Convex Sets

Pontus Giselsson

Today's lecture

Motivation and context

- What is optimization?
- Why optimization?
- Convex vs nonconvex optimization
- Short course outlook

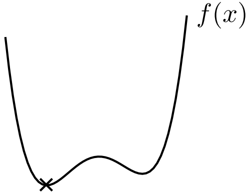
Today's subject: Convex sets

What is optimization?

- Find point $x \in \mathbb{R}^n$ that minimizes a function $f : \mathbb{R}^n \to \mathbb{R}$:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}}\, f(x)$$

- Example in $\mathbb{R}$:

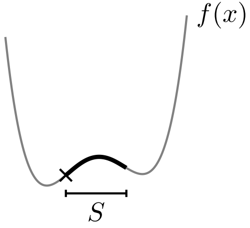


What is optimization?

- Can also require x to belong to a set $S \subset \mathbb{R}^n$:

$$\underset{x \in S}{\text{minimize}}\, f(x)$$

- Example in $\mathbb{R}$:



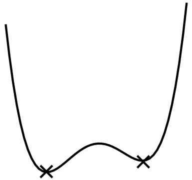
Why optimization?

- Many engineering problems can be modeled using optimization
  - **Supervised learning**
  - Optimal control
  - Signal reconstruction
  - Portfolio selection
  - Image classifiction
  - Circuit design
  - Estimation
  - ...
- Results in "optimal":
  - Model
  - Decision
  - Performance
  - Design
  - Estimate
  - ...

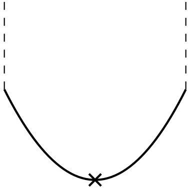
  w.r.t. optimization problem model
- Different question: How good is the model?

Convex vs nonconvex optimization

- Convex optimization if set and function are convex
- Otherwise nonconvex optimization problem
- Why convexity? Local minima are global minima
- Why go nonconvex? Richer modeling capabilities



nonconvex function



convex function

- If convex modeling enough, use it, otherwise try nonconvex

Short course outlook – Convex analysis part

- Set up to arrive at convex duality theory
- Fenchel duality (as opposed to (equivalent) Lagrange duality)
- Dual problem:
  - is companion problem to stated *primal* problem
  - can be easier to solve and than primal (SVM)
  - solution can (sometimes) be used to recover primal solution
  - is based on *conjugate functions* and optimizes over *subgradients*
  - in Fenchel duality assumes primal problem on *composite* form:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\, f(x) + g(x)$$

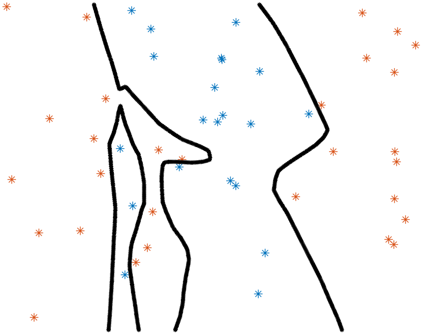
- Will see one algorithm for composite problem form

Short course outlook – Supervised learning part

- Some supervised learning methods from optimization perspective
- Classical supervised learning is based on convexity
  - Least squares, logistic regression, support vector machines (SVM)
  - SVM relies heavily on duality, state of the art until 10 years ago
  - "All local minima good" (if properly regularized)
  - Separates modeling from algorithm design
- Deep learning is based on nonconvex training problems
  - Algorithm can end up in local minima
  - Contemporary deep networks often overparameterized
    - Many global minima, some desired some not
    - Used algorithms (SGD variations) often find a "good" minimum
    - There is *implicit regularization* in SGD – will try to understand
  - No separation between modeling and algorithm

Different global minima generalize differently well

- Binary classification problem with blue and red class
- Black line is decision boundary of trained network with 0 loss
- Perfect fit to data and probably OK generalization



Different global minima generalize differently well

- Binary classification problem with blue and red class
- Decision boundary of another 0 loss network (same problem)
- Perfect fit to data and probably much worse generalization

- SGD has implicit regularization – often finds "good" minima
- Will try to understand why this is the case

Convex Sets

Outline

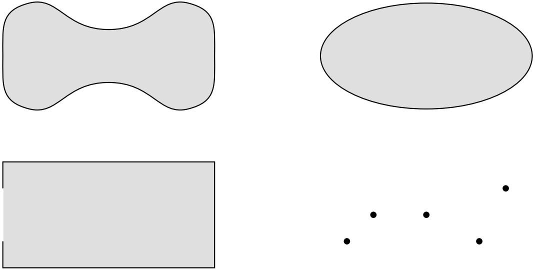
- **Definition and convex hull**
- Examples of convex sets
- Convexity preserving operations
- Concluding convexity – Examples
- Separating and supporting hyperplanes

Convex sets – Definition

- A set C is convex if for every $x, y \in C$ and $\theta \in [0, 1]$:

$$\theta x + (1 - \theta) y \in C$$

- "Every line segment that connect any two points in C is in C"



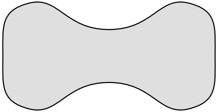
- Will assume that all sets are nonempty and closed

Convex sets – Definition

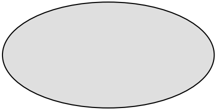
- A set C is convex if for every $x, y \in C$ and $\theta \in [0,1]$:

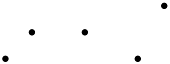
$$\theta x + (1-\theta) y \in C$$

- "Every line segment that connect any two points in C is in C"



Nonconvex





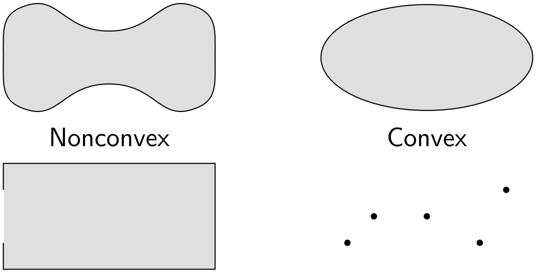
- Will assume that all sets are nonempty and closed

Convex sets – Definition

- A set C is convex if for every $x, y \in C$ and $\theta \in [0, 1]$:

$$\theta x + (1 - \theta) y \in C$$

- "Every line segment that connect any two points in C is in C"



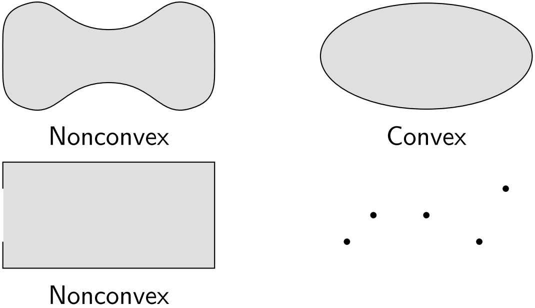
- Will assume that all sets are nonempty and closed

Convex sets – Definition

- A set C is convex if for every $x, y \in C$ and $\theta \in [0,1]$:

$$\theta x + (1-\theta)y \in C$$

- "Every line segment that connect any two points in C is in C"



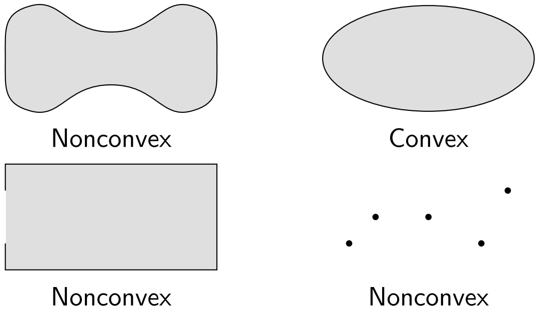
- Will assume that all sets are nonempty and closed

Convex sets – Definition

- A set C is convex if for every $x, y \in C$ and $\theta \in [0, 1]$:

$$\theta x + (1 - \theta) y \in C$$

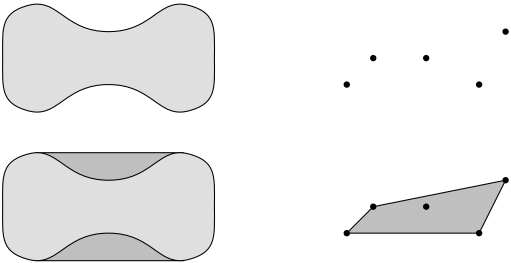
- "Every line segment that connect any two points in C is in C"



- Will assume that all sets are nonempty and closed

Convex combination and convex hull

Convex hull ($\mathrm{conv}S$) of S is smallest convex set that contains S:



Mathematical construction:

- Convex combinations of $x_1, \ldots, x_k$ are all points x of the form

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

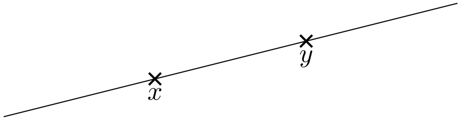
  where $\theta_1 + \ldots + \theta_k = 1$ and $\theta_i \geq 0$
- Convex hull: set of all convex combinations of points in S

Outline

- Definition and convex hull
- **Examples of convex sets**
- Convexity preserving operations
- Concluding convexity – Examples
- Separating and supporting hyperplanes

Affine sets

- Take any two points $x, y \in V$: V is affine if full line in V:



  Lines and planes are affine sets

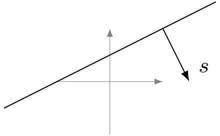
- Definition: A set V is affine if for every $x, y \in V$ and $\alpha \in \mathbb{R}$:

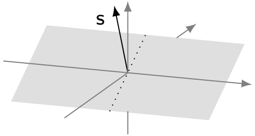
$$\alpha x + (1 - \alpha) y \in V \tag{1}$$

  hence convex this holds in particular for $\alpha \in [0, 1]$

Affine hyperplanes

- Affine hyperplanes in $\mathbb{R}^n$ are affine sets that cut $\mathbb{R}^n$ in two halves





- Dimension of affine hyperplane in $\mathbb{R}^n$ is $n-1$ (If $s \neq 0$)
- All affine sets in $\mathbb{R}^n$ of dimension $n-1$ are hyperplanes
- Mathematical definition:

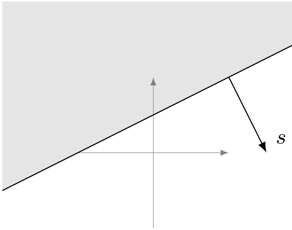
$$h_{s,r} := \{x \in \mathbb{R}^n : s^T x = r\}$$

where $s \in \mathbb{R}^n$ and $r \in \mathbb{R}$, i.e., defined by one *affine function*

- Vector s is called normal to hyperplane

Halfspaces

- A halfspace is one of the halves constructed by a hyperplane



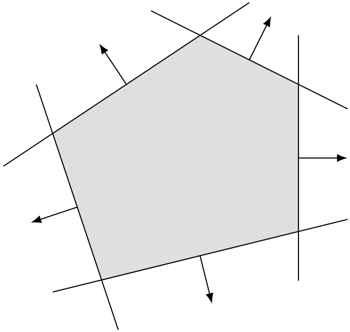
- Mathematical definition:

$$H_{r,s} = \{x \in \mathbb{R}^n : s^T x \leq r\}$$

- Halfspaces are convex, and vector s is called normal to halfspace

Polytopes

- A *polytope* is intersection of halfspaces and hyperplanes



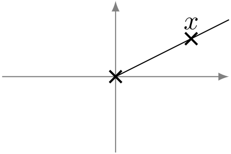
- Mathematical representation:

$$C = \{x \in \mathbb{R}^n : s_i^T x \leq r_i \text{ for } i \in \{1, \ldots, m\} \text{ and}$$
$$s_i^T x = r_i \text{ for } i \in \{m+1, ..., p\}\}$$

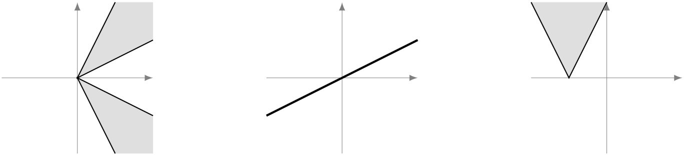
- Polytopes convex since intersection of convex sets

Cones

- A set K is a cone if for all $x \in K$ and $\alpha \geq 0$: $\alpha x \in K$
- If x is in cone K, so is entire ray from origin passing through x:



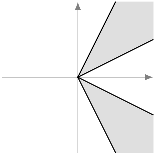
- Examples:

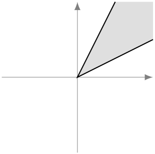


Cone　　Cone　　Not cone

Convex cones

- Cones can be convex or nonconvex:


Nonconvex cone


Convex cone

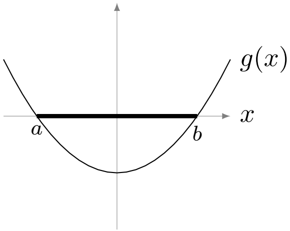
- Convex cone examples:
  - Linear subspaces $\{x \in \mathbb{R}^n : Ax = 0\}$ (but not affine subspaces)
  - Halfspaces based on linear (not affine) hyperplanes $\{x : s^T x \leq 0\}$
  - Positive semi-definite matrices
    $\{X \in \mathbb{R}^{n \times n} : X \text{ symmetric and } z^T X z \geq 0 \text{ for all } z \in \mathbb{R}^n\}$
  - Nonnegative orthant $\{x \in \mathbb{R}^n : x \geq 0\}$
  - Second order cone $\{(x, r) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_2 \leq r\}$

Sublevel sets

- Suppose that $g : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function
- The (0*th*) sublevel set of g is defined as

$$S := \{x \in \mathbb{R}^n : g(x) \leq 0\}$$

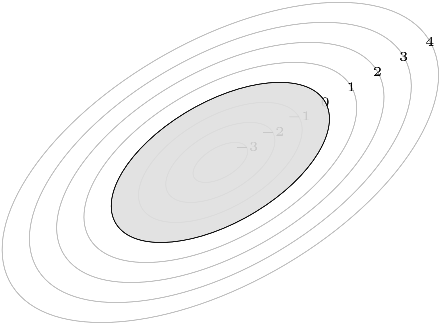
- Example: construction giving 1D interval $S = [a, b]$



- S is a convex set if g is a convex function
- S is not necessarily nonconvex although g is

Sublevel sets – Examples

- Levelset of convex quadratic function



$\{x \in \mathbb{R}^n : \frac{1}{2}x^T P x + q^T x + r \leq 0\}$, with P positive definite

- Norm balls $\{x \in \mathbb{R}^n : \|x\| - r \leq 0\}$
- Second-order cone $\{(x, r) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_2 - r \leq 0\}$
- Halfspaces $\{x \in \mathbb{R}^n : c^T x - r \leq 0\}$

Outline

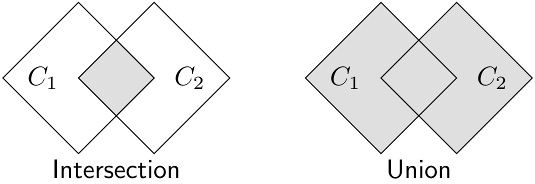
- Definition and convex hull
- Examples of convex sets
- **Convexity preserving operations**
- Concluding convexity – Examples
- Separating and supporting hyperplanes

Convexity preserving operations

- Intersection (but not union)
- Affine image and inverse affine image of a set

Intersection and union

- Intersection $C = C_1 \cap C_2$ means $x \in C$ if $x \in C_1$ **and** $x \in C_2$
  - If no x exists such that $x \in C_1$ and $x \in C_2$ then $C_1 \cap C_2 = \emptyset$
- Union $C = C_1 \cup C_2$ means $x \in C$ if $x \in C_1$ **or** $x \in C_2$



- Intersection of any number of, e.g., infinite, convex sets is convex
- Union of convex sets need not be convex

Image sets and inverse image sets

- Let $L(x) = Ax + b$ be an affine mapping defined by
  - matrix $A \in \mathbb{R}^{m \times n}$
  - vector $b \in \mathbb{R}^m$
- Let C be a convex set in $\mathbb{R}^n$ then the *image set of C under L*

  $$\{Ax + b : x \in C\}$$

  is convex
- Let D be a convex set in $\mathbb{R}^m$ then the *inverse image of D under L*

  $$\{x : Ax + b \in D\}$$

  is convex

Outline

- Definition and convex hull
- Examples of convex sets
- Convexity preserving operations
- **Concluding convexity – Examples**
- Separating and supporting hyperplanes

Ways to conclude convexity

- Use convexity definition
- Show that set is sublevel set of a convex function
- Show that set constructed by convexity preserving operations

Example – Nonnegative orthant

- Nonnegative orthant is set $C = \{x \in \mathbb{R}^n : x \geq 0\}$
- Prove convexity from definition:
  - Let $x \geq 0$ and $y \geq 0$ be arbitrary points in C
  - For all $\theta \in [0, 1]$:

$$\theta x \geq 0 \qquad \text{and} \qquad (1-\theta)y \geq 0$$

  - All convex combinations therefore also satisfy

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

i.e., they belongs to C and the set is convex

Example – Positive semidefinite cone

- The positive semidefinite (PSD) cone is

$$\{X \in \mathbb{R}^{n\times n} : X \text{ symmetric}\} \bigcap \{X \in \mathbb{R}^{n\times n} : z^T X z \geq 0 \text{ for all } z \in \mathbb{R}^n\}$$

- This can be written as the following intersection over all $z \in \mathbb{R}^n$

$$\{X \in \mathbb{R}^{n\times n} : X \text{ symmetric}\} \bigcap \bigcap_{z \in \mathbb{R}^n} \{X \in \mathbb{R}^{n\times n} : z^T X z \geq 0\}$$

which, by noting that $z^T X z = \mathrm{tr}(z^T X z) = \mathrm{tr}(zz^T X)$, is equal to

$$\{X \in \mathbb{R}^{n\times n} : X \text{ symmetric}\} \bigcap \bigcap_{z \in \mathbb{R}^n} \{X \in \mathbb{R}^{n\times n} : \mathrm{tr}(zz^T X) \geq 0\}$$

where $\mathrm{tr}(zz^T X) \geq 0$ is a halfspace in $\mathbb{R}^{n\times n}$ (except when $z = 0$)

- The PSD cone is convex since it is intersection of
  - symmetry set, which is a finite set of (convex) linear equalities
  - an infinite number of (convex) halfspaces in $\mathbb{R}^{n\times n}$
- Notation: If X belongs to the PSD cone, we write $X \succeq 0$

Example – Linear matrix inequality

- Let us consider a linear matrix inequality (LMI) of the form

$$\{x \in \mathbb{R}^k : A + \sum_{i=1}^{k} x_i B_i \succeq 0\}$$

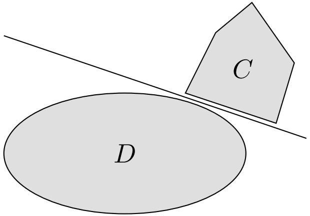
  where A and B_i are fixed matrices in $\mathbb{R}^{n \times n}$
- Convex since inverse image of PSD cone under affine mapping

Outline

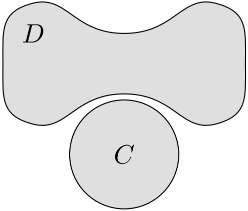
- Definition and convex hull
- Examples of convex sets
- Convexity preserving operations
- Concluding convexity – Examples
- **Separating and supporting hyperplanes**

Separating hyperplane theorem

- Suppose that $C, D \subseteq \mathbb{R}^n$ are two non-intersecting convex sets
- Then there exists hyperplane with C and D in opposite halves



Example



Counter-example
D nonconvex

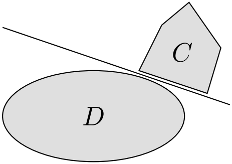
- Mathematical formulation: There exists $s \neq 0$ and r such that

$$s^T x \leq r \qquad \text{for all } x \in C$$
$$s^T x \geq r \qquad \text{for all } x \in D$$

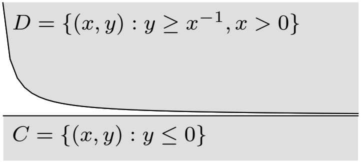
- The hyperplane $\{x : s^T x = r\}$ is called *separating hyperplane*

A strictly separating hyperplane theorem

- Suppose that $C, D \subseteq \mathbb{R}^n$ are non-intersecting closed and convex sets and that one of them is compact (closed and bounded)
- Then there exists hyperplane with strict separation



Example



Counter example
C, D not compact

- Mathematical formulation: There exists $s \neq 0$ and r such that

$$s^T x < r \qquad \text{for all } x \in C$$

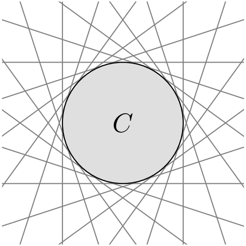
$$s^T x > r \qquad \text{for all } x \in D$$

Consequence – C is intersection of halfspaces

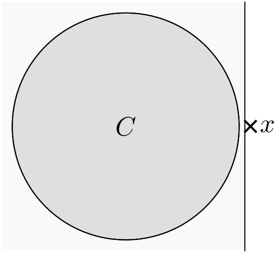
$\boxed{\text{a closed convex set } C \text{ is the intersection of all halfspaces that contain it}}$

proof:

- let H be the intersection of all halfspaces containing C
- $\Rightarrow$: obviously $x \in C \Rightarrow x \in H$
- $\Leftarrow$: assume $x \notin C$, since C closed and convex and $\{x\}$ compact singleton, there exists a strictly separating hyperplane, i.e., $x \notin H$:



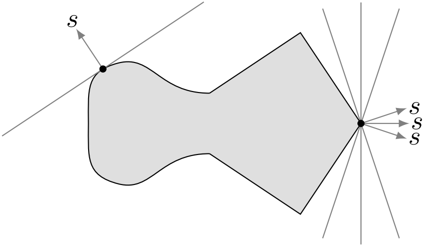
$\Rightarrow$



$\Leftarrow$

Supporting hyperplanes

- Supporting hyperplanes touch set and have full set on one side:



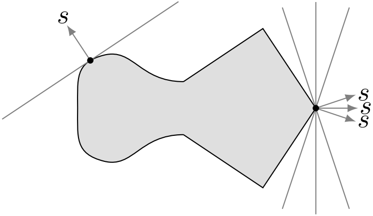
- We call the halfspace that contains the set *supporting halfspace*
- s is called *normal vector* to C at x
- Definition: Hyperplane $\{y : s^T y = r\}$ supports C at $x \in \text{bd } C$ if

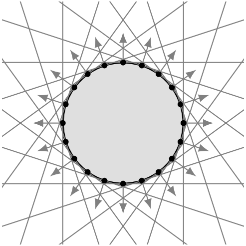
$$s^T x = r \qquad \text{and} \qquad s^T y \leq r \text{ for all } y \in C$$

Supporting hyperplane theorem

Let C be a nonempty convex set and let $x \in \mathrm{bd}(C)$. Then there exists a supporting hyperplane to C at x.

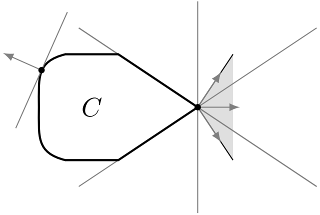
- Does not exist for all point on boundary for nonconvex sets
- Many supporting hyperplanes exist for points of nonsmoothness





Normal cone operator

- Normal cone to C at $x \in \mathrm{bd}(C)$ is set of normals at x



- Normal cone operator N_C to C takes point input and returns set:
  - $x \in \mathrm{bd}(C) \cap C$: set of normal vectors to supporting halfspaces
  - $x \in \mathrm{int}(C)$: returns zero set $\{0\}$
  - $x \notin C$: returns emptyset $\emptyset$
- Mathematical definition: The normal cone operator to a set C is

$$N_C(x) = \begin{cases} \{s : s^T(y - x) \leq 0 \text{ for all } y \in C\} & \text{if } x \in C \\ \emptyset & \text{else} \end{cases}$$

  i.e., vectors that form obtuse angle between s and all $y - x$, $y \in C$
- For all $x \in C$: the N_C outputs a set that contains 0