{x}_0) + \bar{\theta}(\boldsymbol{x}_2 - \boldsymbol{x}_0)\| \\ &\le \|\theta(\boldsymbol{x}_1 - \boldsymbol{x}_0)\| + \|\bar{\theta}(\boldsymbol{x}_2 - \boldsymbol{x}_0)\| \\ &= \theta\|\boldsymbol{x}_1 - \boldsymbol{x}_0\| + \bar{\theta}\|\boldsymbol{x}_2 - \boldsymbol{x}_0\| \le r \end{aligned}$$

Note. True for any norm $\|\cdot\|$.

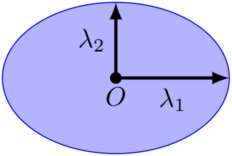
Note. Open balls are also convex.

Example: Ellipsoids

An ellipsoid

$$\mathcal{E} = \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \frac{x_1^2}{\lambda_1^2} + \frac{x_2^2}{\lambda_2^2} \leq 1 \right\}$$

is convex.



Proof. Let $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \lambda_2\}$. Note

$$\mathcal{E} = \{\boldsymbol{x} : \boldsymbol{x}^T \boldsymbol{\Lambda}^{-2} \boldsymbol{x} \leq 1\} = \{\boldsymbol{x} : \|\boldsymbol{\Lambda}^{-1}\boldsymbol{x}\|_2 \leq 1\} = \{\boldsymbol{\Lambda}\boldsymbol{u} : \|\boldsymbol{u}\|_2 \leq 1\}.$$

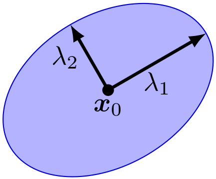
For $\boldsymbol{x}_i = \boldsymbol{\Lambda}\boldsymbol{u}_i \in \mathcal{E}$, and $\theta \in [0, 1]$,

$$\theta \boldsymbol{x}_1 + \bar{\theta} \boldsymbol{x}_2 = \boldsymbol{\Lambda}(\theta \boldsymbol{u}_1 + \bar{\theta} \boldsymbol{u}_2).$$

Recall the unit ball is convex, so $\|\theta \boldsymbol{u}_1 + \bar{\theta} \boldsymbol{u}_2\|_2 \leq 1$ and $\theta \boldsymbol{x}_1 + \bar{\theta} \boldsymbol{x}_2 \in \mathcal{E}$.

Example: Ellipsoids (cont'd)

An ellipsoid $\mathcal{E} = \{\boldsymbol{x}_0 + \boldsymbol{A}\boldsymbol{u} : \|\boldsymbol{u}\|_2 \leq 1\}$, $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{A} \succ \boldsymbol{O}$ is convex.



$\boldsymbol{A}$ has eigendecomposition $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$, where $\boldsymbol{\Lambda}$ is diagonal and $\boldsymbol{Q}$ is orthogonal. With $\tilde{\boldsymbol{u}} = \boldsymbol{Q}^T\boldsymbol{u}$,

$$\mathcal{E} = \{\boldsymbol{x}_0 + \boldsymbol{Q}\boldsymbol{\Lambda}\tilde{\boldsymbol{u}} : \|\tilde{\boldsymbol{u}}\|_2 \leq 1\},$$

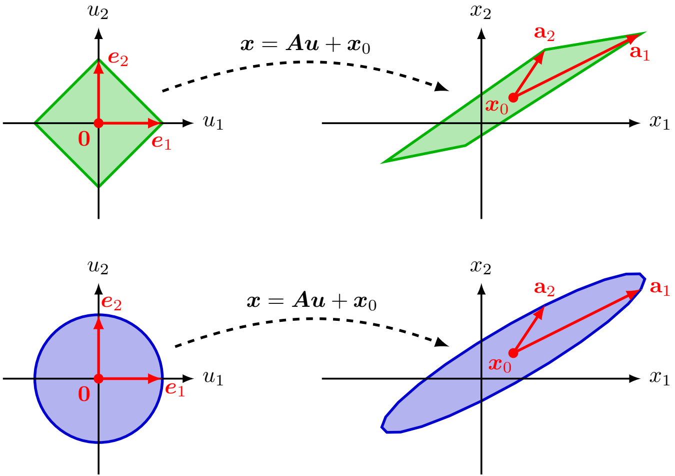
which is a rotated and shifted version of $\mathcal{E}' = \{\boldsymbol{\Lambda}\tilde{\boldsymbol{u}} : \|\tilde{\boldsymbol{u}}\|_2 \leq 1\}$.

Note. The lengths of semi-axes are eigenvalues of $\boldsymbol{A}$

Note. Also often written as $\mathcal{E} = \{\boldsymbol{x} : (\boldsymbol{x} - \boldsymbol{x}_0)^T\boldsymbol{P}^{-1}(\boldsymbol{x} - \boldsymbol{x}_0) \leq 1\}$, $\boldsymbol{P} = \boldsymbol{A}^2$.

Affine transformation preserves convexity

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{x}_0, \quad \text{where } \boldsymbol{A} = \begin{pmatrix} 2 & 0.5 \\ 1 & 0.75 \end{pmatrix}, \quad \boldsymbol{x}_0 = (0.5, 0.4)^T$$



Affine transformation preserves convexity

Proposition. The image of a convex set under an affine transformation is convex.

Proof. Let $C \subset \mathbb{R}^n$ be a convex set and $f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ an affine transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Given $\boldsymbol{y}_1, \boldsymbol{y}_2 \in f(C) = \{f(\boldsymbol{x}) : \boldsymbol{x} \in C\}$ and $\theta \in [0, 1]$, need to show $\theta \boldsymbol{y}_1 + \bar{\theta} \boldsymbol{y}_2 \in f(C)$.

1. By definition, $\boldsymbol{y}_i = f(\boldsymbol{x}_i)$ for some $\boldsymbol{x}_i \in C$, $i = 1, 2$.
2. Since f is affine,

$$\begin{aligned}\theta \boldsymbol{y}_1 + \bar{\theta} \boldsymbol{y}_2 &= \theta f(\boldsymbol{x}_1) + \bar{\theta} f(\boldsymbol{x}_2) \\ &= \theta(\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}) + \bar{\theta}(\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}) \\ &= \boldsymbol{A}(\theta \boldsymbol{x}_1 + \bar{\theta}_2 \boldsymbol{x}_2) + \boldsymbol{b}\end{aligned}$$

3. Since C is convex, $\boldsymbol{z} \triangleq \theta \boldsymbol{x}_1 + \bar{\theta} \boldsymbol{x}_2 \in C$, so $\theta \boldsymbol{y}_1 + \bar{\theta} \boldsymbol{y}_2 = f(\boldsymbol{z}) \in f(C)$.

Proposition. The inverse image of a convex set under an affine transformation is convex.

Example: Positive semidefinite matrices

The set of positive semidefinite matrices

$$\mathcal{S}_+^n = \{\boldsymbol{A} \in \mathbb{R}^{n \times n} : \boldsymbol{A} \succeq \boldsymbol{O}\}$$

is convex.

Proof. For arbitrary $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{S}_+^n$ and $\theta \in [0, 1]$, $\boldsymbol{x} \in \mathbb{R}^n$, need to show $\theta\boldsymbol{A} + \bar{\theta}\boldsymbol{B} \in \mathcal{S}_+^n$. Check the definition of positive semidefiniteness.

1. $\theta\boldsymbol{A} + \bar{\theta}\boldsymbol{B}$ is symmetric,

$$(\theta\boldsymbol{A} + \bar{\theta}\boldsymbol{B})^T = \theta\boldsymbol{A}^T + \bar{\theta}\boldsymbol{B}^T = \theta\boldsymbol{A} + \bar{\theta}\boldsymbol{B}$$

2. $\boldsymbol{x}^T(\theta\boldsymbol{A} + \bar{\theta}\boldsymbol{B})\boldsymbol{x} \geq 0$ for any $\boldsymbol{x}$,

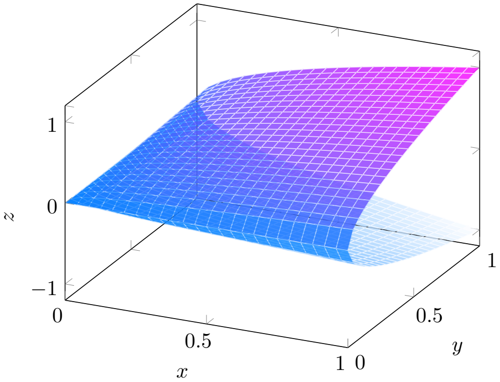
$$\boldsymbol{x}^T(\theta\boldsymbol{A} + \bar{\theta}\boldsymbol{B})\boldsymbol{x} = \theta(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) + \bar{\theta}(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}) \geq 0$$

Example: Positive semidefinite matrices (cont'd)

For $n = 2$, can identify $\mathcal{S}_+^2$ with a subset of $\mathbb{R}^3$. By Sylvester's Theorem,

$$\boldsymbol{A} = \begin{pmatrix} x & z \\ z & y \end{pmatrix} \in \mathcal{S}_+^2 \iff x \geq 0,\ y \geq 0,\ xy \geq z^2$$

Boundary $\partial\mathcal{S}_+^2 = \{(x, y, z) : x \geq 0, y \geq 0, z^2 = xy\}$



Convex combination

A convex combination of $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ is a point of the form

$$\sum_{i=1}^{m} \theta_i \boldsymbol{x}_i = \theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 + \cdots + \theta_m \boldsymbol{x}_m$$

where $\theta_i \geq 0$ for all i and $\sum_{i=1}^{m} \theta_i = 1$.

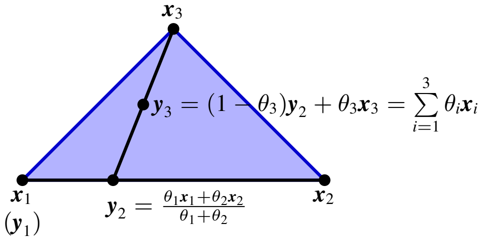
Theorem. If C is convex, and $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in C$, then any convex combination $\sum_{i=1}^{m} \theta_i \boldsymbol{x}_i \in C$.

In general, $\boldsymbol{y}_1 = \boldsymbol{x}_1$, and

$$\boldsymbol{y}_k = \frac{\sigma_{k-1}}{\sigma_k} \boldsymbol{y}_{k-1} + \frac{\theta_k}{\sigma_k} \boldsymbol{x}_k, \quad k \geq 2$$

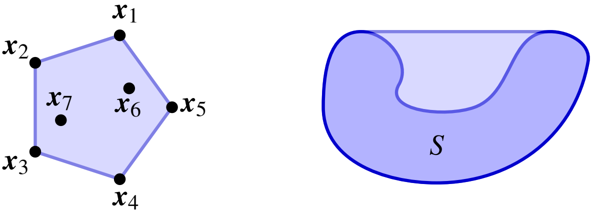
where

$$\sigma_k = \sum_{i=1}^{k} \theta_i$$



Convex hull

The convex hull of a set $S \subset \mathbb{R}^n$, denoted $\operatorname{conv} S$, is the smallest convex set containing S.



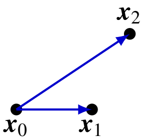
Theorem. $\operatorname{conv} S$ is the set of all convex combinations of points in S, i.e.

$$\operatorname{conv} S = \left\{ \sum_{i=1}^{m} \theta_i \boldsymbol{x}_i : m \in \mathbb{N}; \boldsymbol{x}_i \in S, \theta_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^{m} \theta_i = 1 \right\}$$

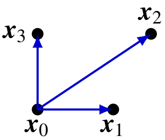
Note. Actually we can replace $m \in \mathbb{N}$ by $m \leq n+1$ in the above representation, i.e. each $\boldsymbol{x} \in \operatorname{conv} S$ is the convex combination of at most $n+1$ points in S.

Affinely independent points

$m+1$ points $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ are affinely independent if $\boldsymbol{x}_1 - \boldsymbol{x}_0, \ldots, \boldsymbol{x}_m - \boldsymbol{x}_0$ are linearly independent.



affinely independent points in $\mathbb{R}^2$



affinely dependent points in $\mathbb{R}^2$

Proposition. $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \mathbb{R}^n$ are affinely independent iff

$$\sum_{i=0}^{m} c_i \boldsymbol{x}_i = \mathbf{0} \text{ and } \sum_{i=0}^{m} c_i = 0 \implies c_i = 0 \text{ for } i = 0, 1, \ldots, m$$

Note. In $\mathbb{R}^n$, the maximum number of linearly independent vectors is n, so the maximum number of affinely independent points is $n+1$.

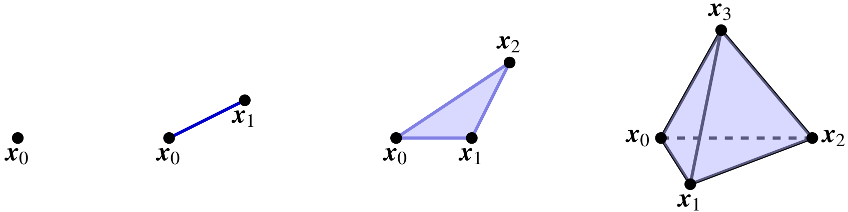
Simplexes

An m-dimensional simplex, also called an m-simplex, is the convex hull of $m+1$ affinely independent points. More specifically, the simplex determined by affinely independent points $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ is

$$\operatorname{conv}\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_m\} = \{\theta_0 \boldsymbol{x}_0 + \theta_1 \boldsymbol{x}_1 + \cdots + \theta_m \boldsymbol{x}_m : \boldsymbol{\theta} \geq \mathbf{0}, \mathbf{1}^T \boldsymbol{\theta} = 1\}$$

Note. $\mathbb{R}^n$ only has m-simplexes with $m \leq n$

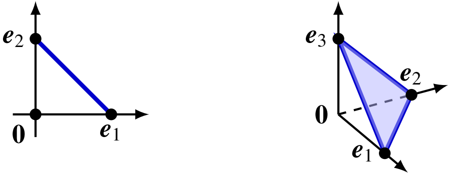
- 0-simplexes are points
- 1-simplexes are line segments
- 2-simplexes are triangles
- 3-simplexes are tetrahedra



Simplexes (cont'd)

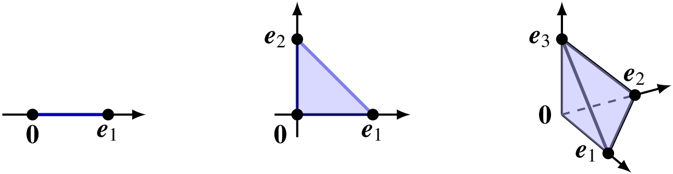
Example. The probability n-simplex is the n-simplex in $\mathbb{R}^{n+1}$ determined by the standard basis vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n+1}$,

$$\Delta_n = \{\boldsymbol{\theta} \in \mathbb{R}^{n+1} : \boldsymbol{\theta} \geq \mathbf{0},\ \mathbf{1}^T\boldsymbol{\theta} = 1\}$$



Example. The unit n-simplex in $\mathbb{R}^n$ is the n-simplex determined by $\mathbf{0} \in \mathbb{R}^n$ and the standard basis vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n \in \mathbb{R}^n$,

$$\Delta'_n = \{\boldsymbol{\theta}' \in \mathbb{R}^n : \boldsymbol{\theta}' \geq \mathbf{0},\ \mathbf{1}^T\boldsymbol{\theta}' \leq 1\}$$



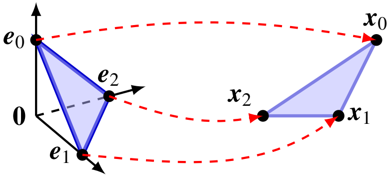
Simplexes (cont'd)

The m-simplex in $\mathbb{R}^n$ determined by affinely independent points $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ is the image of Δ_m under a linear transformation

$$\boldsymbol{\theta} = \sum_{i=0}^{m} \theta_i \boldsymbol{e}_i \mapsto \boldsymbol{x} = \sum_{i=0}^{m} \theta_i \boldsymbol{x}_i = \boldsymbol{X}\boldsymbol{\theta}$$

where

$$\boldsymbol{X} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \in \mathbb{R}^{n \times (m+1)}, \quad \boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_m)^T \in \Delta_m$$



Simplexes (cont'd)

Note

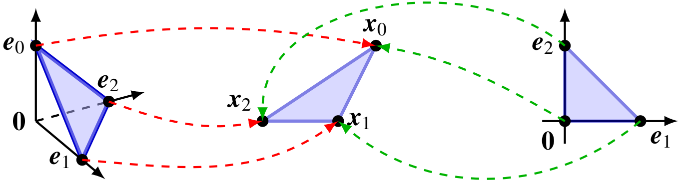
$$\boldsymbol{x} = \sum_{i=0}^{m} \theta_i \boldsymbol{x}_i = \boldsymbol{x}_0 + \sum_{i=1}^{m} \theta_i (\boldsymbol{x}_i - \boldsymbol{x}_0)$$

and $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_m)^T \in \Delta'_m$.

The simplex $\operatorname{conv}\{\boldsymbol{x}_0, \ldots, \boldsymbol{x}_m\}$ is also the image of Δ'_m under the affine transformation

$$\boldsymbol{\theta}' \mapsto \boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{B}\boldsymbol{\theta}'$$

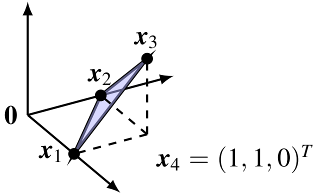
where $\boldsymbol{B} = (\boldsymbol{x}_1 - \boldsymbol{x}_0, \ldots, \boldsymbol{x}_m - \boldsymbol{x}_0) \in \mathbb{R}^{n \times m}$.



Simplexes (cont'd)

Example. Let $\boldsymbol{x}_1 = (1,0,0)^T$, $\boldsymbol{x}_2 = (0,1,0)^T$ and $\boldsymbol{x}_3 = (1,1,1)^T$. Points in the 2-simplex determined by $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are of the form

$$\boldsymbol{x} = \sum_{i=1}^{3} \theta_i \boldsymbol{x}_i = (\theta_1 + \theta_3, \theta_2 + \theta_3, \theta_3)^T$$



where $\boldsymbol{\theta} \in \Delta_2$, i.e. $\theta_i \geq 0$, $\theta_1 + \theta_2 + \theta_3 = 1$.

Alternatively,

$$\boldsymbol{x} = \boldsymbol{x}_1 + \theta_2(\boldsymbol{x}_2 - \boldsymbol{x}_1) + \theta_3(\boldsymbol{x}_3 - \boldsymbol{x}_1) = (1 - \theta_2, \theta_2 + \theta_3, \theta_3)^T = \boldsymbol{x}_1 + \boldsymbol{B}\boldsymbol{\theta}',$$

where

$$\boldsymbol{B} = (\boldsymbol{x}_2 - \boldsymbol{x}_1, \boldsymbol{x}_3 - \boldsymbol{x}_1) = \begin{pmatrix} -1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}, \ \boldsymbol{\theta}' = (\theta_2, \theta_3)^T \in \Delta_2'$$

Note $\boldsymbol{B}$ has full column rank by the affine independence of the $\boldsymbol{x}_i$'s.

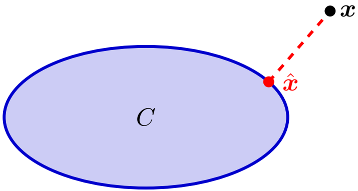
Outline

- Convex Sets
- Supporting and separating hyperplanes

Projection onto convex set

Given a set $C \subset \mathbb{R}^n$, the distance between a point $\boldsymbol{x}$ and C is

$$\mathrm{dist}(\boldsymbol{x}, C) = \inf_{\boldsymbol{z} \in C} \|\boldsymbol{x} - \boldsymbol{z}\|$$



Theorem. If $C \subset \mathbb{R}^n$ is nonempty, closed and convex, then for any $\boldsymbol{x}$, there is a unique $\hat{\boldsymbol{x}} \in C$ s.t.

$$\mathrm{dist}(\boldsymbol{x}, C) = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|$$

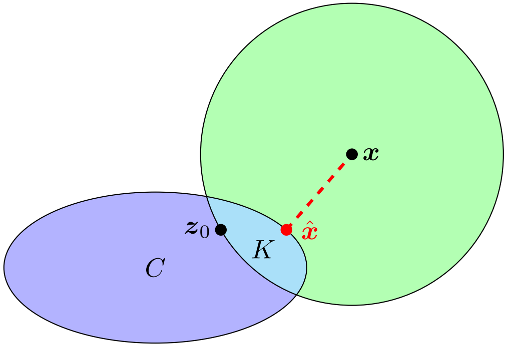
$\hat{\boldsymbol{x}}$ is called the projection of $\boldsymbol{x}$ onto C and denoted by $\hat{\boldsymbol{x}} = \mathcal{P}_C(\boldsymbol{x})$.

Note. $\mathcal{P}_C(\boldsymbol{x}) = \boldsymbol{x}$ iff $\boldsymbol{x} \in C$.

Projection onto convex set (cont'd)

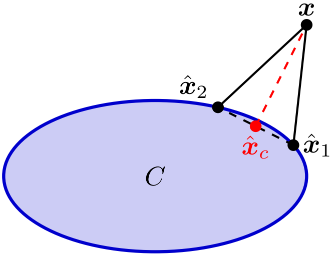
Proof. First show existence.

- Let $\boldsymbol{z}_0 \in C$. Then $\mathrm{dist}(\boldsymbol{x}, C) \le \|\boldsymbol{x} - \boldsymbol{z}_0\|$.
- Let $K = \{\boldsymbol{z} \in C : \|\boldsymbol{x} - \boldsymbol{z}\| \le \|\boldsymbol{x} - \boldsymbol{z}_0\|\} = C \cap \bar{B}(\boldsymbol{x}, \|\boldsymbol{x} - \boldsymbol{z}_0\|)$.
- $\|\boldsymbol{x} - \boldsymbol{z}\|$ is continuous in $\boldsymbol{z}$, K compact $\implies \exists \hat{\boldsymbol{x}} \in K$ such that $\mathrm{dist}(\boldsymbol{x}, C) = \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|$



Projection onto convex set (cont'd)

Proof (cont'd). Now show uniqueness. Suppose $\hat{x}_1, \hat{x}_2 \in C$ satisfy $\mathrm{dist}(x, C) = \|x - \hat{x}_1\| = \|x - \hat{x}_2\|$.



- $\hat{x}_c := \frac{\hat{x}_1 + \hat{x}_2}{2} \in C$ by the convexity of C, so

$$\|x - \hat{x}_c\| \geq \mathrm{dist}(x, C)$$

- By the polarization identity

$$\|y + z\|^2 + \|y - z\|^2 = 2\|y\|^2 + 2\|z\|^2$$

we obtain

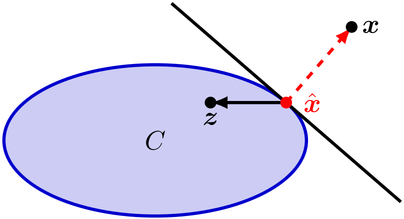
$$\begin{aligned} \|\hat{x}_1 - \hat{x}_2\|^2 &= 2\|x - \hat{x}_1\|^2 + 2\|x - \hat{x}_2\|^2 - \|2x - (\hat{x}_1 + \hat{x}_2)\|^2 \\ &= 4[\mathrm{dist}(x, C)]^2 - 4\|x - \hat{x}_c\|^2 \leq 0 \end{aligned}$$

so $\hat{x}_1 = \hat{x}_2$.

Projection onto convex set (cont'd)

Proposition. Let $C \subset \mathbb{R}^n$ be nonempty, closed and convex. Given $\hat{\boldsymbol{x}} \in C$, $\hat{\boldsymbol{x}} = \mathcal{P}_C(\boldsymbol{x})$ iff

$$\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{z} - \hat{\boldsymbol{x}} \rangle \leq 0, \quad \forall \boldsymbol{z} \in C$$



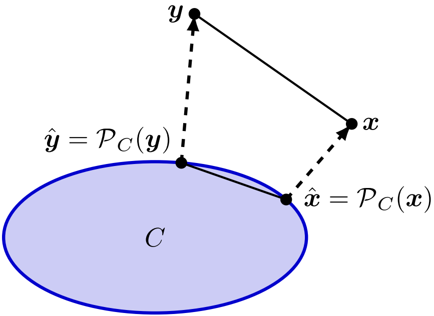
Proof. Note $\hat{\boldsymbol{x}} + t(\boldsymbol{z} - \hat{\boldsymbol{x}}) \in C$, $\forall \boldsymbol{z} \in C, t \in [0,1]$.

$$\begin{aligned}
\hat{\boldsymbol{x}} = \mathcal{P}_C(\boldsymbol{x}) &\iff \|\boldsymbol{x} - \hat{\boldsymbol{x}} - t(\boldsymbol{z} - \hat{\boldsymbol{x}})\|^2 \geq \|\boldsymbol{x} - \hat{\boldsymbol{x}}\|^2, \quad \forall \boldsymbol{z} \in C, t \in [0,1] \\
&\iff -2t\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{z} - \hat{\boldsymbol{x}} \rangle + t^2 \|\boldsymbol{z} - \hat{\boldsymbol{x}}\|^2 \geq 0, \quad \forall \boldsymbol{z} \in C, t \in [0,1] \\
&\iff \langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \boldsymbol{z} - \hat{\boldsymbol{x}} \rangle \leq 0, \quad \forall \boldsymbol{z} \in C
\end{aligned}$$

Projection onto convex set (cont'd)

Corollary. The projection operator is nonexpansive, i.e.

$$\| \mathcal{P}_C(\boldsymbol{x}) - \mathcal{P}_C(\boldsymbol{y}) \| \le \|\boldsymbol{x} - \boldsymbol{y}\|.$$



Proof. Let $\hat{\boldsymbol{x}} = \mathcal{P}_C(\boldsymbol{x}), \hat{\boldsymbol{y}} = \mathcal{P}_C(\boldsymbol{y})$. By the proposition on slide 30,

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{y}\|^2 &= \|\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\|^2 + \|\boldsymbol{x} - \boldsymbol{y} - (\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}})\|^2 + 2\langle \boldsymbol{x} - \boldsymbol{y} - (\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}), \hat{\boldsymbol{x}} - \hat{\boldsymbol{y}} \rangle \\ &\ge \|\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\|^2 - 2 \underbrace{\langle \boldsymbol{y} - \hat{\boldsymbol{y}}, \hat{\boldsymbol{x}} - \hat{\boldsymbol{y}} \rangle}_{\le 0} - 2 \underbrace{\langle \boldsymbol{x} - \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} - \hat{\boldsymbol{x}} \rangle}_{\le 0} \ge \|\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}}\|^2 \end{aligned}$$

Projection onto convex set (cont'd)

Corollary. Let $C \subset \mathbb{R}^n$ be nonempty, closed and convex. For $\boldsymbol{x}_0 \notin C$, there exists a $\boldsymbol{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ s.t.

$$\sup_{\boldsymbol{x} \in C} \langle \boldsymbol{w}, \boldsymbol{x} \rangle < \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle.$$

Note. Special case of the separating hyperplane theorem on slide 36.

Proof. Let $\hat{\boldsymbol{x}}_0 = \mathcal{P}_C(\boldsymbol{x}_0)$ and $\boldsymbol{w} = \boldsymbol{x}_0 - \hat{\boldsymbol{x}}_0$. Since $\boldsymbol{x}_0 \notin C$ and $\hat{\boldsymbol{x}}_0 \in C$, $\boldsymbol{w} \neq \mathbf{0}$. By the proposition on slide 30, for any $\boldsymbol{x} \in C$,

$$\langle \boldsymbol{w}, \boldsymbol{x} - \hat{\boldsymbol{x}}_0 \rangle \leq 0$$

so

$$\langle \boldsymbol{w}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{w}, \hat{\boldsymbol{x}}_0 \rangle = \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle - \langle \boldsymbol{w}, \boldsymbol{w} \rangle$$

Taking supremum over C,

$$\sup_{\boldsymbol{x} \in C} \langle \boldsymbol{w}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle - \langle \boldsymbol{w}, \boldsymbol{w} \rangle < \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle$$

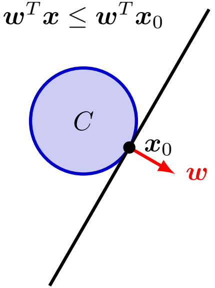
Supporting hyperplane

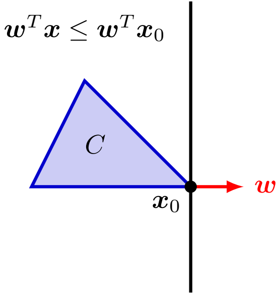
The boundary of a set C is $\partial C = \overline{C} \setminus \operatorname{int} C$.

Supporting hyperplane theorem. If C is a nonempty, convex set in $\mathbb{R}^n$ and $\boldsymbol{x}_0 \in \partial C$, then there exists $\boldsymbol{w} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ s.t.

$$\langle \boldsymbol{w}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle, \quad \forall \boldsymbol{x} \in C$$

$P = \{\boldsymbol{x} : \boldsymbol{w}^T \boldsymbol{x} = \boldsymbol{w}^T \boldsymbol{x}_0\}$ is called a supporting hyperplane to C at $\boldsymbol{x}_0$.



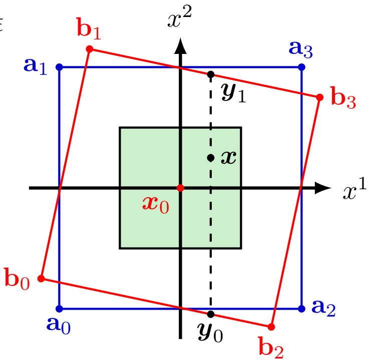


Supporting hyperplane (cont'd)

Lemma. If C is convex, then $\operatorname{int} C = \operatorname{int} \overline{C}$ and $\partial C = \partial \overline{C}$.

Proof sketch for 2D. First show $\operatorname{int} C = \operatorname{int} \overline{C}$. Let $\boldsymbol{x}_0 \in \operatorname{int} \overline{C}$. Show $\boldsymbol{x}_0 \in \operatorname{int} C$. WLOG, assume $\boldsymbol{x}_0 = \mathbf{0}$. Let $\bar{B}(r) = \{\boldsymbol{x} : \|\boldsymbol{x}\|_\infty \le r\}$.

1. $\exists \epsilon > 0$ s.t. $\bar{B}(2\epsilon) \subset \overline{C}$
2. Let $\boldsymbol{a}_i$ denote the vertices of $\bar{B}(2\epsilon)$
3. $\boldsymbol{a}_i \in \overline{C} \implies \exists \boldsymbol{b}_i \in C$ s.t. $\|\boldsymbol{a}_i - \boldsymbol{b}_i\|_\infty < \epsilon$
4. Show $\bar{B}(\epsilon) \subset \operatorname{conv}\{\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ and hence $\boldsymbol{x}_0 \in \operatorname{int} C$
   - Let $\boldsymbol{x} \in \bar{B}(\epsilon)$
   - Find θ_i s.t. $x^1 = \theta_0 b_0^1 + \theta_2 b_2^1 = \theta_1 b_1^1 + \theta_3 b_3^1$
   - Let $\boldsymbol{y}_0 = \theta_0 \boldsymbol{b}_0 + \theta_2 \boldsymbol{b}_2$, $\boldsymbol{y}_1 = \theta_1 \boldsymbol{b}_1 + \theta_3 \boldsymbol{b}_3$
   - Find α s.t. $x^2 = \alpha y_0^2 + \bar{\alpha} y_1^2$
   - $\boldsymbol{x} = \alpha \boldsymbol{y}_0 + \bar{\alpha} \boldsymbol{y}_1 \in \operatorname{conv}\{\boldsymbol{b}_0, \boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$



Then $\partial C = \overline{C} \setminus \operatorname{int} C = \overline{C} \setminus \operatorname{int} \overline{C} = \partial \overline{C}$.

Supporting hyperplane (cont'd)

Proof of theorem.

- $\boldsymbol{x}_0 \in \partial C \implies \boldsymbol{x}_0 \in \partial \overline{C}$ by the previous lemma.
- There exists a sequence $\{\boldsymbol{x}_k\}$ s.t. $\boldsymbol{x}_k \notin \overline{C}$ and $\boldsymbol{x}_k \to \boldsymbol{x}_0$ as $k \to \infty$.
- By the corollary on slide 32, there exists $\boldsymbol{w}_k \neq 0$ s.t.

$$\langle \boldsymbol{w}_k, \boldsymbol{x} \rangle < \langle \boldsymbol{w}_k, \boldsymbol{x}_k \rangle, \quad \forall \boldsymbol{x} \in C$$

  By rescaling, we can assume $\|\boldsymbol{w}_k\| = 1$.
- By the Bolzano-Weierstrass Theorem, $\{\boldsymbol{w}_k\}$ has a convergent subsequence $\boldsymbol{w}_{k_i} \to \boldsymbol{w}$.
- Taking the limit $i \to \infty$ along the subsequence,

$$\langle \boldsymbol{w}_{k_i}, \boldsymbol{x} \rangle < \langle \boldsymbol{w}_{k_i}, \boldsymbol{x}_{k_i} \rangle, \quad \forall \boldsymbol{x} \in C \implies \langle \boldsymbol{w}, \boldsymbol{x} \rangle \leq \langle \boldsymbol{w}, \boldsymbol{x}_0 \rangle, \quad \forall \boldsymbol{x} \in C$$

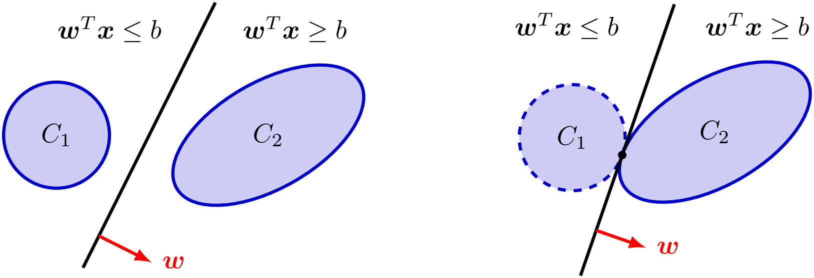
$$\|\boldsymbol{w}_{k_i}\| = 1 \implies \|\boldsymbol{w}\| = 1$$

Separating hyperplane

Separating hyperplane theorem. If C_1, C_2 are nonempty, convex sets in $\mathbb{R}^n$ with $C_1 \cap C_2 = \emptyset$, then C_1 and C_2 can be separated by a hyperplane, i.e. there exists $\boldsymbol{w} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $b \in \mathbb{R}$ s.t.

$$\boldsymbol{w}^T\boldsymbol{x} \leq b, \quad \forall \boldsymbol{x} \in C_1$$
$$\boldsymbol{w}^T\boldsymbol{x} \geq b, \quad \forall \boldsymbol{x} \in C_2$$

$P = \{\boldsymbol{x} : \boldsymbol{w}^T\boldsymbol{x} = b\}$ is called a separating hyperplane for C_1 and C_2.



Separating hyperplane (cont'd)

Lemma. If C_1, C_2 are two nonempty convex sets s.t. $C_1 \cap C_2 = \emptyset$, then $C = C_1 - C_2 = \{\boldsymbol{x}_1 - \boldsymbol{x}_2 : \boldsymbol{x}_1 \in C_1, \boldsymbol{x}_2 \in C_2\}$ is a nonempty convex set and $\mathbf{0} \notin C$.

Proof of theorem.

- It suffices to show there exists a $\boldsymbol{w} \neq \mathbf{0}$ s.t.

$$\boldsymbol{w}^T \boldsymbol{x}_1 \leq \boldsymbol{w}^T \boldsymbol{x}_2, \quad \forall \boldsymbol{x}_1 \in C_1, \boldsymbol{x}_2 \in C_2. \tag{$\dagger$}$$

  Then we can take $b = \sup_{\boldsymbol{x}_1 \in C_1} \boldsymbol{w}^T \boldsymbol{x}_1 \in \mathbb{R}$.
- Let $C = C_1 - C_2$. Then ($\dagger$) is equivalent to

$$\langle \boldsymbol{w}, \boldsymbol{x} \rangle \leq 0, \quad \forall \boldsymbol{x} \in C. \tag{$\ddagger$}$$

- Since $\mathbf{0} \notin C$, there are two cases. 0不在闭包内，取x0=0
  - If $\mathbf{0} \notin \overline{C}$, ($\ddagger$) will follow from the corollary on slide 32.
  - If $\mathbf{0} \in \partial C$, ($\ddagger$) will follow from the supporting hyperplane theorem.