Lecture 2:
Convex optimization

Optimization for data sciences

Rémy Sun
remy.sun@inria.fr

Course organization

What can we optimize?

- Reduce the complexity/overhead of a problem
  - E.g. Network quantization
  - E.g. Computational optimization

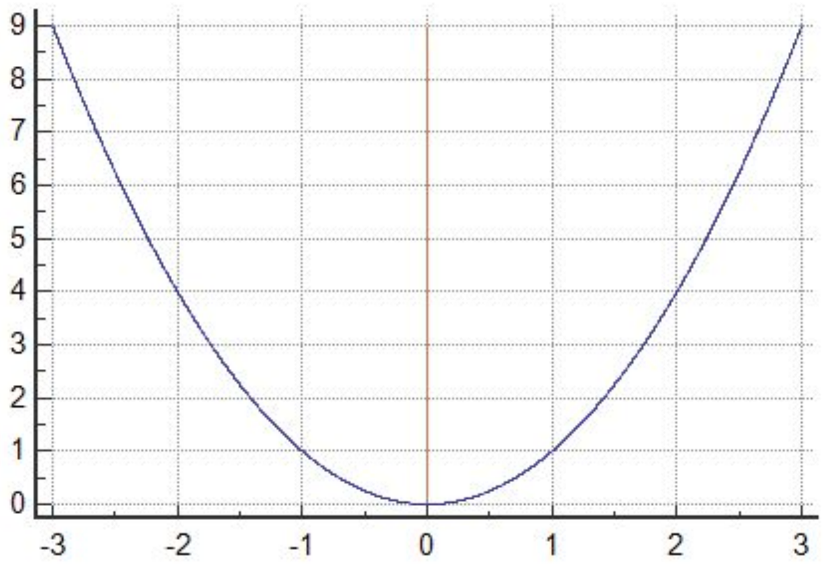
- **Find the best solution to a problem**
  - **Numerical optimization**
  - **Evaluate solutions according to a criterion**
  - **Look at solutions from some given space of possible solutions to consider**

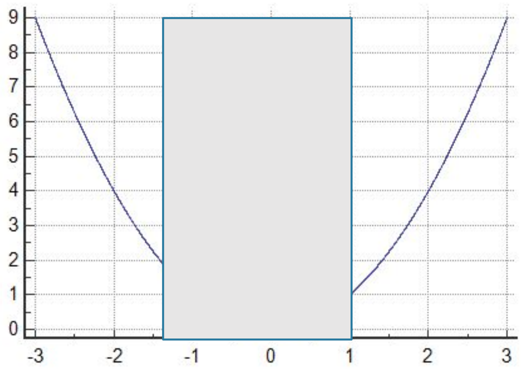
Defining an optimization problem

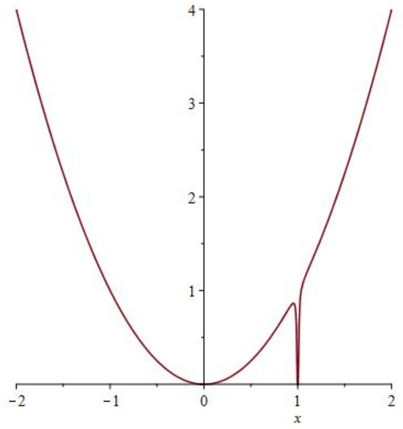
- Minimize a quantity $f_0(x)$
  - Under inequality and equality constraints
  - Constraints define a domain D
  - Could have no constraint except x in D

$$
\begin{array}{ll}
\text{minimize} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& g_i(x) = 0, \quad i = 1, \ldots, p
\end{array}
$$

Can you formalize these problems?







Course organization

- Introduction to optimization
  - A few problems of interest
  - Quick mathematical refresher

- **Convex problems (Following Stephen Boyd)**
  - **Quick refresher on last week**
  - **Convex sets**
  - **Convex functions**
  - **Convex problems**
  - **Simplex algorithm for Linear Programming**

Course organization

- Duality (for convex problems)
- Newton's Descent and Barrier methods for convex case
- (First order) descent methods for the general case
- Backpropagation
- Some more properties on stochastic gradient descent

Evaluation

- Reports on lab sessions
  - Labs on jupyter notebooks
    - Not every session
  - Explain the code done in the session
  - Summarize what is done in the practical

- Written Exam
  - Theoretical questions
  - We will do exercises in class

Refresher on last week

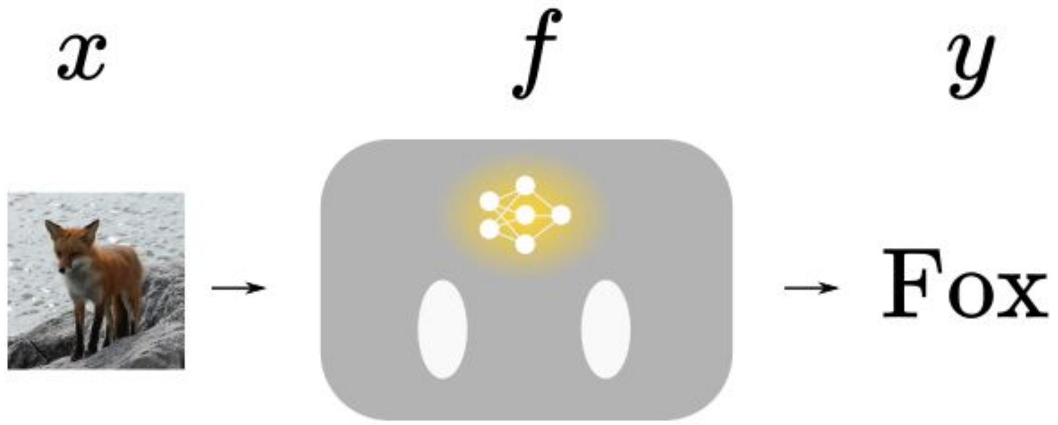
Optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & g_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- $x \in \mathbf{R}^n$ is (vector) variable to be chosen (n scalar variables $x_1, \ldots, x_n$)
- f_0 is the **objective function**, to be minimized
- $f_1, \ldots, f_m$ are the **inequality constraint functions**
- $g_1, \ldots, g_p$ are the **equality constraint functions**

- variations: maximize objective, multiple objectives, ...

Problem statement: Ideal case



- Find (robot) f that classifies images well
  - Often based on neural networks

$$\forall (x, y) \in \mathcal{D}, f(x) = y$$

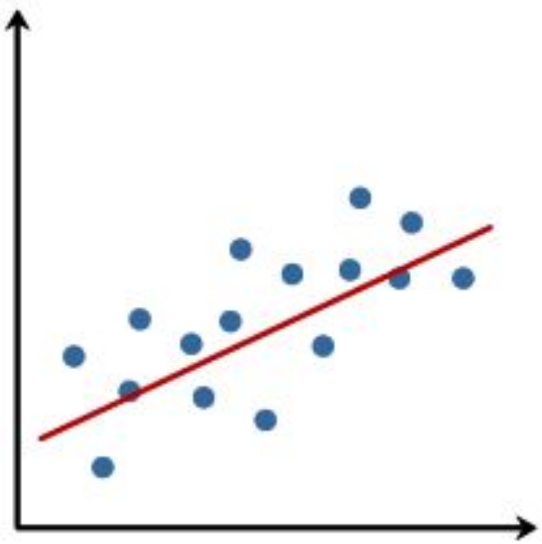
Minimizing Risk

- Problem: we do not know $\mathcal{D}$!
  - Solved problem otherwise…
  - Evaluating the risk requires this distribution

- Solution: Use a dataset D of (x,y) sampled from $\mathcal{D}$
  - ***Empirical Risk Minimization***
  - If the (x,y) are i.i.d drawn from $\mathcal{D}$ can be expressed as a mean over the dataset

$$min_\theta \hat{\mathcal{R}_\theta} = \frac{1}{N} \sum_{i=0,\ldots,N-1} l(f_\theta(x_i), y_i)$$

Linear regression

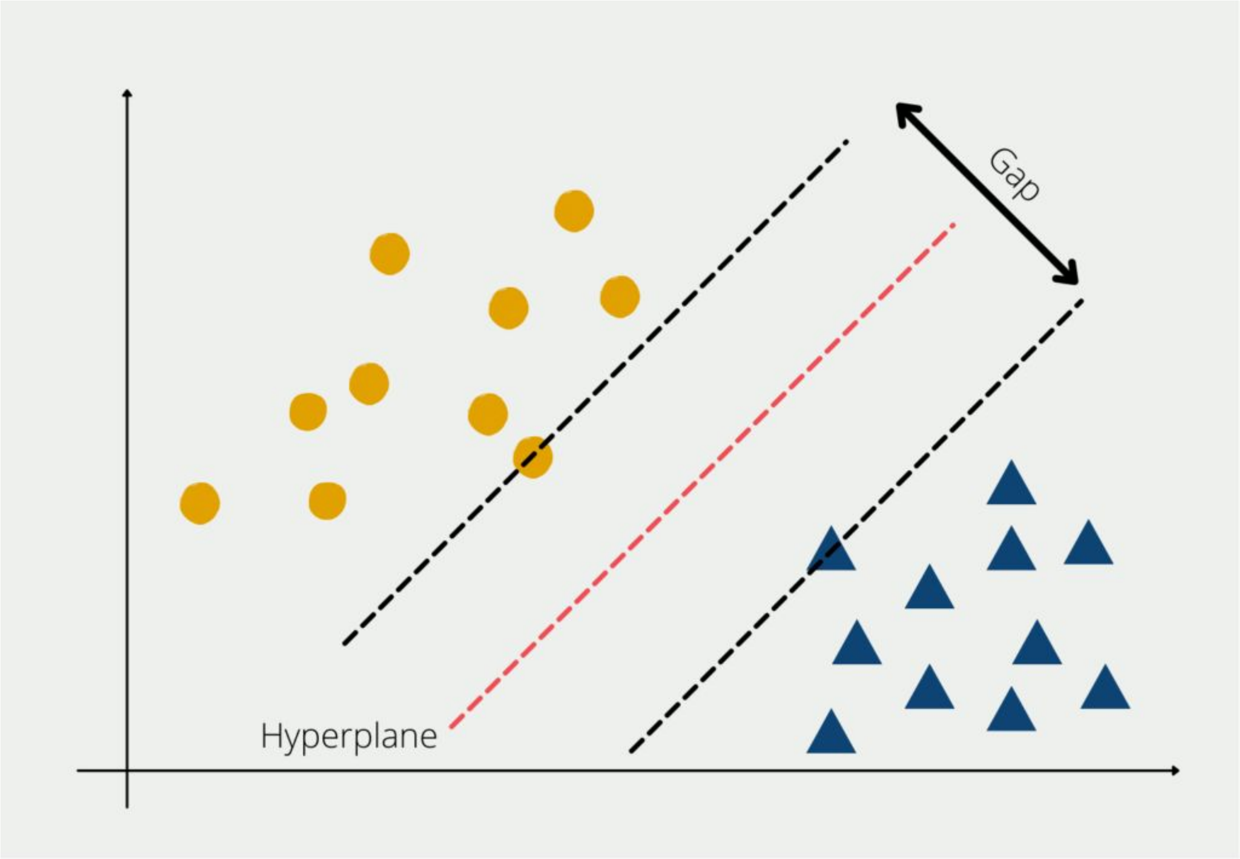


- Linearly correlated data
  - Input x (e.g. Voltage)
  - Output y (e.g. Intensity)

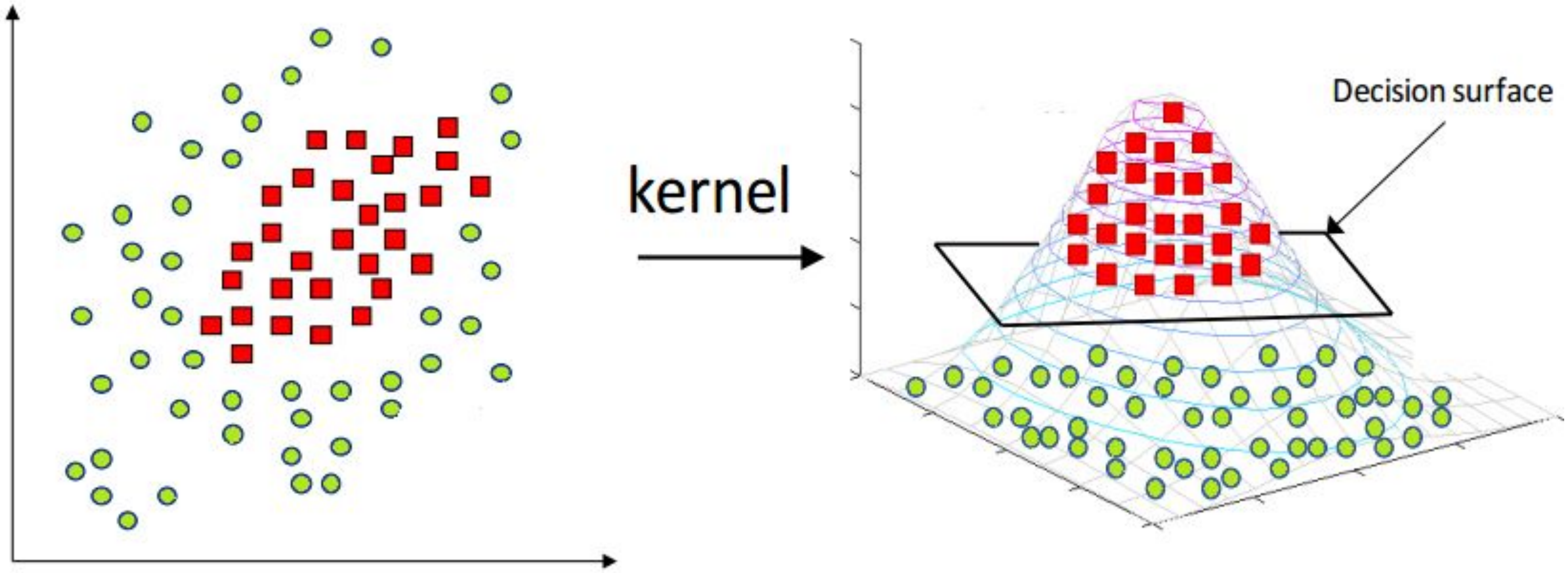
- Simple family of linear functions
  - Find linear coefficient

$$f_\theta(x) = \theta x$$

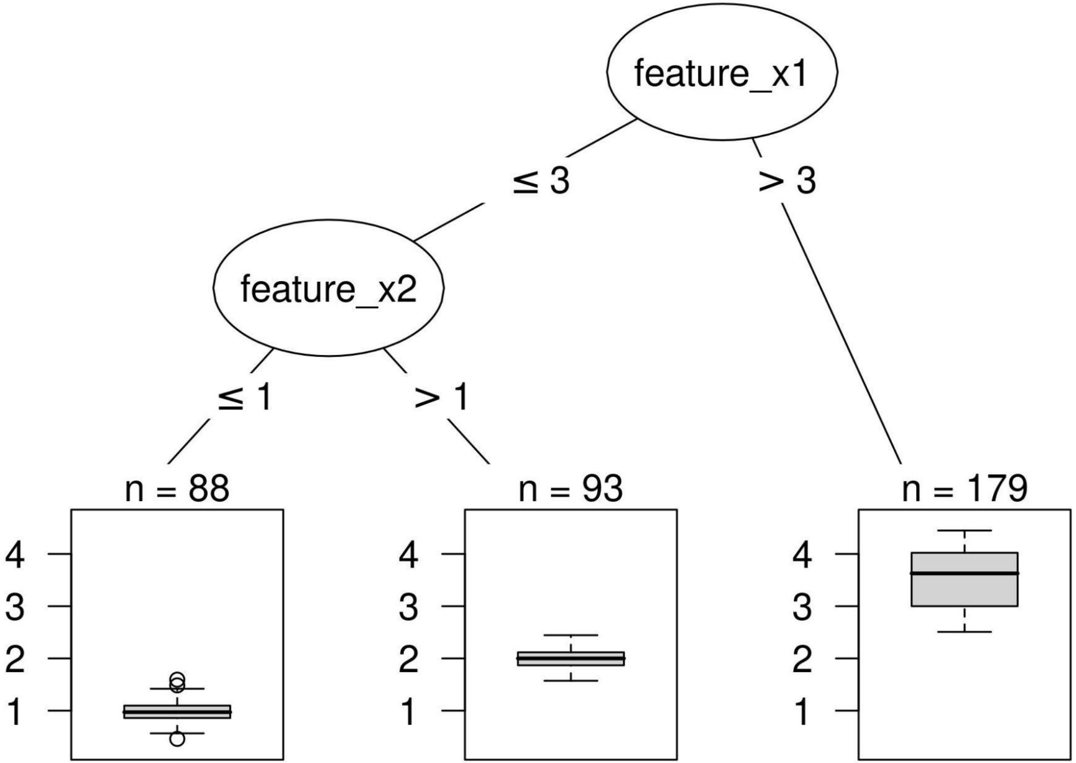
Separating hyperplane (SVM)



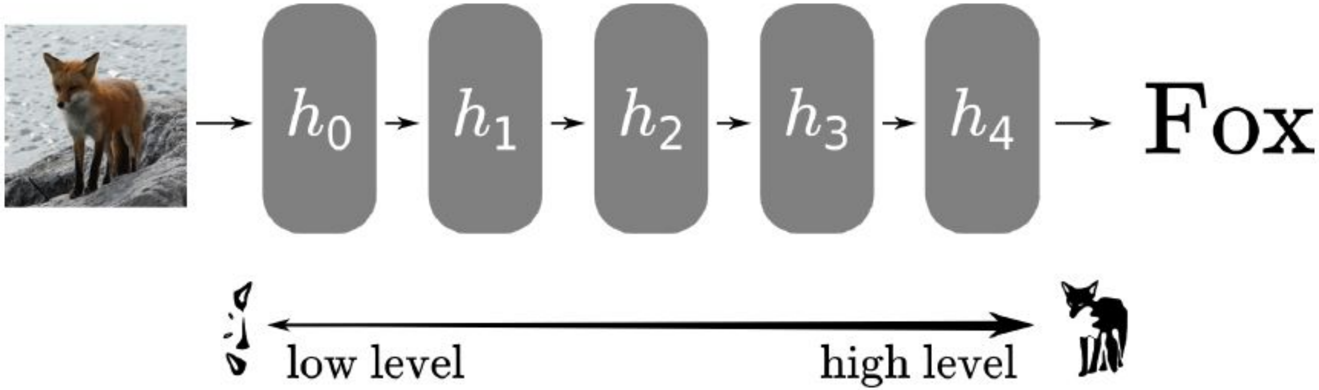
Separating hyperplane (SVM with kernel)



Decision trees



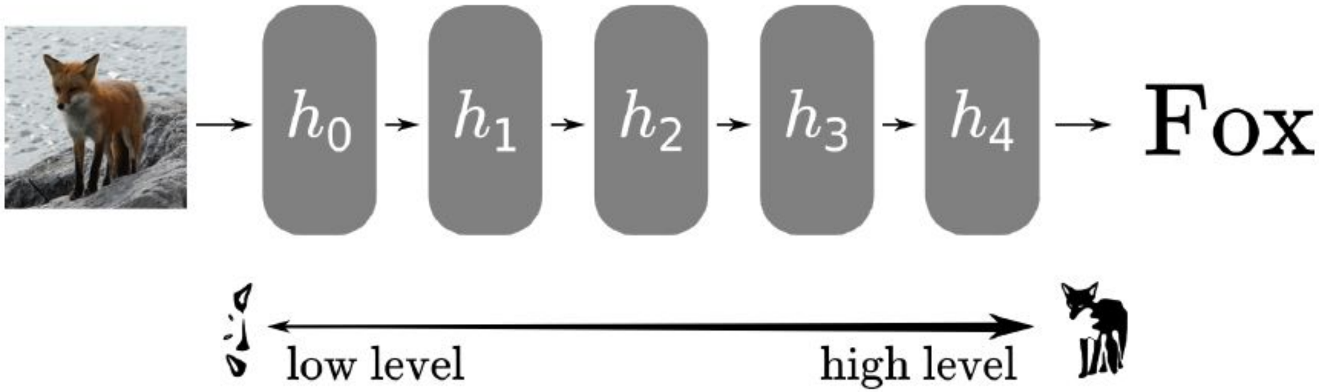
Neural network functions

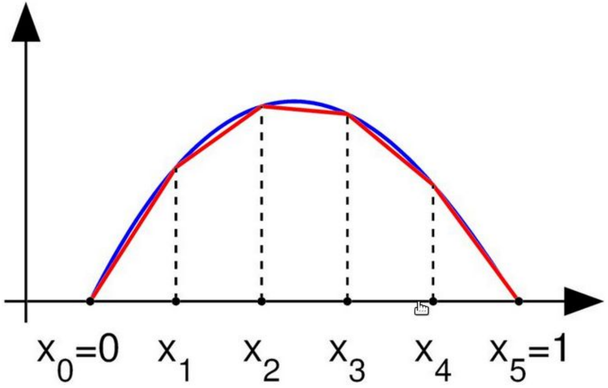


- Neural networks are sequences of simple functions

$$f_\theta = h_\theta^0 \circ h_\theta^1 \circ \cdots \circ h_\theta^{L-1}$$

Interlude: Formalization of dense networks





- Highly expressive
  - Can fit many types of distributions

0. Convex problem?

Convex problem

convex optimization problem:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

- variable $x \in \mathbf{R}^n$
- equality constraints are linear
- $f_0, \ldots, f_m$ are **convex**: for $\theta \in [0, 1]$,

$$f_i(\theta x + (1 - \theta)y) \leq \theta f_i(x) + (1 - \theta) f_i(y)$$

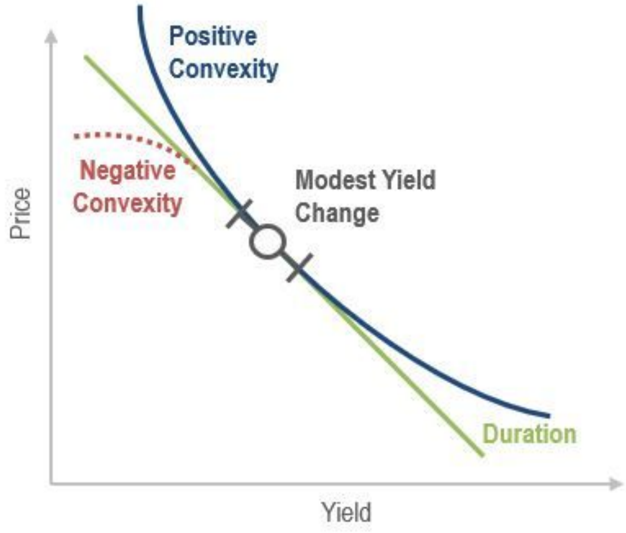
i.e., f_i have nonnegative (upward) curvature

Easy problem

- classical view:
  - linear (zero curvature) is easy
  - nonlinear (nonzero curvature) is hard

- the classical view is **wrong**

- the correct view:
  - convex (nonnegative curvature) is easy
  - nonconvex (negative curvature) is hard



Easy to solve!

- many different algorithms (that run on many platforms)
  - interior-point methods for up to 10000s of variables
  - first-order methods for larger problems
  - do not require initial point, babysitting, or tuning
- can develop and deploy quickly using modeling languages such as CVXPY
- solvers are reliable, so can be embedded
- code generation yields real-time solvers that execute in milliseconds (*e.g.*, on Falcon 9 and Heavy for landing)

Brief history

UNIVERSITÉ CÔTE D'AZUR

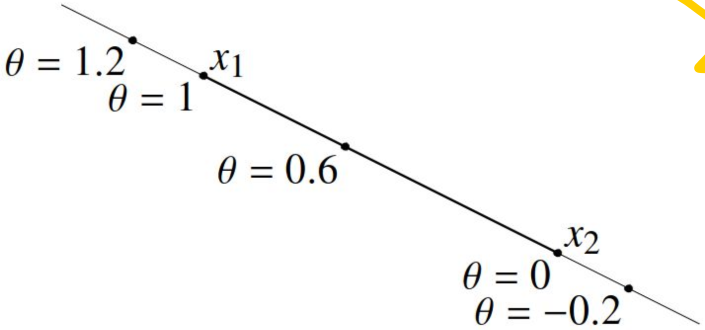
- **theory (convex analysis)**: 1900–1970
- **algorithms**
  - 1947: simplex algorithm for linear programming (Dantzig)
  - 1960s: early interior-point methods (Fiacco & McCormick, Dikin, . . .)
  - 1970s: ellipsoid method and other subgradient methods
  - 1980s & 90s: interior-point methods (Karmarkar, Nesterov & Nemirovski)
  - since 2000s: many methods for large-scale convex optimization
- **applications**
  - before 1990: mostly in operations research, a few in engineering
  - since 1990: many applications in engineering (control, signal processing, communications, circuit design, . . .)
  - since 2000s: machine learning and statistics, finance

1. Convex sets

Affine sets

line through x_1, x_2: all points of form $x = \theta x_1 + (1 - \theta) x_2$, with $\theta \in \mathbf{R}$

$\sum \theta_i = 1$



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

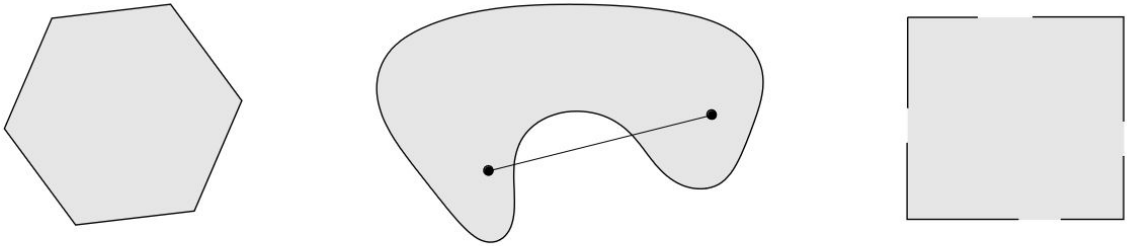
UNIVERSITÉ CÔTE D'AZUR

line segment between x_1 and x_2: all points of form $x = \theta x_1 + (1 - \theta)x_2$, with $0 \leq \theta \leq 1$

convex set: contains line segment between any two points in the set

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \Longrightarrow \quad \theta x_1 + (1 - \theta)x_2 \in C$$

examples (one convex, two nonconvex sets)



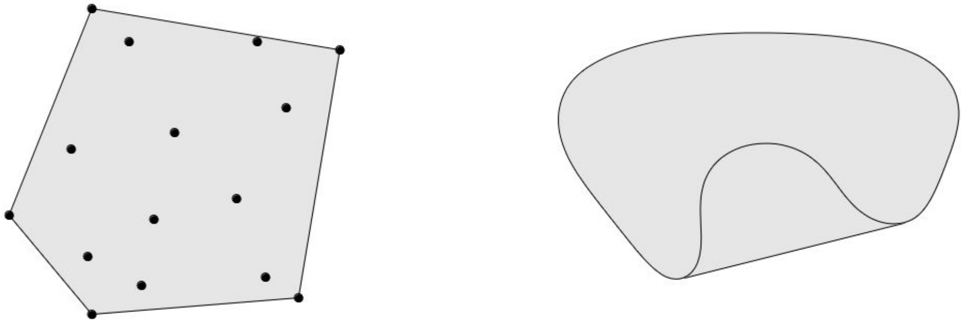
Convex combination and hull

convex combination of $x_1, \ldots, x_k$: any point x of the form

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

with $\theta_1 + \cdots + \theta_k = 1$, $\theta_i \geq 0$

convex hull $\mathbf{conv}\, S$: set of all convex combinations of points in S

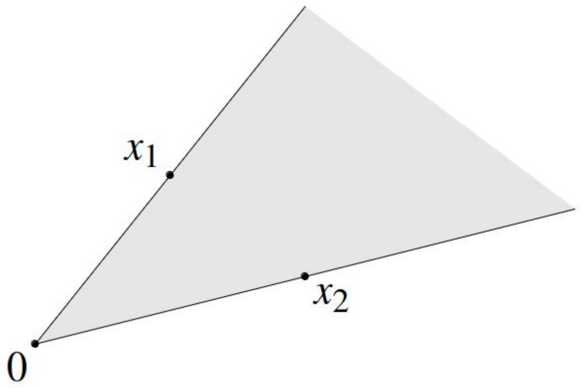


Convex cone

conic (nonnegative) combination of x_1 and x_2: any point of the form

$$\boxed{x = \theta_1 x_1 + \theta_2 x_2}$$

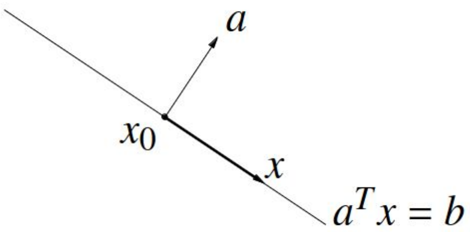
with $\theta_1 \geq 0$, $\theta_2 \geq 0$



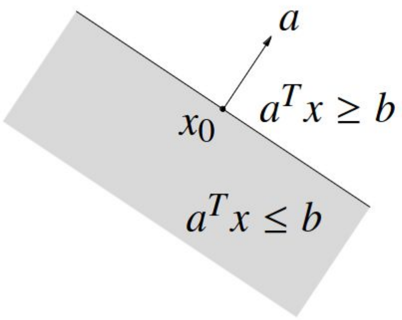
convex cone: set that contains all conic combinations of points in the set

Hyperplanes and halfspaces

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



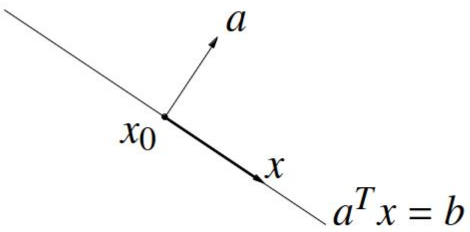
halfspace: set of the form $\{x \mid a^T x \leq b\}$, with $a \neq 0$



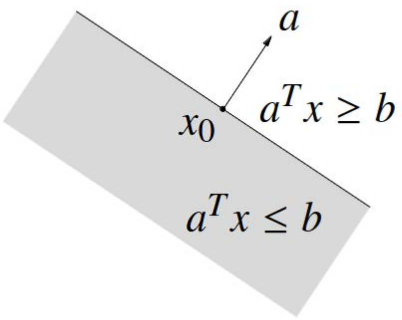
- a is the normal vector
- hyperplanes are affine and convex; halfspaces are convex

Hyperplanes and halfspaces

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



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



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

Hyperplanes and halfspaces

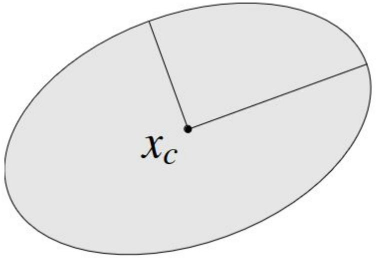
(Euclidean) ball with center x_c and radius r:

$$\boxed{B(x_c, r) = \{x \mid \|x - x_c\|_2 \le r\}} = \{x_c + ru \mid \|u\|_2 \le 1\}$$

ellipsoid: set of the form

$$\{x \mid (x - x_c)^T P^{-1} (x - x_c) \le 1\}$$

with $P \in \mathbf{S}^n_{++}$ (*i.e.*, P symmetric positive definite)



another representation: $\{x_c + Au \mid \|u\|_2 \le 1\}$ with A square and nonsingular

UNIVERSITÉ CÔTE D'AZUR

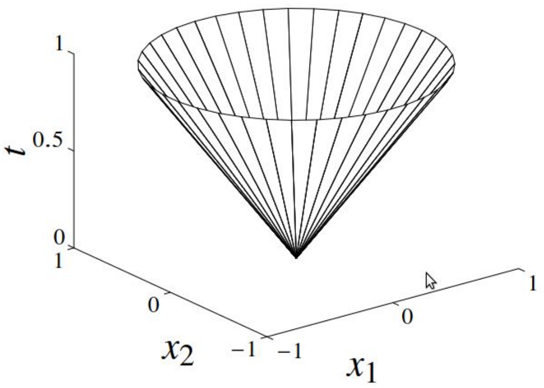
Norm balls and norm cones

- **norm:** a function $\|\cdot\|$ that satisfies
  - $\|x\| \geq 0$; $\|x\| = 0$ if and only if $x = 0$
  - $\|tx\| = |t| \, \|x\|$ for $t \in \mathbf{R}$
  - $\|x + y\| \leq \|x\| + \|y\|$
- notation: $\|\cdot\|$ is general (unspecified) norm; $\|\cdot\|_{\text{symb}}$ is particular norm
- **norm ball** with center x_c and radius r: $\{x \mid \|x - x_c\| \leq r\}$
- **norm cone:** $\{(x, t) \mid \|x\| \leq t\}$
- norm balls and cones are convex

Euclidean norm cone

$$\{(x, t) \mid \|x\|_2 \leq t\} \subset \mathbf{R}^{n+1}$$

is called **second-order cone**

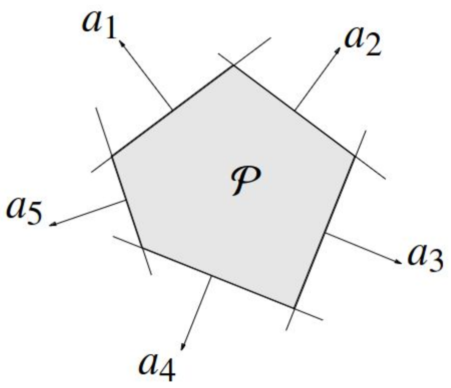


Polyhedron

- **polyhedron** is solution set of finitely many linear inequalities and equalities

$$\{x \mid Ax \preceq b, \; Cx = d\}$$

  ($A \in \mathbf{R}^{m\times n}$, $C \in \mathbf{R}^{p\times n}$, $\preceq$ is componentwise inequality)
- intersection of finite number of halfspaces and hyperplanes
- example with no equality constraints; a_i^T are rows of A



Positive semidefinite cone

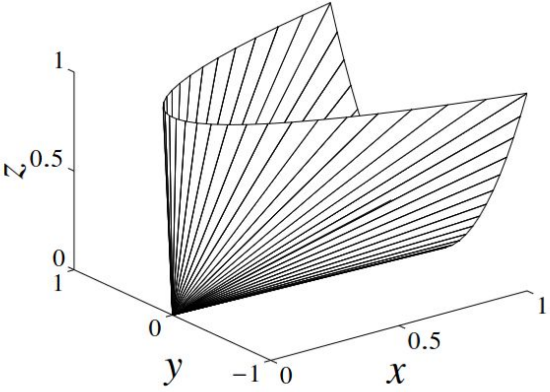
notation:

- $\mathbf{S}^n$ is set of symmetric $n \times n$ matrices
- $\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}$: positive semidefinite (symmetric) $n \times n$ matrices

$$X \in \mathbf{S}^n_+ \iff z^T X z \geq 0 \text{ for all } z$$

- $\mathbf{S}^n_+$ is a convex cone, the **positive semidefinite cone**
- $\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}$: positive definite (symmetric) $n \times n$ matrices

example: $\begin{bmatrix} x & y \\ y & z \end{bmatrix} \in \mathbf{S}^2_+$



Showing a set is convex

methods for establishing convexity of a set C

1. apply definition: show $x_1, x_2 \in C, 0 \le \theta \le 1 \implies \theta x_1 + (1-\theta)x_2 \in C$
   - recommended only for **very simple** sets

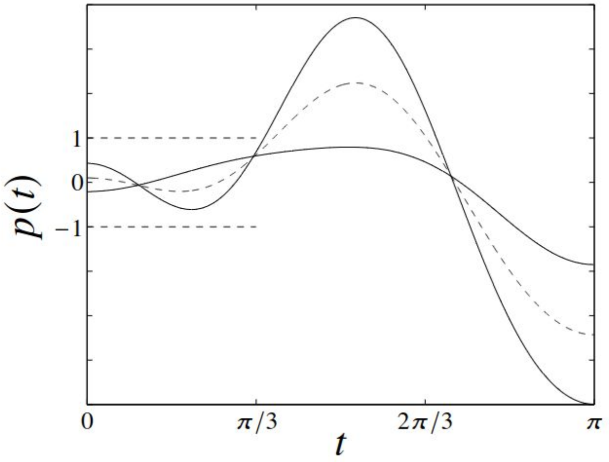
2. use convex functions (next lecture)

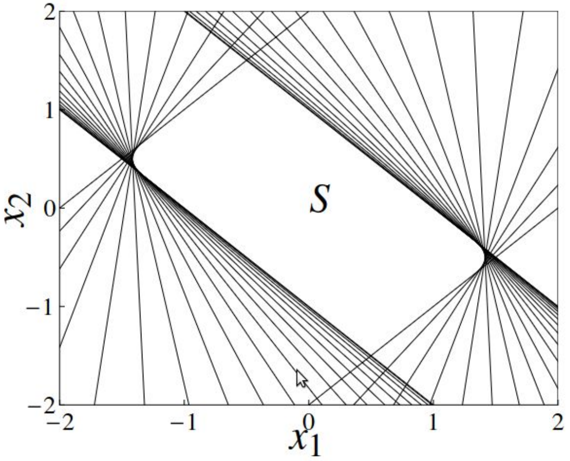
3. show that C is obtained from simple convex sets (hyperplanes, halfspaces, norm balls, ...) by operations that preserve convexity
   - intersection
   - affine mapping
   - perspective mapping
   - linear-fractional mapping

you'll mostly use methods 2 and 3

Showing a set is convex

- the intersection of (any number of) convex sets is convex
- **example:**
  - $S = \{x \in \mathbf{R}^m \mid |p(t)| \leq 1 \text{ for } |t| \leq \pi/3\}$, with $p(t) = x_1 \cos t + \cdots + x_m \cos mt$
  - write $S = \bigcap_{|t| \leq \pi/3} \{x \mid |p(t)| \leq 1\}$, *i.e.*, an intersection of (convex) slabs
- picture for $m = 2$:





Showing a set is convex

- suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ is affine, *i.e.*, $f(x) = Ax + b$ with $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$

- the **image** of a convex set under f is convex

$$S \subseteq \mathbf{R}^n \text{ convex} \quad \Longrightarrow \quad f(S) = \{f(x) \mid x \in S\} \text{ convex}$$

- the **inverse image** $f^{-1}(C)$ of a convex set under f is convex

$$C \subseteq \mathbf{R}^m \text{ convex} \quad \Longrightarrow \quad f^{-1}(C) = \{x \in \mathbf{R}^n \mid f(x) \in C\} \text{ convex}$$

Showing a set is convex

- **perspective function** $P : \mathbf{R}^{n+1} \rightarrow \mathbf{R}^n$:

$$P(x,t) = x/t, \qquad \textbf{dom}\, P = \{(x,t) \mid t > 0\}$$

- images and inverse images of convex sets under perspective are convex

- **linear-fractional function** $f : \mathbf{R}^n \rightarrow \mathbf{R}^m$:

$$f(x) = \frac{Ax+b}{c^T x + d}, \qquad \textbf{dom}\, f = \{x \mid c^T x + d > 0\}$$

- images and inverse images of convex sets under linear-fractional functions are convex

Proper cones

a convex cone $K \subseteq \mathbf{R}^n$ is a **proper cone** if

- K is closed (contains its boundary)
- K is solid (has nonempty interior)
- K is pointed (contains no line)

examples

- nonnegative orthant $K = \mathbf{R}^n_+ = \{x \in \mathbf{R}^n \mid x_i \geq 0,\ i = 1, \ldots, n\}$
- positive semidefinite cone $K = \mathbf{S}^n_+$
- nonnegative polynomials on $[0, 1]$:

$$K = \{x \in \mathbf{R}^n \mid x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1} \geq 0 \text{ for } t \in [0, 1]\}$$

Generalized inequality

- (nonstrict and strict) **generalized inequality** defined by a proper cone K:

$$x \preceq_K y \iff y - x \in K, \qquad x \prec_K y \iff y - x \in \textbf{int}\, K$$

- **examples**
  - componentwise inequality ($K = \mathbf{R}^n_+$): $x \preceq_{\mathbf{R}^n_+} y \iff x_i \leq y_i, \quad i = 1, \ldots, n$
  - matrix inequality ($K = \mathbf{S}^n_+$): $X \preceq_{\mathbf{S}^n_+} Y \iff Y - X$ positive semidefinite

  these two types are so common that we drop the subscript in $\preceq_K$

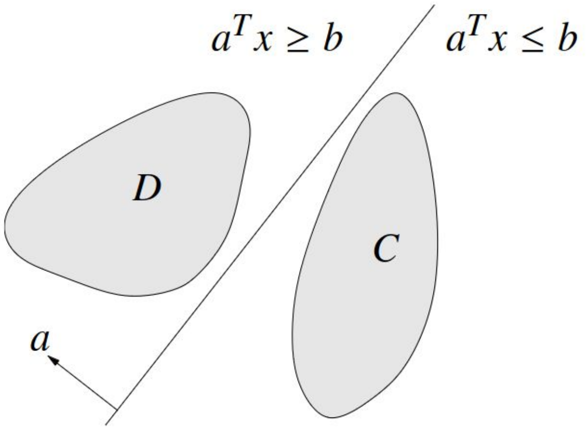
- many properties of $\preceq_K$ are similar to $\leq$ on $\mathbf{R}$, *e.g.*,

$$x \preceq_K y, \quad u \preceq_K v \implies x + u \preceq_K y + v$$

Separating hyperplane

- if C and D are nonempty disjoint (*i.e.*, $C \cap D = \emptyset$) convex sets, there exist $a \neq 0$, b s.t.

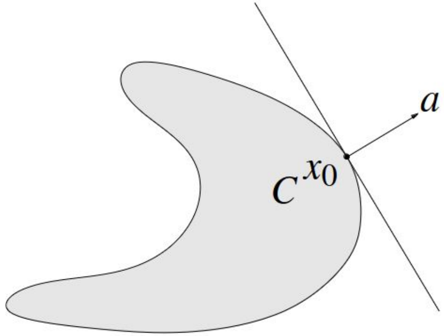
$$a^T x \leq b \text{ for } x \in C, \qquad a^T x \geq b \text{ for } x \in D$$



- the hyperplane $\{x \mid a^T x = b\}$ **separates** C and D
- strict separation requires additional assumptions (*e.g.*, C is closed, D is a singleton)

Separating hyperplane

- suppose x_0 is a boundary point of set $C \subset \mathbf{R}^n$
- **supporting hyperplane** to C at x_0 has form $\{x \mid a^T x = a^T x_0\}$, where $a \neq 0$ and $a^T x \leq a^T x_0$ for all $x \in C$



- **supporting hyperplane theorem:** if C is convex, then there exists a supporting hyperplane at every boundary point of C

Takeaway

UNIVERSITÉ CÔTE D'AZUR

$$x_1, x_2 \in C, \quad 0 \leq \theta \leq 1 \quad \Longrightarrow \quad \theta x_1 + (1-\theta)x_2 \in C$$

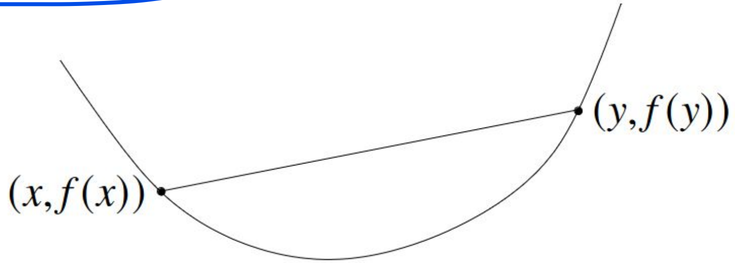
- Classic convex sets
  - Affine sets, hyperplanes, cones, balls, polyhedrons
- Convexity preserving operations
  - Intersection
  - Affine mapping
  - Perspective
  - Linear Fractional mapping

2. Convex functions

Convex functions

- $f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and for all $x, y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$,

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$$



- f is concave if $-f$ is convex
- f is strictly convex if $\mathbf{dom}\, f$ is convex and for $x, y \in \mathbf{dom}\, f$, $x \ne y$, $0 < \theta < 1$,

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$$

Convex functions (Examples)

convex functions:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- exponential: e^{ax}, for any $a \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$, for $\alpha \geq 1$ or $\alpha \leq 0$
- powers of absolute value: $|x|^p$ on $\mathbf{R}$, for $p \geq 1$
- positive part (relu): $\max\{0, x\}$

concave functions:

- affine: $ax + b$ on $\mathbf{R}$, for any $a, b \in \mathbf{R}$
- powers: x^α on $\mathbf{R}_{++}$, for $0 \leq \alpha \leq 1$
- logarithm: $\log x$ on $\mathbf{R}_{++}$
- entropy: $-x \log x$ on $\mathbf{R}_{++}$
- negative part: $\min\{0, x\}$

Extended value

- suppose f is convex on $\mathbf{R}^n$, with domain $\mathbf{dom}\, f$
- its extended-value extension $\tilde{f}$ is function $\tilde{f} : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathbf{dom}\, f \\ \infty & x \notin \mathbf{dom}\, f \end{cases}$$

- often simplifies notation; for example, the condition

$$0 \le \theta \le 1 \quad \Longrightarrow \quad \tilde{f}(\theta x + (1-\theta)y) \le \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y)$$

(as an inequality in $\mathbf{R} \cup \{\infty\}$), means the same as the two conditions

- $\mathbf{dom}\, f$ is convex
- $x, y \in \mathbf{dom}\, f,\ 0 \le \theta \le 1 \quad \Longrightarrow \quad f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$

Line restriction

- $f : \mathbf{R}^n \to \mathbf{R}$ is convex if and only if the function $g : \mathbf{R} \to \mathbf{R}$,

$$g(t) = f(x + tv), \qquad \mathbf{dom}\, g = \{t \mid x + tv \in \mathbf{dom}\, f\}$$

  is convex (in t) for any $x \in \mathbf{dom}\, f$, $v \in \mathbf{R}^n$

- can check convexity of f by checking convexity of functions of one variable

First order condition

- f is **differentiable** if $\mathbf{dom}\, f$ is open and the gradient

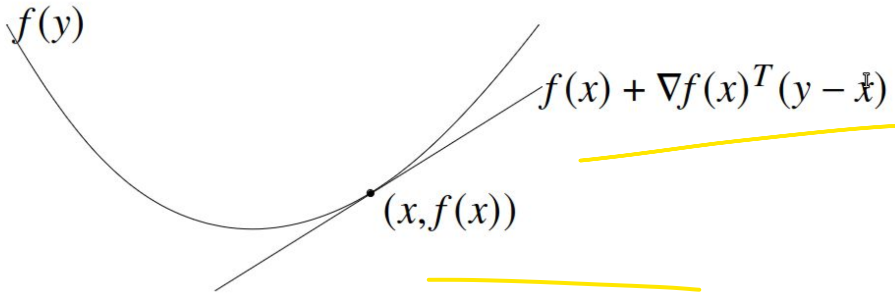
$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right) \in \mathbf{R}^n$$

  exists at each $x \in \mathbf{dom}\, f$

- **1st-order condition:** differentiable f with convex domain is convex if and only if

$$\boxed{f(y) \geq f(x) + \nabla f(x)^T (y - x)} \quad \text{for all } x, y \in \mathbf{dom}\, f$$

- first order Taylor approximation of convex f is a **global underestimator** of f



Second order condition

- f is **twice differentiable** if $\mathbf{dom}\, f$ is open and the Hessian $\nabla^2 f(x) \in \mathbf{S}^n$,

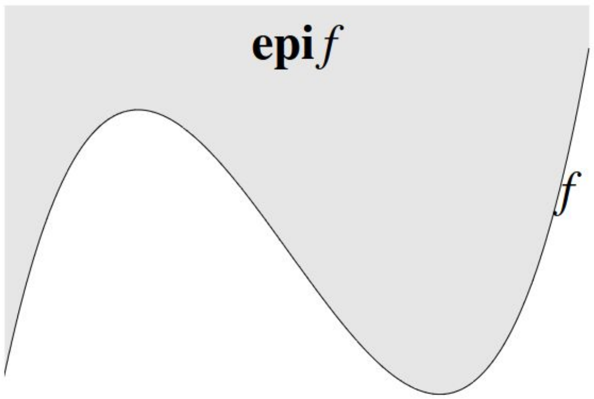
$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \qquad i, j = 1, \ldots, n,$$

  exists at each $x \in \mathbf{dom}\, f$

- **2nd-order conditions:** for twice differentiable f with convex domain
  - f is convex if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbf{dom}\, f$
  - if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex

Epigraph and sublevel set

- α**-sublevel set** of $f : \mathbf{R}^n \to \mathbf{R}$ is $C_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$
- sublevel sets of convex functions are convex sets (but converse is false)
- **epigraph** of $f : \mathbf{R}^n \to \mathbf{R}$ is $\mathbf{epi}\, f = \{(x, t) \in \mathbf{R}^{n+1} \mid x \in \mathbf{dom}\, f,\ f(x) \leq t\}$



- f is convex if and only if $\mathbf{epi}\, f$ is a convex set

Jensen's inequality

- **basic inequality:** if f is convex, then for $x, y \in \mathbf{dom}\, f$, $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

- **extension:** if f is convex and z is a random variable on $\mathbf{dom}\, f$,

$$f(\mathbf{E}\, z) \leq \mathbf{E}\, f(z)$$

- basic inequality is special case with discrete distribution

$$\mathbf{prob}(z = x) = \theta, \qquad \mathbf{prob}(z = y) = 1 - \theta$$

Showing a function is convex

methods for establishing convexity of a function f

1. verify definition (often simplified by restricting to a line)
2. for twice differentiable functions, show $\nabla^2 f(x) \succeq 0$
   - recommended only for **very simple** functions
3. show that f is obtained from simple convex functions by operations that preserve convexity
   - nonnegative weighted sum
   - composition with affine function
   - pointwise maximum and supremum
   - composition
   - minimization
   - perspective

you'll mostly use methods 2 and 3

Showing a function is convex

- **nonnegative multiple:** αf is convex if f is convex, $\alpha \geq 0$
- **sum:** $f_1 + f_2$ convex if f_1, f_2 convex
- **infinite sum:** if $f_1, f_2, \ldots$ are convex functions, infinite sum $\sum_{i=1}^{\infty} f_i$ is convex
- **integral:** if $f(x, \alpha)$ is convex in x for each $\alpha \in \mathcal{A}$, then $\displaystyle\int_{\alpha \in \mathcal{A}} f(x, \alpha)\, d\alpha$ is convex
- there are analogous rules for concave functions

Showing a function is convex

(pre-)composition with affine function: $f(Ax+b)$ is convex if f is convex

examples

- log barrier for linear inequalities

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x), \qquad \mathbf{dom}\, f = \{x \mid a_i^T x < b_i, i = 1, \ldots, m\}$$

- norm approximation error: $f(x) = \|Ax - b\|$ (any norm) if $f_1, \ldots, f_m$ are convex, then $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ is convex

examples

- piecewise-linear function: $f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$
- sum of r largest components of $x \in \mathbf{R}^n$:

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

($x_{[i]}$ is ith largest component of x)

proof: $f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$

Showing a function is convex

suprimum

if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, then $g(x) = \sup_{y \in \mathcal{A}} f(x, y)$ is convex

examples

- distance to farthest point in a set C: $f(x) = \sup_{y \in C} \|x - y\|$
- maximum eigenvalue of symmetric matrix: for $X \in \mathbf{S}^n$, $\lambda_{\max}(X) = \sup_{\|y\|_2=1} y^T X y$ is convex
- support function of a set C: $S_C(x) = \sup_{y \in C} y^T x$ is convex

Showing a function is convex

- the function $g(x) = \inf_{y \in C} f(x, y)$ is called the **partial minimization** of f (w.r.t. y)
- if $f(x, y)$ is convex in (x, y) and C is a convex set, then partial minimization g is convex

examples

- $f(x, y) = x^T Ax + 2x^T By + y^T Cy$ with

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0, \qquad C \succ 0$$

  minimizing over y gives $g(x) = \inf_y f(x, y) = x^T (A - BC^{-1} B^T)x$
  g is convex, hence Schur complement $A - BC^{-1} B^T \succeq 0$

- distance to a set: $\mathbf{dist}(x, S) = \inf_{y \in S} \|x - y\|$ is convex if S is convex

Showing a function is convex

- composition of $g : \mathbf{R}^n \to \mathbf{R}$ and $h : \mathbf{R} \to \mathbf{R}$ is $f(x) = h(g(x))$ (written as $f = h \circ g$)
- composition f is convex if
  - g convex, h convex, $\tilde{h}$ nondecreasing
  - or g concave, h convex, $\tilde{h}$ nonincreasing

  (monotonicity must hold for extended-value extension $\tilde{h}$)
- proof (for $n = 1$, differentiable g, h)

$$f''(x) = h''(g(x))g'(x)^2 + h'(g(x))g''(x)$$

examples

- $f(x) = \exp g(x)$ is convex if g is convex
- $f(x) = 1/g(x)$ is convex if g is concave and positive

Showing a function is convex

UNIVERSITÉ CÔTE D'AZUR

- composition of $g : \mathbf{R}^n \to \mathbf{R}^k$ and $h : \mathbf{R}^k \to \mathbf{R}$ is $f(x) = h(g(x)) = h(g_1(x), g_2(x), \ldots, g_k(x))$
- f is convex if h is convex and for each i one of the following holds
  - g_i convex, $\tilde{h}$ nondecreasing in its ith argument
  - g_i concave, $\tilde{h}$ nonincreasing in its ith argument
  - g_i affine
- you will use this composition rule **constantly** throughout this course
- you need to commit this rule to memory

Perspective

UNIVERSITÉ CÔTE D'AZUR

- the **perspective** of a function $f : \mathbf{R}^n \to \mathbf{R}$ is the function $g : \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$,

$$g(x,t) = tf(x/t), \qquad \mathbf{dom}\, g = \{(x,t) \mid x/t \in \mathbf{dom}\, f,\ t > 0\}$$

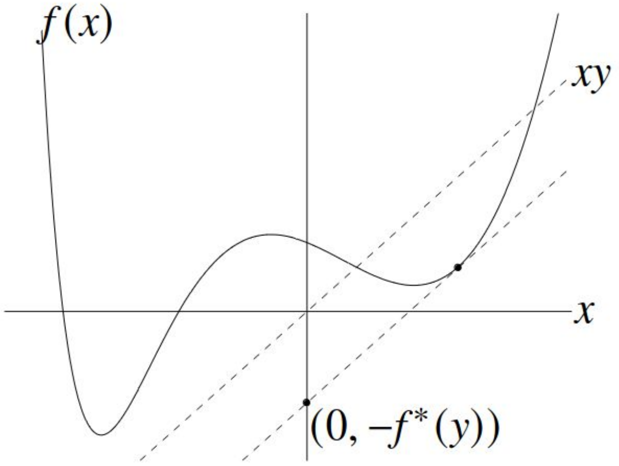
- g is convex if f is convex

examples

- $f(x) = x^T x$ is convex; so $g(x,t) = x^T x/t$ is convex for $t > 0$
- $f(x) = -\log x$ is convex; so relative entropy $g(x,t) = t\log t - t\log x$ is convex on $\mathbf{R}^2_{++}$

Conjugate

- the **conjugate** of a function f is $f^*(y) = \sup_{x \in \mathbf{dom} f} (y^T x - f(x))$



- f^* is convex (even if f is not)
- will be useful in chapter 5

Takeaway

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$$

- Classic convex functions
  - Affine, exponential, norms, max, …
- Convexity preserving operations
  - Non negative weighted sum, composition with affine
  - Pointwise maximum and supremum
  - Composition
  - Minimization
  - Perspective