CS257 Linear and Convex Optimization

Lecture 3

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

September 21, 2020

Recap: Optimality Conditions

Consider unconstrained optimization problem, i.e. $X = \mathbb{R}^n$.

Theorem (1st-order necessary condition). If f is continuously differentiable and $\boldsymbol{x}^*$ is a local minimum of f, then its gradient at $\boldsymbol{x}^*$ vanishes, i.e. $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$.

Theorem (2nd-order necessary condition). If f is twice continuously differentiable and $\boldsymbol{x}^*$ is a local minimum of f, then its Hessian matrix $\nabla^2 f(\boldsymbol{x}^*)$ is positive semidefinite, i.e.

$$\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} \geq 0, \quad \forall \boldsymbol{d} \in \mathbb{R}^n.$$

Theorem (2nd-order sufficient condition). Suppose f is twice continuously differentiable. If

1. $\nabla f(\boldsymbol{x}^*) = 0$
2. $\nabla^2 f(\boldsymbol{x}^*)$ is positive definite, i.e.

$$\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} > 0, \quad \forall \boldsymbol{d} \neq \mathbf{0}$$

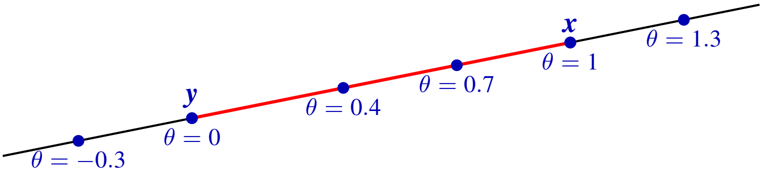
then $\boldsymbol{x}^*$ is a local minimum.

Contents

Lines, Line Segments and Rays

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

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



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

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

The line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$ consists of points

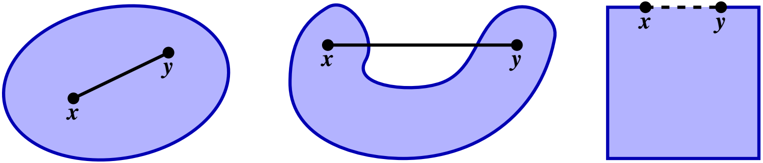
$$\theta\boldsymbol{x} + (1 - \theta)\boldsymbol{y}, \quad 0 \leq \theta \leq 1$$

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.

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



convex nonconvex nonconvex

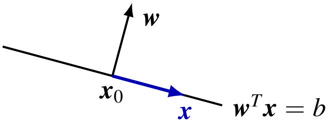
For $\theta \in [0,1]$, $\theta \boldsymbol{x} + \bar{\theta} \boldsymbol{y}$ is called a convex combination of $\boldsymbol{x}$ and $\boldsymbol{y}$. In a more symmetric form, a convex combination is

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

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 $\boldsymbol{w} \in \mathbb{R}^n$, $b \in \mathbb{R}$.

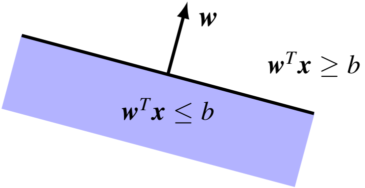


Proof. For $\boldsymbol{x}_1, \boldsymbol{x}_2 \in P$ and $\theta \in [0, 1]$,

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

Example: Halfspaces

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



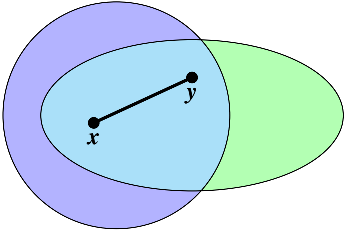
Note. $H = \{\boldsymbol{x} : f(\boldsymbol{x}) \leq b\}$ is the so-called sublevel set of $f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{x}$.
Note $\nabla f(\boldsymbol{x}) = \boldsymbol{w}$.

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

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

Intersection of Convex Sets

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.

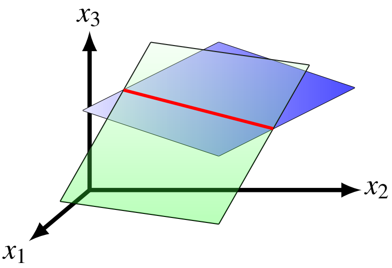
Can verify convexity by definition; here use the intersection property.

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

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

where

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



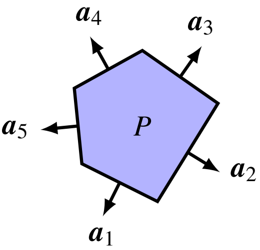
Note. An affine space S actually contains the line through any $\boldsymbol{x}, \boldsymbol{y} \in S$.

Example: Polyhedra

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

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

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

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

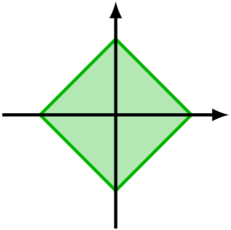
More generally, $\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}\}$ is a polyhedron.

Example: Polyhedra (cont'd)

The 1-norm unit ball is a polyhedron is convex,

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

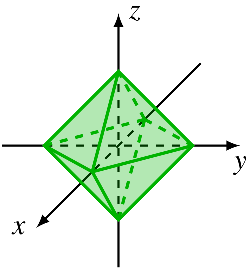
In 2d,



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

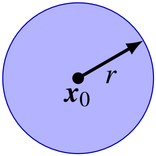
In 3d,

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



Example: Norm Balls

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



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

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

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

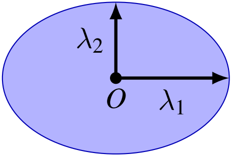
Note. Open balls are also convex.

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

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

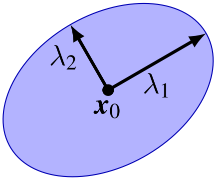
For $\boldsymbol{x}_i = \boldsymbol{\Lambda u}_i \in \mathcal{E}$, and $\theta \in [0, 1]$,

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

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 Transformations and Convex Sets

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,

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

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 $\boldsymbol{x}$,

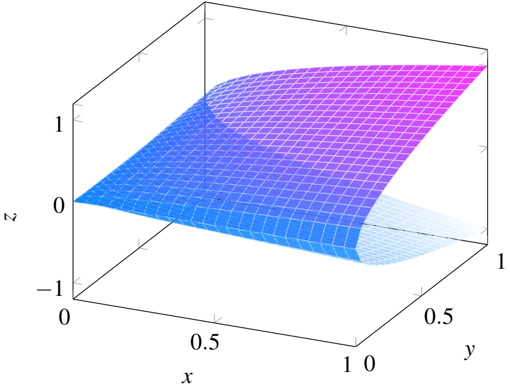
$$\boldsymbol{x}^T(\theta\boldsymbol{A} + \bar{\theta}\boldsymbol{B})\boldsymbol{x} = \theta(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}) + \bar{\theta}(\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}) \geq 0$$

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 $\partial \mathcal{S}_2^n = \{(x, y, z) : x \geq 0, y \geq 0, z^2 = xy\}$



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 $\sum_{i=1}^{m} \theta_i = 1$.

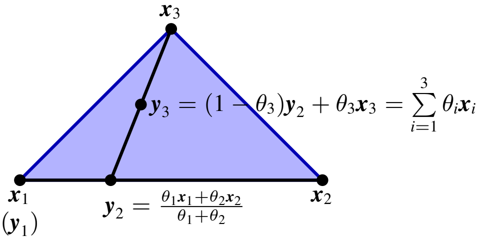
Theorem. If C is convex, and $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m \in C$, then any convex combination $\sum_{i=1}^{m} \theta_i \boldsymbol{x}_i \in C$.

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

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



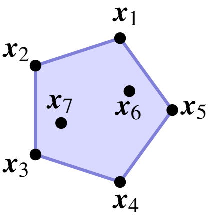
Convex Hull

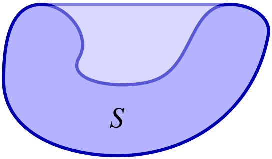
The convex hull of a set $S \subset \mathbb{R}^n$, denoted $\operatorname{conv} S$, is the smallest convex set containing S.

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

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

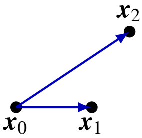
Note. Actually we need at most $n + 1$ points here, i.e. we can impose the condition $m \leq n + 1$ in the above representation.



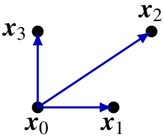


Affinely Independent Points

$m+1$ 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

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

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

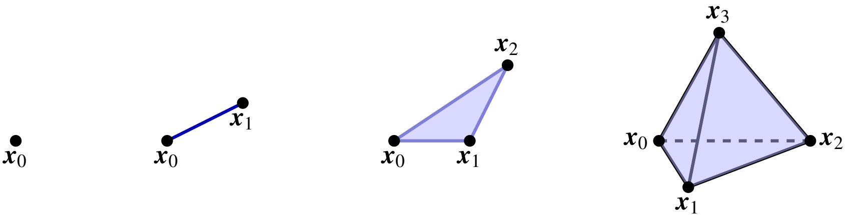
Simplexes

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

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

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

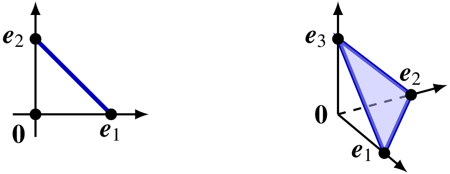
- 0-simplexes are points
- 1-simplexes are line segments
- 2-simplexes are triangles
- 3-simplexes are tetrahedra



Simplexes (cont'd)

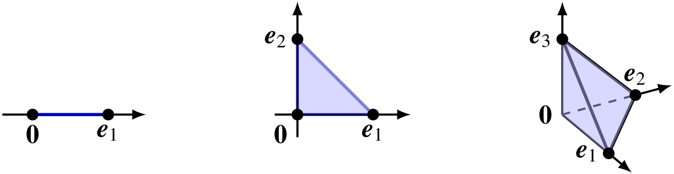
Example. The probability n-simplex is the n-simplex in $\mathbb{R}^{n+1}$ determined by the standard basis vectors $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n+1}$,

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



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

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



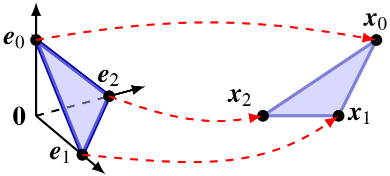
Simplexes (cont'd)

The m-simplex in $\mathbb{R}^n$ determined by affinely independent points $\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_m$ 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

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



Simplexes (cont'd)

Note

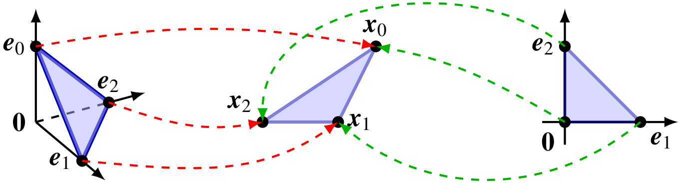
$$\boldsymbol{x} = \sum_{i=0}^{m} \theta_i \boldsymbol{x}_i = \boldsymbol{x}_0 + \sum_{i=1}^{m} \theta_i (\boldsymbol{x}_i - \boldsymbol{x}_0)$$

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

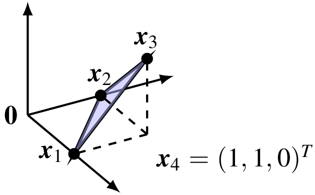


Simplexes (cont'd)

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,

$$\boldsymbol{x} = \boldsymbol{x}_1 + \theta_2(\boldsymbol{x}_2 - \boldsymbol{x}_1) + \theta_3(\boldsymbol{x}_3 - x_1) = (1 - \theta_2, \theta_2 + \theta_3, \theta_3)^T = \boldsymbol{x}_1 + \boldsymbol{B}\boldsymbol{\theta}',$$

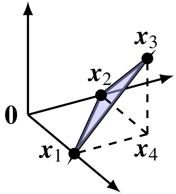
where

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

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

Simplex as Polyhedron

A simplex is a polyhedron, i.e. its points $\boldsymbol{x}$ can be specified by linear equalities and inequalities in $\boldsymbol{x}$.



Example. Recall for the 2-simplex determined by $\boldsymbol{x}_1 = (1,0,0)^T$, $\boldsymbol{x}_2 = (0,1,0)^T$ and $\boldsymbol{x}_3 = (1,1,1)^T$,

$$\boldsymbol{x} = \boldsymbol{X\theta} = (\theta_1 + \theta_3, \theta_2 + \theta_3, \theta_3)^T \implies \boldsymbol{\theta} = \boldsymbol{X}^{-1}\boldsymbol{x} = (x_1 - x_3, x_2 - x_3, x_3)^T$$

Since $\boldsymbol{\theta} \in \Delta_2$, $\boldsymbol{x}$ satisfies

$$\begin{cases} x_1 - x_3 \geq 0 \\ x_2 - x_3 \geq 0 \\ x_3 \geq 0 \\ x_1 + x_2 - x_3 = 1 \end{cases}$$

Note. This derivation does not work in general, as $\boldsymbol{X}$ may not even be a square matrix, let alone an invertible matrix.

Simplex as Polyhedron (cont'd)

Example (cont'd). Recall the other representation,

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{B}\boldsymbol{\theta}',\ \boldsymbol{B} = (\boldsymbol{x}_2 - \boldsymbol{x}_1, \boldsymbol{x}_3 - \boldsymbol{x}_1) = \begin{pmatrix} -1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix},\ \boldsymbol{\theta}' \in \Delta_2'$$

Note $\boldsymbol{B}$ has full column rank by affine independence of the $\boldsymbol{x}_i$'s, so it can be augmented to an invertible matrix $\tilde{\boldsymbol{B}}$,

$$(\boldsymbol{I}, \tilde{\boldsymbol{B}}) = (\boldsymbol{I}, \boldsymbol{B}, *) = \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & -1 & 0 & * \\ 0 & 1 & 0 & 1 & 1 & * \\ 0 & 0 & 1 & 0 & 1 & * \end{array}\right]$$

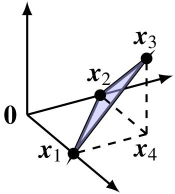
By elementary row operations,

$$(\boldsymbol{I}, \boldsymbol{B}, *) \to \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{I} & * \\ \boldsymbol{A}_2 & \boldsymbol{O} & * \end{pmatrix} = \left[\begin{array}{ccc:cc:c} -1 & 0 & 0 & 1 & 0 & * \\ 1 & 1 & 0 & 0 & 1 & * \\ \hdashline -1 & -1 & 1 & 0 & 0 & * \end{array}\right]$$

Simplex as Polyhedron (cont'd)

By those elementary row operations,

$$\boldsymbol{I}\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x}_1 + \boldsymbol{B}\boldsymbol{\theta}' \implies \begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{pmatrix} \boldsymbol{x} = \begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \end{pmatrix} \boldsymbol{x}_1 + \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{O} \end{pmatrix} \boldsymbol{\theta}'$$



or $\boldsymbol{A}_1\boldsymbol{x} = \boldsymbol{A}_1\boldsymbol{x}_1 + \boldsymbol{\theta}',\ \boldsymbol{A}_2\boldsymbol{x} = \boldsymbol{A}_2\boldsymbol{x}_1$

Since $\boldsymbol{\theta}' \in \Delta_2'$, the points $\boldsymbol{x} = (x_1, x_2, x_3)^T$ in the 2-simplex satisfy

$$\boldsymbol{A}_1\boldsymbol{x} \geq \boldsymbol{A}_1\boldsymbol{x}_1, \boldsymbol{1}^T\boldsymbol{A}_1(\boldsymbol{x} - \boldsymbol{x}_1) \leq 1, \boldsymbol{A}_2\boldsymbol{x} = \boldsymbol{A}_2\boldsymbol{x}_1$$

Using $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ from the previous slide,

$$\begin{cases} x_1 \leq 1 \\ x_1 + x_2 \geq 1 \\ x_2 \leq 1 \\ x_1 + x_2 - x_3 = 1 \end{cases}$$

Note. This method works for any simplex.