Optimization for Machine Learning

Lecture 1: Overview; Convex sets & functions

SUVRIT SRA

Massachusetts Institute of Technology

16th February, 2021

Course organization

Canvas

Course materials: optml.mit.edu/teach/6881

Instructor: *Suvrit Sra*

TAs: *Kwangjun Ahn (kjahn@mit.edu)*

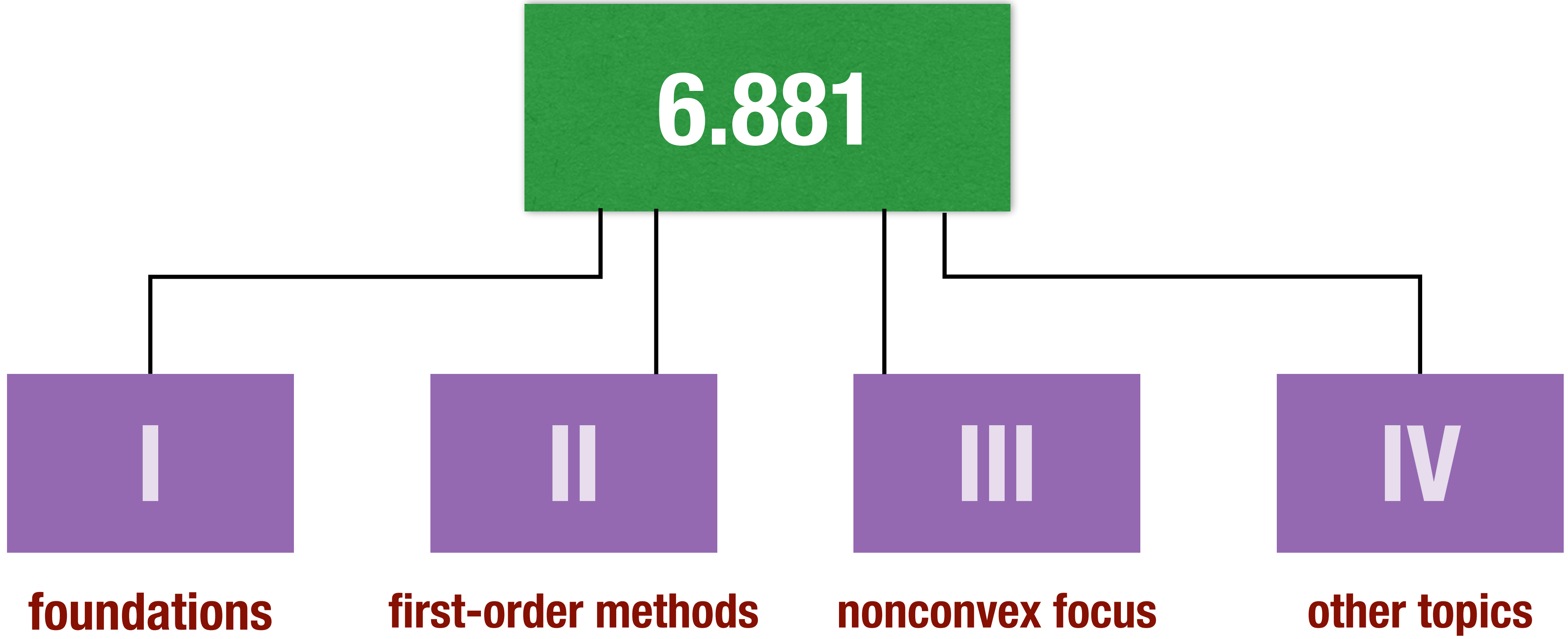
Grading: **Homework (45%), Project (50%), Peer Review (5%)**

Homework: 5 HWs. 1st one going out today

Project: The ideal project should be growable into a publishable paper at a top-tier conference or journal.

Slack: mit-optml2021.slack.com

All this information in greater detail avail via course webpage


6.881
I
II
III
IV
foundations
first-order methods
nonconvex focus
other topics

6.881

I foundations

II first-order methods

III nonconvex focus

IV other topics

1. Convex sets, functions
2. Convex conjugates, subdifferentials
3. Weak and strong duality
4. Optimality conditions - convex
5. Optimality conditions - nonconvex

Large-scale ML
Parallel and
Distributed
Big-Data

Deep Learning
GANs
Algorithms
Theory

ML & Control
Reinforcement
Learning
Robotics

Probability and Statistics
Modeling data
Unsupervised
learning

Optimization in ML

Applications
Healthcare
Engineering
Signal Proc.
Biology

Automating ML
Learning
Models and
Architectures

Interpretable ML
Interactive
Learning
Fairness

Basic terminology

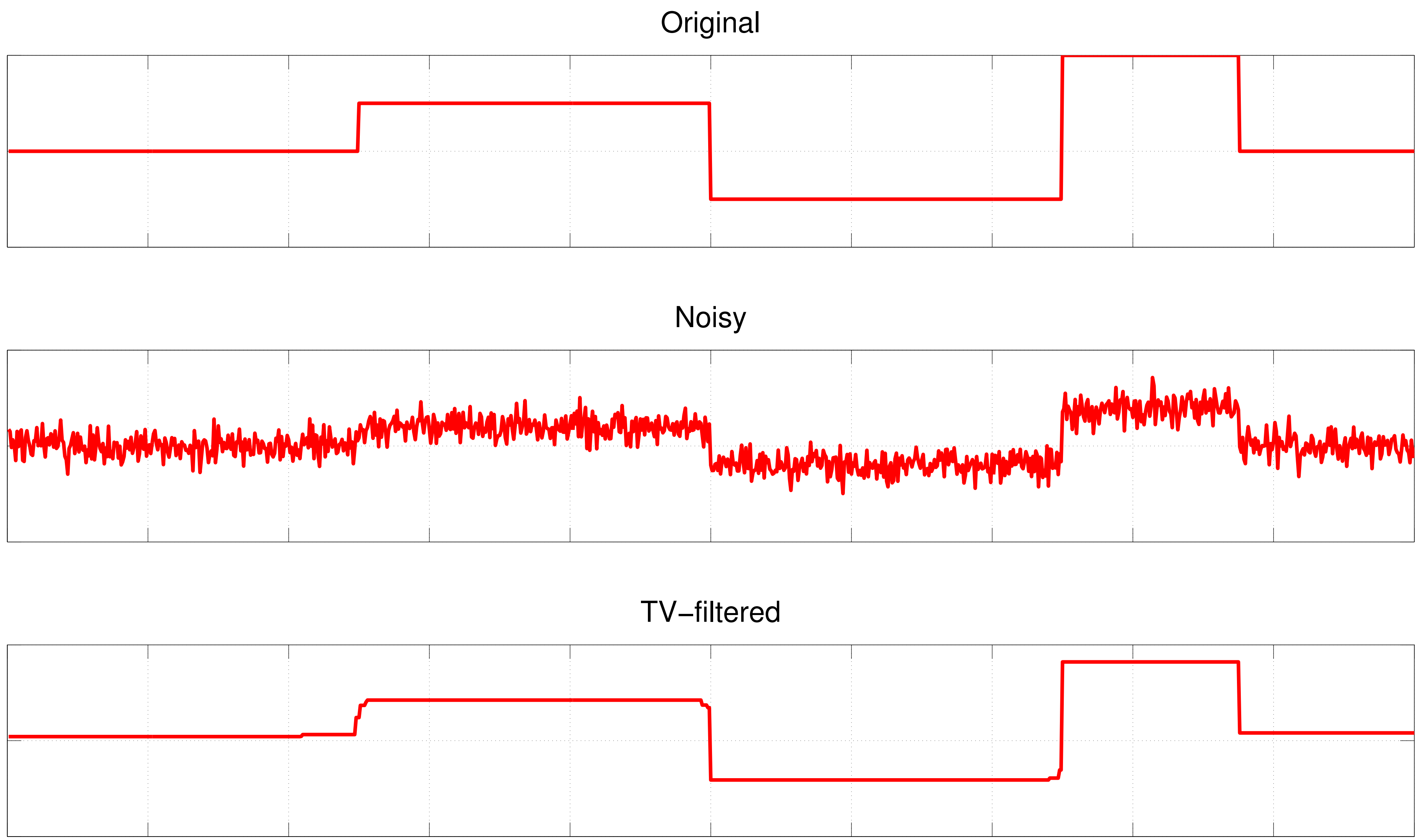
$$\min_{x} \quad f(x)$$

$$\underbrace{\textsc{Constr}(x) = \texttt{true}}$$

variable
parameter
weight vector

constraints that must
be satisfied, exactly or
approximately
(feasible set)

cost
objective
loss function



$$\frac{1}{2}\|x-y\|^2+\lambda\sum_i|x_{i+1}-x_i|$$

Trend detection via total variation denoising

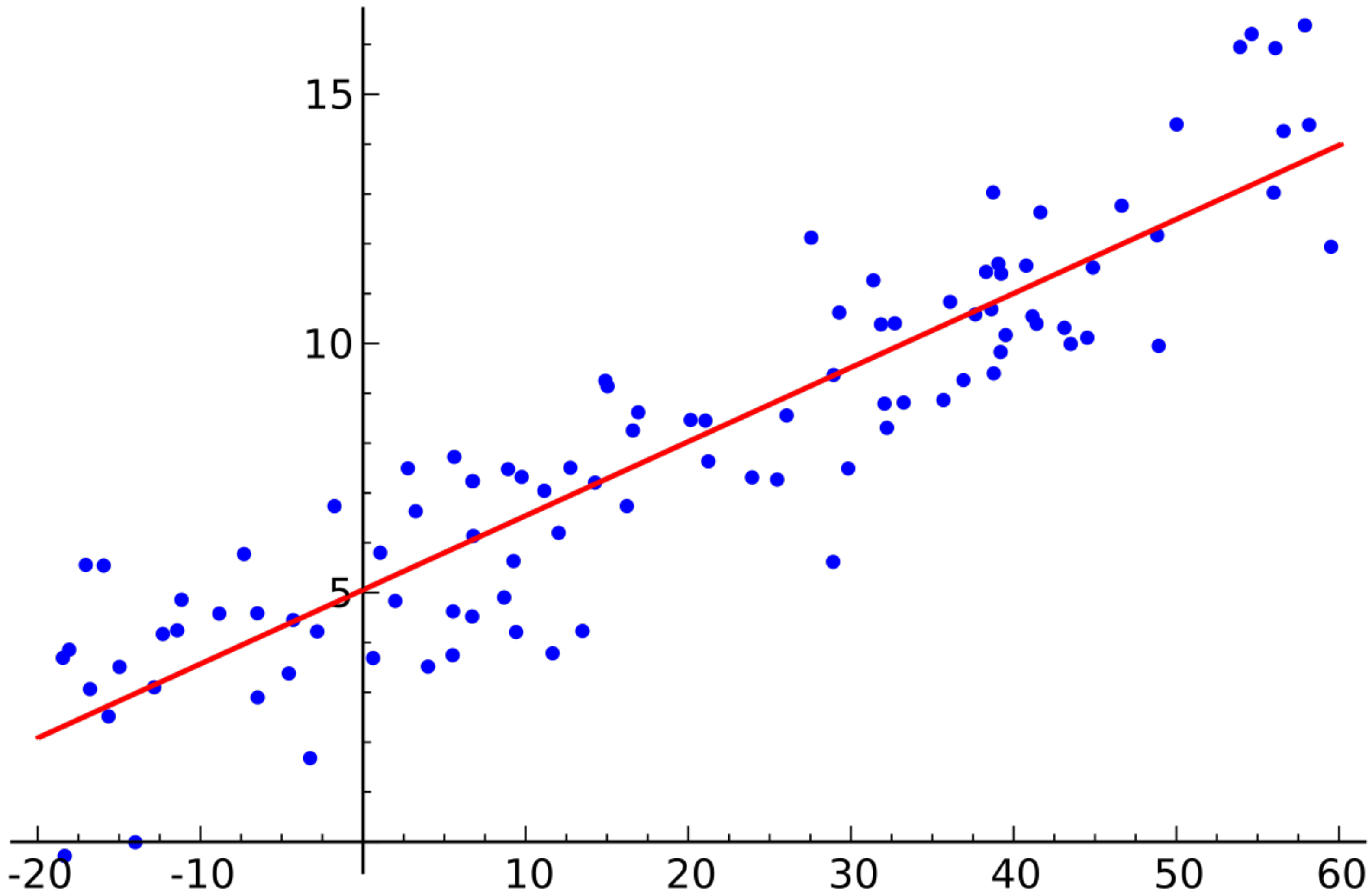


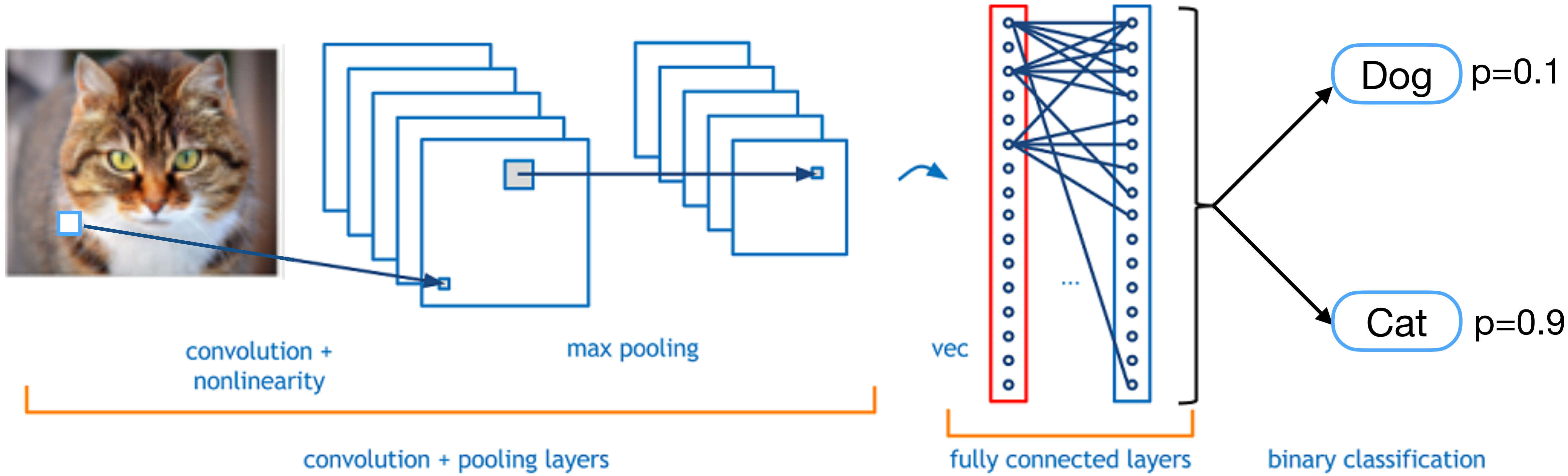
Image: Wikipedia

$$\min_x \quad \frac{1}{2}\|Ax - b\|^2$$

Least-squares Regression

$$\min_{D,x_1,\ldots,x_n} \sum_{i=1}^{n} \tfrac{1}{2}\|Dx_i - y_i\|^2 + r(x_i) + h(D)$$

Image Processing via "Dictionary Learning"



$$\min_{x} \frac{1}{n} \sum_{i=1}^{n} \ell\left(y_i, \mathrm{net}(x, a_i)\right)$$

Classification using CNNs

Model: convolutional neural net

Params x: weights of the network

Data (a_i, y_i): ( ,yes),...,( ,no)

Note: This is the simplified Empirical Risk problem; ideally want to min loss over unseen data.

If you bought this,
you may like to add …

??

Let V be the set of all items.
Let current set of items be S.
Find new item ‘i’ by solving:

$$\max_{i \in V} \; F(S \cup \{i\})$$

Recommender Systems

F: “value of information”

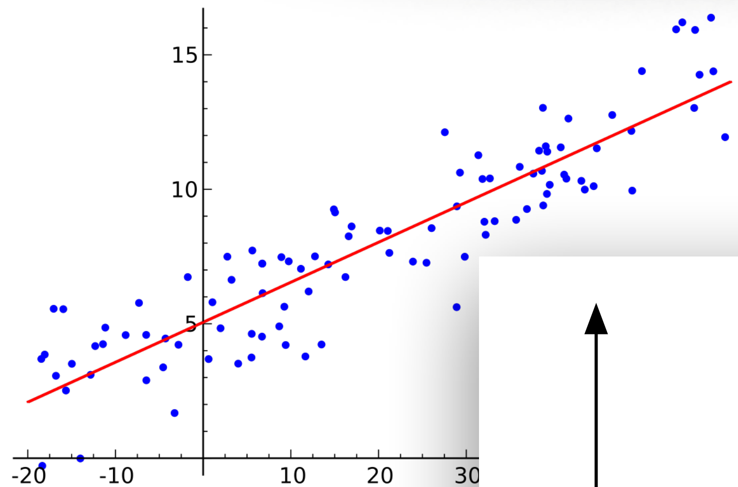
Objective functions

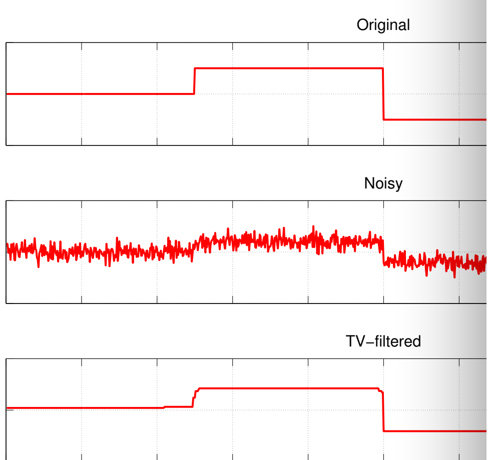
$$\min_x \; f(x)$$

Some questions:

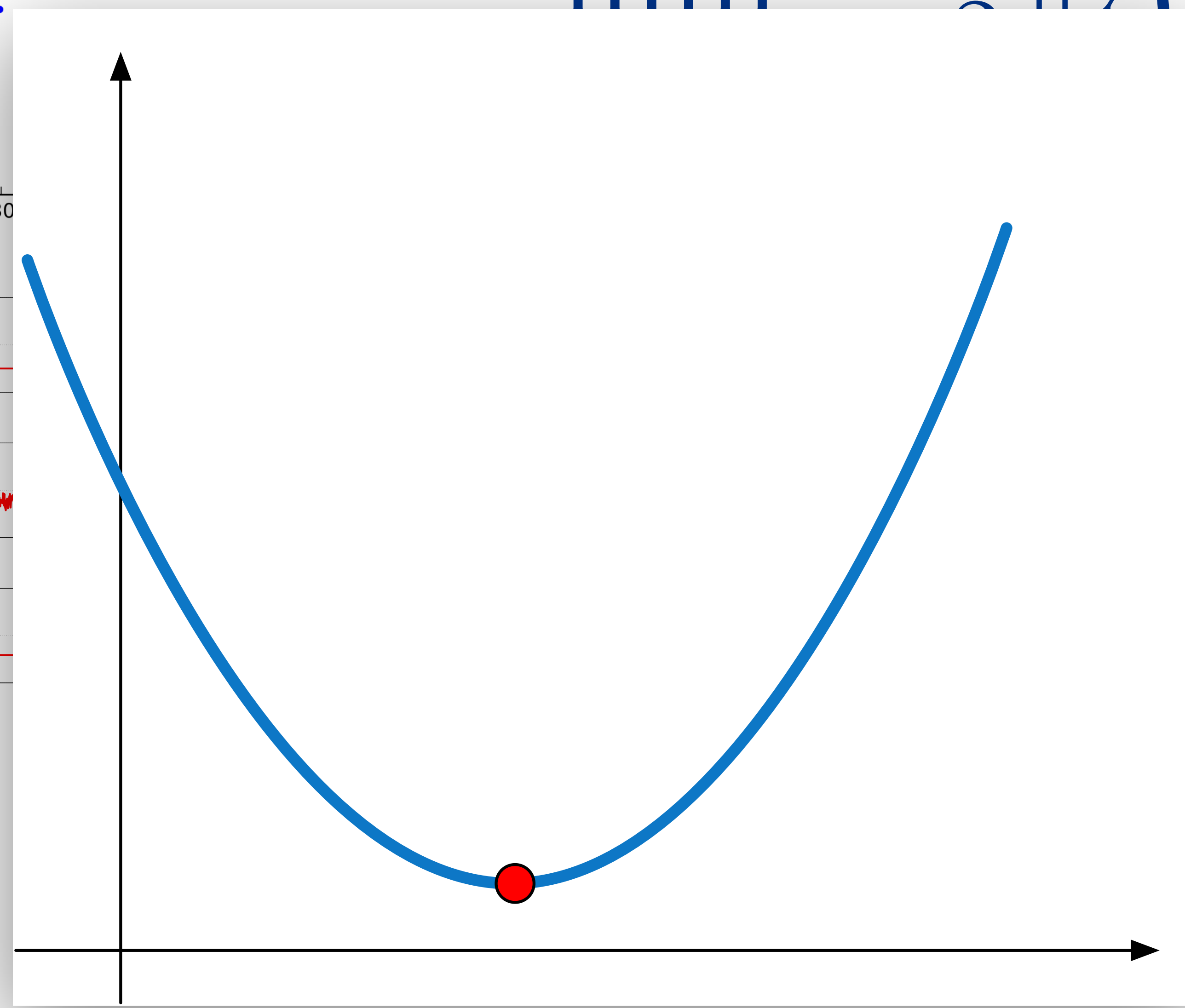
- ➡ How to select?
- ➡ Where does it come from?
- ➡ What properties may be important?
- ➡ How to actually optimize it?

Convex functions





$$\min \quad \frac{1}{2}\|Ax - b\|^2$$



$$|x_{i+1} - x_i|$$

plot of a convex function

Outline

- Convex Sets
- Convex Functions
- Recognizing, Constructing cvxfns
- Important examples: indicators, vector & matrix norms
- Exercises and Challenges

Outline

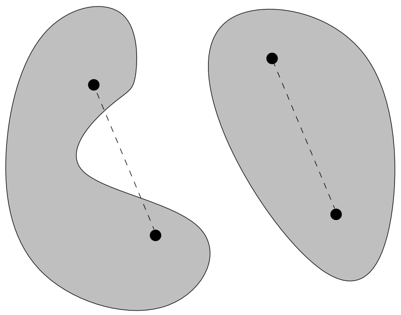
- Convex Sets
- Convex Functions
- Recognizing, Constructing cvxfns
- Important examples: indicators, vector & matrix norms
- Exercises and Challenges

Reading suggestion

cvx sets: Read BV Chapter 2
cvx func: Read BV Chapter 3
examples: Read BV Chapter 4

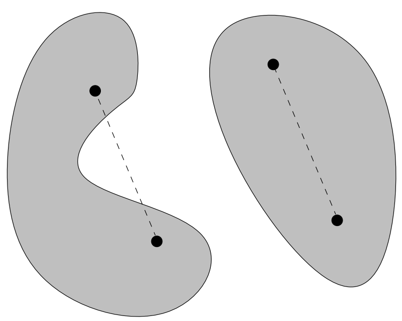
Convex sets

Def. A set $C \subset \mathbb{R}^d$ is called **convex**, if for any $x, y \in C$, the line-segment $\theta x + (1-\theta)y$ (here $0 \le \theta \le 1$) also lies in C.



Convex sets

Def. A set $C \subset \mathbb{R}^d$ is called **convex**, if for any $x, y \in C$, the line-segment $\theta x + (1-\theta)y$ (here $0 \leq \theta \leq 1$) also lies in C.



Combinations of vectors

- Linear: $\theta_1 x + \theta_2 y$ for $x, y \in C$
- Conic: if we restrict $\theta_1, \theta_2 \geq 0$
- Convex: if we restrict $\theta_1, \theta_2 \geq 0$ and $\theta_1 + \theta_2 = 1$.

Convex sets

Theorem. (Intersection).
Let C_1, C_2 be convex sets. Then, $C_1 \cap C_2$ is also convex.

Convex sets

Theorem. (Intersection).
Let C_1, C_2 be convex sets. Then, $C_1 \cap C_2$ is also convex.

Proof.

If $C_1 \cap C_2 = \emptyset$, then true vacuously.
Let $x, y \in C_1 \cap C_2$. Then, $x, y \in C_1$ and $x, y \in C_2$.
But C_1, C_2 convex; hence $\theta x + (1-\theta)y$ in C_1 and in C_2.
Thus, $\theta x + (1-\theta)y \in C_1 \cap C_2$.
(Inductively follows that $\cap_{i=1}^{m} C_i$ is also convex.)

Convex sets: examples

♡ Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^d$. Their ***convex hull*** is

$$\mathrm{co}(x_1, \ldots, x_m) := \{\sum\nolimits_i \theta_i x_i \mid \theta_i \geq 0, \sum\nolimits_i \theta_i = 1\}.$$

Convex sets: examples

♡ Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^d$. Their ***convex hull*** is

$$\mathrm{co}(x_1, \ldots, x_m) := \{\sum_i \theta_i x_i \mid \theta_i \geq 0, \sum_i \theta_i = 1\}.$$

♡ ***halfspace*** $\{x \mid a^T x \leq b\}$.

Convex sets: examples

♡ Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^d$. Their ***convex hull*** is

$$\mathrm{co}(x_1, \ldots, x_m) := \{\textstyle\sum_i \theta_i x_i \mid \theta_i \geq 0, \sum_i \theta_i = 1\}.$$

♡ ***halfspace*** $\{x \mid a^T x \leq b\}$.

♡ ***polyhedron*** $\{x \mid Ax \leq b, Cx = d\}$.

Convex sets: examples

♡ Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^d$. Their ***convex hull*** is

$$\mathrm{co}(x_1, \ldots, x_m) := \{\sum_i \theta_i x_i \mid \theta_i \geq 0, \sum_i \theta_i = 1\}.$$

♡ ***halfspace*** $\{x \mid a^T x \leq b\}$.

♡ ***polyhedron*** $\{x \mid Ax \leq b, Cx = d\}$.

♡ ***ellipsoid*** $\{x \mid (x - x_0)^T A(x - x_0) \leq 1\}$, ($A$: semidefinite)

Convex sets: examples

♡ Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^d$. Their ***convex hull*** is

$$\mathrm{co}(x_1, \ldots, x_m) := \{\sum_i \theta_i x_i \mid \theta_i \geq 0, \sum_i \theta_i = 1\}.$$

♡ ***halfspace*** $\{x \mid a^T x \leq b\}$.

♡ ***polyhedron*** $\{x \mid Ax \leq b, Cx = d\}$.

♡ ***ellipsoid*** $\{x \mid (x - x_0)^T A(x - x_0) \leq 1\}$, ($A$: semidefinite)

♡ ***probability simplex*** $\{x \mid x \geq 0, \sum_i x_i = 1\}$

Convex sets: examples

♡ Let $x_1, x_2, \dots, x_m \in \mathbb{R}^d$. Their ***convex hull*** is

$$\mathrm{co}(x_1, \dots, x_m) := \{\sum_i \theta_i x_i \mid \theta_i \geq 0, \sum_i \theta_i = 1\}.$$

♡ ***halfspace*** $\{x \mid a^T x \leq b\}$.

♡ ***polyhedron*** $\{x \mid Ax \leq b, Cx = d\}$.

♡ ***ellipsoid*** $\{x \mid (x - x_0)^T A(x - x_0) \leq 1\}$, ($A$: semidefinite)

♡ ***probability simplex*** $\{x \mid x \geq 0, \sum_i x_i = 1\}$

♡ ***Convex Cones.*** A convex set $K \subset \mathbb{R}^n$ is called a ***cone*** if for $x \in K$, the ray αx is in K for all $\alpha > 0$.

Examples: *nonneg orthant* $\mathbb{R}^n_+$;
Lorentz cone $\{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid \|\boldsymbol{x}\|_2 \leq t\}$;
PSD cone $\mathbb{S}^n_+ := \{X \in \mathbb{R}^{n \times n} \mid X = X^T, \ \mathrm{Eig}(X) \geq 0\}$.

Convex sets: examples

♡ Let $x_1, x_2, \ldots, x_m \in \mathbb{R}^d$. Their ***convex hull*** is

$$\mathrm{co}(x_1, \ldots, x_m) := \left\{\sum\nolimits_i \theta_i x_i \mid \theta_i \geq 0, \sum\nolimits_i \theta_i = 1\right\}.$$

♡ ***halfspace*** $\{x \mid a^T x \leq b\}$.

♡ ***polyhedron*** $\{x \mid Ax \leq b, Cx = d\}$.

♡ ***ellipsoid*** $\{x \mid (x - x_0)^T A (x - x_0) \leq 1\}$, ($A$: semidefinite)

♡ ***probability simplex*** $\{x \mid x \geq 0, \sum_i x_i = 1\}$

♡ ***Convex Cones.*** A convex set $K \subset \mathbb{R}^n$ is called a ***cone*** if for $x \in K$, the ray αx is in K for all $\alpha > 0$.
Examples: *nonneg orthant* $\mathbb{R}^n_+$;
Lorentz cone $\{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid \|\boldsymbol{x}\|_2 \leq t\}$;
PSD cone $\mathbb{S}^n_+ := \{X \in \mathbb{R}^{n \times n} \mid X = X^T,\ \mathrm{Eig}(X) \geq 0\}$.

Exercise: Prove that these sets are convex.

Convex sets challenge (for fun)

Challenge A. *Let* $A, B \in \mathbb{S}^d$ *(*$d \times d$ *real symmetric). Then,*

$$K := \left\{(x^T Ax, x^T Bx) \mid x \in \mathbb{R}^d\right\},$$

is a convex cone in $\mathbb{R}^2$.

Challenge B. *Let* $A, B \in \mathbb{S}^d$ *(*$d \times d$ *real symmetric). Then, the set*

$$R(A, B) := \left\{(x^T Ax, x^T Bx) \mid \|x\|_2 = 1\right\} \subset \mathbb{R}^2,$$

is a compact convex set for $d \geq 3$. *Moreover, Challenge B implies A.*

These results imply tractability of some impt. nonconvex probs

Exercises: verify the following

- Intersection of arbitrary collection of cvx cones is a cvx cone
- Let $\{\boldsymbol{b}_j\}_{j\in J}$ be vectors in $\mathbb{R}^n$. Then, $\left\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{b}_j \rangle \le 0,\ j \in J\right\}$ is a convex cone (if J is finite, then this cone is polyhedral).
- A cone K is convex if and only if $K + K \subset K$.
- $\{(\boldsymbol{x}, t) \in \mathbb{R}^n \times \mathbb{R}_{++} \mid \|\boldsymbol{x}\| \le t\}$ is a cone for any norm $\|\cdot\|$
- A real symmetric matrix $\boldsymbol{A}$ is called ***copositive*** if for every nonnegative vector $\boldsymbol{x}$ we have $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \ge 0$. Verify that the set CP_n of $n \times n$ copositive matrices forms a convex cone.
- ***Spectrahedron***: the set $S := \{\boldsymbol{x} \in \mathbb{R}^n \mid x_1A_1 + \ldots + x_nA_n \succeq 0\}$ is convex for symmetric matrices $A_1, \ldots, A_n \in \mathbb{R}^{m\times m}$. Additionally, observe that the spectrahedron is the inverse image of $\mathbb{S}^m_+$ under the affine map $A(\boldsymbol{x}) = \sum_i x_iA_i$.
- The convex hull of $S = \left\{\boldsymbol{x}\boldsymbol{x}^T \mid \boldsymbol{x} \in \mathbb{R}^n\right\}$ is $\mathbb{S}^n_+$.

Convex functions

Def. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called **convex** if its domain $\mathrm{dom}(f)$ is a convex set and for any $x, y \in \mathrm{dom}(f)$ and $\theta \in [0, 1]$ we have

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

Convex functions

Def. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called **convex** if its domain $\mathrm{dom}(f)$ is a convex set and for any $x, y \in \mathrm{dom}(f)$ and $\theta \in [0, 1]$ we have

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

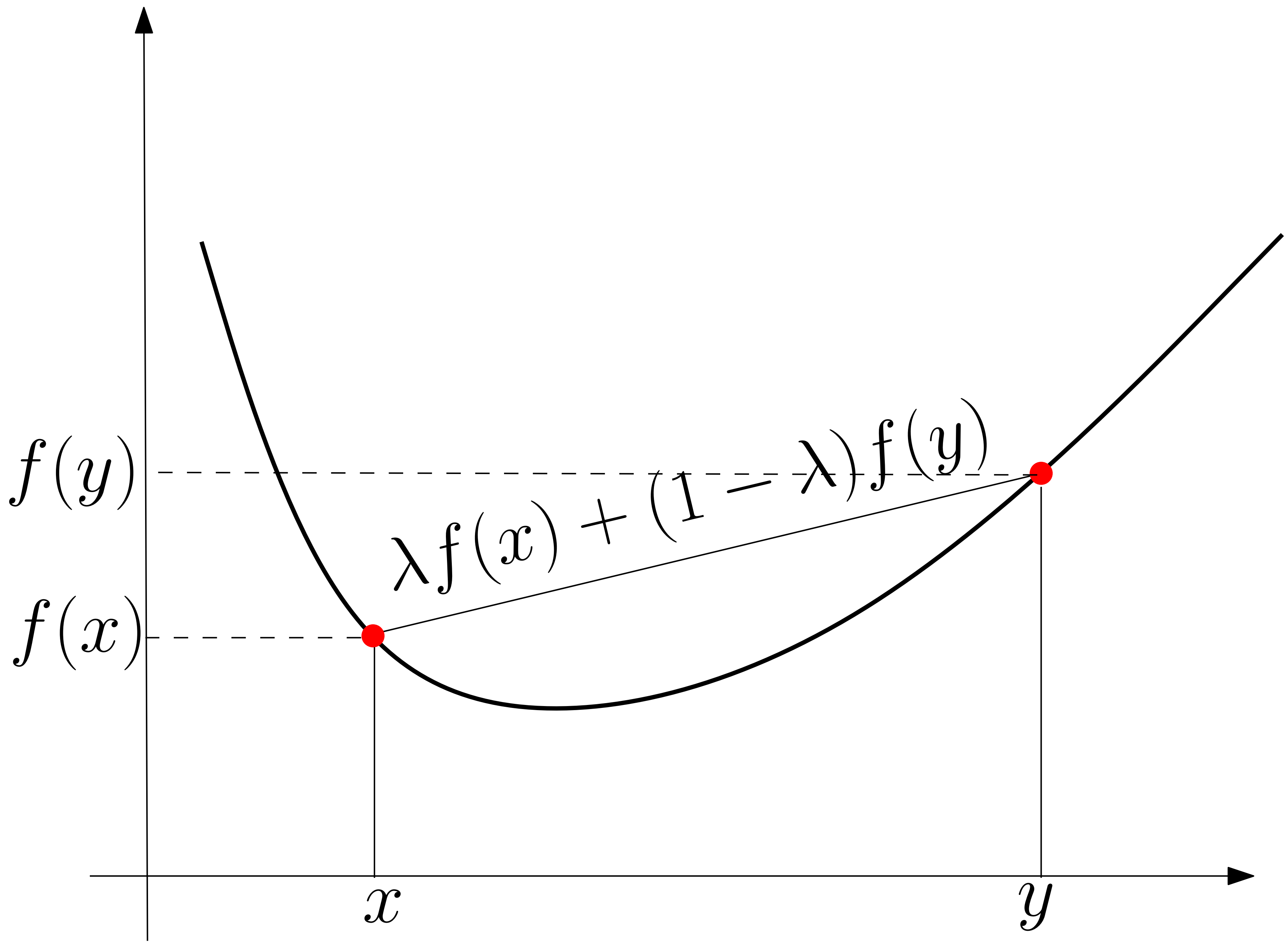
Theorem. (J.L.W.V. Jensen). Let $f : I \to \mathbb{R}$ be continuous. Then, f is convex *if and only if* it is midpoint convex, i.e., if $f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}$ for all $x, y \in I$.

Exercise: Prove Jensen's Theorem for $f : \mathcal{X} \subseteq \mathbb{R}^d \to \mathbb{R}$.

Note: Midpoint convexity often useful for checking convexity.

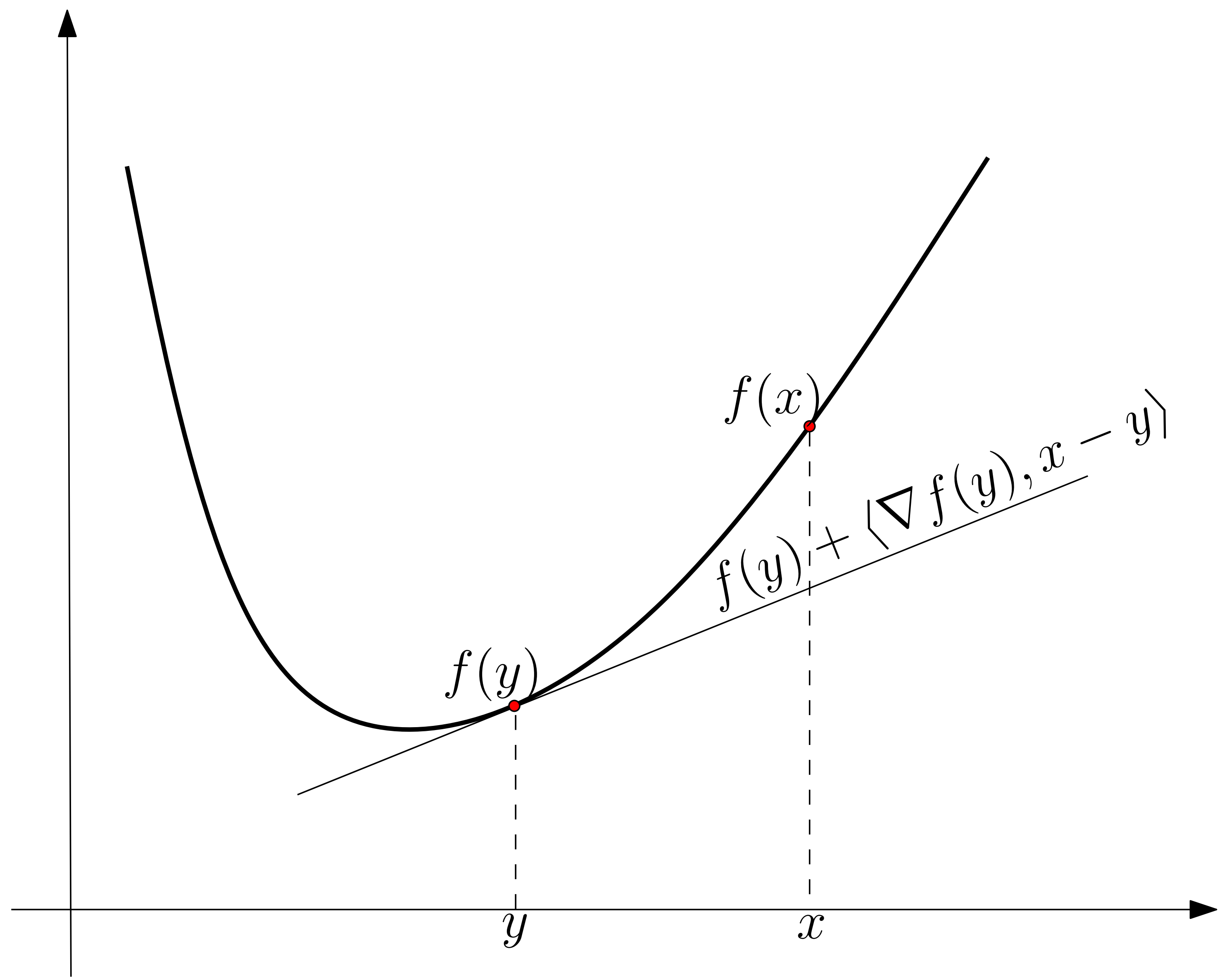
Explore: Check out Jensen's original paper on convexity!

Convex functions: curve lies below line



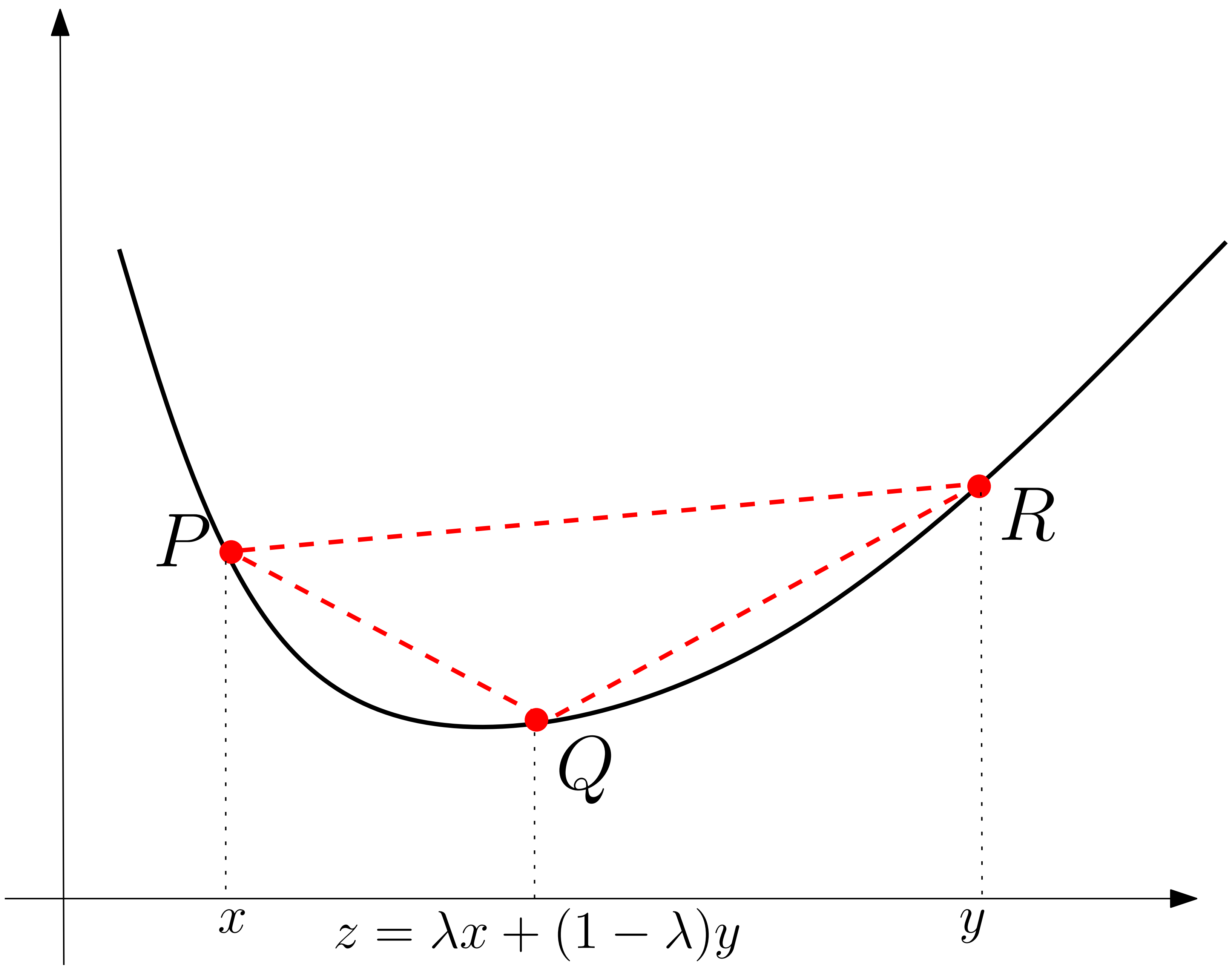
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Convex functions: curve above tangent



$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$$

Convex functions: increasing derivatives



slope PQ $\leq$ slope PR $\leq$ slope QR

Recognizing convex functions

♠ If f is continuous and midpoint convex, then it is convex.

Recognizing convex functions

♠ If f is continuous and midpoint convex, then it is convex.

♠ If f is differentiable, then f is convex *if and only if* $\operatorname{dom} f$ is convex and $f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle$ for all $x, y \in \operatorname{dom} f$.

Recognizing convex functions

- If f is continuous and midpoint convex, then it is convex.
- If f is differentiable, then f is convex *if and only if* $\mathrm{dom}\, f$ is convex and $f(x) \geq f(y) + \langle \nabla f(y),\, x - y \rangle$ for all $x, y \in \mathrm{dom}\, f$.
- If f is twice differentiable, then f is convex *if and only if* $\mathrm{dom}\, f$ is convex and $\nabla^2 f(x) \succeq 0$ at every $x \in \mathrm{dom}\, f$.
- What if f is not twice differentiable? or not even C^1?

Matrix convexity

Recall the Löwner partial order: Let A and B be Hermitian. We write $A \succeq B$ if $A - B \succeq 0$.

Suppose $f : \mathbb{S}^d \to \mathbb{S}^d$. We say f is ***matrix convex*** if

$$f((1-\lambda)X + \lambda Y) \preceq (1-\lambda)f(X) + \lambda f(Y).$$

Matrix convexity

Recall the Löwner partial order: Let A and B be Hermitian. We write $A \succeq B$ if $A - B \succeq 0$.

Suppose $f : \mathbb{S}^d \to \mathbb{S}^d$. We say f is ***matrix convex*** if

$$f((1-\lambda)X + \lambda Y) \preceq (1-\lambda)f(X) + \lambda f(Y).$$

Example. For HPD matrices $f(X) = X^2$ is matrix convex as is $f(X) = X^{-1}$. What about $-\log X$ and $\exp(X)$?

Challenge C. *X^p is matrix convex for $p \in (1, 2)$.*

Matrix convexity

Recall the Löwner partial order: Let A and B be Hermitian. We write $A \succeq B$ if $A - B \succeq 0$.

Suppose $f : \mathbb{S}^d \to \mathbb{S}^d$. We say f is ***matrix convex*** if

$$f((1-\lambda)X + \lambda Y) \preceq (1-\lambda)f(X) + \lambda f(Y).$$

Example. For HPD matrices $f(X) = X^2$ is matrix convex as is $f(X) = X^{-1}$. What about $-\log X$ and $\exp(X)$?

Challenge C. *X^p is matrix convex for $p \in (1, 2)$.*

More generally, convexity wrt a cone (see BV Chs 2,3).

Fundamental example: pointwise sup

Example. The **pointwise maximum** of a family of convex functions is convex. That is, if $f(x;y)$ is a convex function of x for every y in an arbitrary "index set" $\mathcal{Y}$, then

$$f(x) := \max_{y \in \mathcal{Y}} f(x;y)$$

is a convex function of x (set $\mathcal{Y}$ is arbitrary).

Exercise: Verify this claim!

Fundamental example: pointwise sup

Example. The **pointwise maximum** of a family of convex functions is convex. That is, if $f(x;y)$ is a convex function of x for every y in an arbitrary "index set" $\mathcal{Y}$, then

$$f(x) := \max_{y \in \mathcal{Y}} f(x;y)$$

is a convex function of x (set $\mathcal{Y}$ is arbitrary).

Exercise: Verify this claim!

Example. Let $f : \mathbb{R}^d \to \mathbb{R}$ be convex. Let $A \in \mathbb{R}^{m \times n}$, and $b \in \mathbb{R}^m$. Prove that $g(x) = f(Ax + b)$ is convex.

Exercise: Verify this claim!

Fundamental example: partial minimization

Theorem. Let $\mathcal{Y}$ be a nonempty convex set. Suppose $L(x, y)$ is convex in (x, y), then,

$$f(x) := \inf_{y \in \mathcal{Y}} \quad L(x, y)$$

is a convex function of x, provided $f(x) > -\infty$.

Fundamental example: partial minimization

Theorem. Let $\mathcal{Y}$ be a nonempty convex set. Suppose $L(x, y)$ is convex in (x, y), then,

$$f(x) := \inf_{y \in \mathcal{Y}} \quad L(x, y)$$

is a convex function of x, provided $f(x) > -\infty$.

Proof. Let $u, v \in \operatorname{dom} f$. Since $f(u) = \inf_y L(u, y)$, for each $\epsilon > 0$, there is a $y_1 \in \mathcal{Y}$, s.t. $f(u) + \frac{\epsilon}{2}$ is not the infimum. Thus, $L(u, y_1) \leq f(u) + \frac{\epsilon}{2}$.
Similarly, there is $y_2 \in \mathcal{Y}$, such that $L(v, y_2) \leq f(v) + \frac{\epsilon}{2}$.
Now we prove that $f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda) f(v)$ directly.

$$\begin{aligned}
f(\lambda u + (1-\lambda)v) &= \inf_{y \in \mathcal{Y}} L(\lambda u + (1-\lambda)v, y) \\
&\leq L(\lambda u + (1-\lambda)v, \lambda y_1 + (1-\lambda) y_2) \\
&\leq \lambda L(u, y_1) + (1-\lambda) L(v, y_2) \\
&\leq \lambda f(u) + (1-\lambda) f(v) + \epsilon.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, claim follows.

Cool application: Schur complements

Let A, B, C be matrices such that $C \succ 0$, and let

$$Z := \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0,$$

then the **Schur complement**[1] $A - BC^{-1}B^T \succeq 0$.

[1] Extremely important object in linear algebra and matrix theory.

Cool application: Schur complements

Let A, B, C be matrices such that $C \succ 0$, and let

$$Z := \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0,$$

then the **Schur complement**[1] $A - BC^{-1}B^T \succeq 0$.

Proof. (Skip ahead and solve $\nabla_y L(x,y) = 0$ to minimize!)

$L(x,y) = [x,y]^T Z [x,y]$ is convex in (x,y) since $Z \succeq 0$.

Observe that $f(x) = \inf_y L(x,y) = x^T(A - BC^{-1}B^T)x$ is convex.
Thus, its Hessian $\nabla^2 f(x) = A - BC^{-1}B^T \succeq 0$.

[1] Extremely important object in linear algebra and matrix theory.

Fundamental example: indicator function

Let $\mathbb{1}_{\mathcal{X}}$ be the *indicator function* for $\mathcal{X}$ defined as:

$$\mathbb{1}_{\mathcal{X}}(x) := \begin{cases} 0 & \text{if } x \in \mathcal{X}, \\ \infty & \text{otherwise.} \end{cases}$$

Exercise: Verify $\mathbb{1}_{\mathcal{X}}(x)$ is convex if and only if $\mathcal{X}$ is convex.

Fundamental example: indicator function

Let $\mathbb{1}_{\mathcal{X}}$ be the *indicator function* for $\mathcal{X}$ defined as:

$$\mathbb{1}_{\mathcal{X}}(x) := \begin{cases} 0 & \text{if } x \in \mathcal{X}, \\ \infty & \text{otherwise.} \end{cases}$$

Exercise: Verify $\mathbb{1}_{\mathcal{X}}(x)$ is convex if and only if $\mathcal{X}$ is convex.

Example. Using $\mathbb{1}_{\mathcal{X}}(x)$ we can rewrite the *constrained problem*

$$\min_x f(x), \quad x \in \mathcal{X},$$

as the following *unconstrained problem*

$$\min_x f(x) + \mathbb{1}_{\mathcal{X}}(x).$$

Important example: Norms

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a function that satisfies

1. $f(x) \geq 0$, and $f(x) = 0$ if and only if $x = 0$ (definiteness)
2. $f(\lambda x) = |\lambda| f(x)$ for any $\lambda \in \mathbb{R}$ (positive homogeneity)
3. $f(x + y) \leq f(x) + f(y)$ (subadditivity)

Such a function is called a *norm*.
We usually denote norms by $\|\cdot\|$.

Theorem. Norms are convex.

Proof. Immediate from subadditivity and positive homogeneity.

Important example: Distance function

Example. Let $\mathcal{Y}$ be a convex set. Let $x \in \mathbb{R}^d$ be some point. The distance of x to the set $\mathcal{Y}$ is defined as

$$\operatorname{dist}(x, \mathcal{Y}) := \inf_{y \in \mathcal{Y}} \quad \|x - y\|.$$

Because $\|x - y\|$ is jointly convex in (x, y), the function $\operatorname{dist}(x, \mathcal{Y})$ is a convex function of x.

Vector norms

Example. The **Euclidean** or ℓ_2-norm is $\|x\|_2 = \left(\sum_i x_i^2\right)^{1/2}$

Example. Let $p \geq 1$; ℓ_p-norm is $\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$

Exercise: Verify that $\|x\|_p$ is indeed a norm.

Vector norms

Example. The **Euclidean** or ℓ_2-norm is $\|x\|_2 = \left(\sum_i x_i^2\right)^{1/2}$

Example. Let $p \geq 1$; ℓ_p-norm is $\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$

Exercise: Verify that $\|x\|_p$ is indeed a norm.

Example. (ℓ_∞-norm): $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

Vector norms

Example. The **Euclidean** or ℓ_2-norm is $\|x\|_2 = \left(\sum_i x_i^2\right)^{1/2}$

Example. Let $p \geq 1$; ℓ_p-norm is $\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$

Exercise: Verify that $\|x\|_p$ is indeed a norm.

Example. (ℓ_∞-norm): $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$

Example. (Frobenius-norm): Let $A \in \mathbb{C}^{m \times n}$. The **Frobenius** norm of A is $\|A\|_{\mathrm{F}} := \sqrt{\sum_{ij} |a_{ij}|^2}$; that is, $\|A\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(A^*A)}$.

Mixed norms

Def. Let $x \in \mathbb{R}^{n_1+n_2+\cdots+n_G}$ be a vector partitioned into subvectors $x_j \in \mathbb{R}^{n_j}$, $1 \le j \le G$. Let $\boldsymbol{p} := (p_0, p_1, p_2, \ldots, p_G)$, where $p_j \ge 1$. We define the **mixed-norm** of x as

$$\|x\|_{\boldsymbol{p}} := \big\|(\|x_1\|_{p_1}, \cdots, \|x_G\|_{p_G})\big\|_{p_0}.$$

Mixed norms

Def. Let $x \in \mathbb{R}^{n_1+n_2+\cdots+n_G}$ be a vector partitioned into subvectors $x_j \in \mathbb{R}^{n_j}$, $1 \le j \le G$. Let $\boldsymbol{p} := (p_0, p_1, p_2, \ldots, p_G)$, where $p_j \ge 1$. We define the **mixed-norm** of x as

$$\|x\|_{\boldsymbol{p}} := \left\| (\|x_1\|_{p_1}, \cdots, \|x_G\|_{p_G}) \right\|_{p_0}.$$

Example. $\ell_{1,q}$-norm: Let x be as above.

$$\|x\|_{1,q} := \sum\nolimits_{i=1}^{G} \|x_i\|_q.$$

Used in machine learning (e.g., in multi-task learning). Also shows up in combinatorics, Banach space theory, statistics, and other areas!

Matrix Norms: induced norm

Let $A \in \mathbb{R}^{m \times n}$, and let $\|\cdot\|$ be any vector norm. We define an ***induced matrix norm*** as

$$\|A\| := \sup_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Matrix Norms: induced norm

Let $A \in \mathbb{R}^{m\times n}$, and let $\|\cdot\|$ be any vector norm. We define an ***induced matrix norm*** as

$$\|A\| := \sup_{\|x\|\neq 0} \frac{\|Ax\|}{\|x\|}.$$

Verify it is a norm

- Clearly, $\|A\| = 0$ iff $A = 0$ (definiteness)
- $\|\alpha A\| = |\alpha|\ \|A\|$ (homogeneity)
- $\|A + B\| = \sup \frac{\|(A+B)x\|}{\|x\|} \le \sup \frac{\|Ax\|+\|Bx\|}{\|x\|} \le \|A\| + \|B\|$.

Operator norm

Example. Let A be any matrix. Its **operator norm** is

$$\|A\|_2 := \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

It can be shown that $\|A\|_2 = \sigma_{\max}(A)$, where $\sigma_{\max}$ is the largest singular value of A.

Operator norm

Example. Let A be any matrix. Its **operator norm** is

$$\|A\|_2 := \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

It can be shown that $\|A\|_2 = \sigma_{\max}(A)$, where $\sigma_{\max}$ is the largest singular value of A.

- **Warning!** Generally, largest eigenvalue **not** a norm!

Operator norm

Example. Let A be any matrix. Its **operator norm** is

$$\|A\|_2 := \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

It can be shown that $\|A\|_2 = \sigma_{\max}(A)$, where $\sigma_{\max}$ is the largest singular value of A.

- **Warning!** Generally, largest eigenvalue **not** a norm!
- $\|A\|_1$ and $\|A\|_\infty$—max-abs-column and max-abs-row sums.

Operator norm

Example. Let A be any matrix. Its **operator norm** is

$$\|A\|_2 := \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

It can be shown that $\|A\|_2 = \sigma_{\max}(A)$, where $\sigma_{\max}$ is the largest singular value of A.

- **Warning!** Generally, largest eigenvalue **not** a norm!
- $\|A\|_1$ and $\|A\|_\infty$—max-abs-column and max-abs-row sums.
- $\|A\|_p$ generally NP-Hard to compute for $p \notin \{1, 2, \infty\}$

Operator norm

Example. Let A be any matrix. Its **operator norm** is

$$\|A\|_2 := \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

It can be shown that $\|A\|_2 = \sigma_{\max}(A)$, where $\sigma_{\max}$ is the largest singular value of A.

- **Warning!** Generally, largest eigenvalue **not** a norm!
- $\|A\|_1$ and $\|A\|_\infty$—max-abs-column and max-abs-row sums.
- $\|A\|_p$ generally NP-Hard to compute for $p \notin \{1, 2, \infty\}$
- *Schatten p-norm:* ℓ_p-norm of vector of singular value.

Operator norm

Example. Let A be any matrix. Its **operator norm** is

$$\|A\|_2 := \sup_{\|x\|_2 \neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

It can be shown that $\|A\|_2 = \sigma_{\max}(A)$, where $\sigma_{\max}$ is the largest singular value of A.

- **Warning!** Generally, largest eigenvalue **not** a norm!
- $\|A\|_1$ and $\|A\|_\infty$—max-abs-column and max-abs-row sums.
- $\|A\|_p$ generally NP-Hard to compute for $p \notin \{1, 2, \infty\}$
- *Schatten p-norm:* ℓ_p-norm of vector of singular value.
- **Exercise:** Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ be singular values of a matrix $A \in \mathbb{R}^{m \times n}$. Prove that

$$\|A\|_{(k)} := \sum\nolimits_{i=1}^{k} \sigma_i(A),$$

is a norm; $1 \leq k \leq n$.

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \leq 1\right\}.$$

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \le 1\right\}.$$

Exercise: Verify that $\|u\|_*$ is a norm.

Dual norms

Def. Let $\|\cdot\|$ be a norm on $\mathbb{R}^d$. Its **dual norm** is

$$\|u\|_* := \sup\left\{u^T x \mid \|x\| \le 1\right\}.$$

Exercise: Verify that $\|u\|_*$ is a norm.

Note. The generalized *Hölder inequality* $u^T x \le \|u\|\|x\|_*$ follows immediately directly from definition of dual norm!

Exercise: Let $1/p + 1/q = 1$, where $p, q \ge 1$. Show that $\|\cdot\|_q$ is dual to $\|\cdot\|_p$. In particular, the ℓ_2-norm is self-dual.

Exercises and Challenges

Exercises

Prove that the following functions are convex.

- ♢ $f(x, y) = x^2/y$ for $y > 0$ on $\mathbb{R} \times \mathbb{R}_{++}$
- ♢ $f^*(y) = \sup_{x \in \mathrm{dom} f} \langle x, y \rangle - f(x)$
- ♢ $\mathrm{Tr} f(X)$, where f is scalar cvx, X Hermitian
- ♢ $f(X) = -\log\det(X)$ on positive definite matrices
- ♢ $f(x) = \log(1 + e^x)$ – logistic loss, on $\mathbb{R}$
- ♢ $f(x) = \log(\sum_j e^{a_j^T x})$ – log-sum-exp on $\mathbb{R}^d$
- ♢ $f(x) = \log \frac{1-x^a}{1-x}$ for $a \geq 5$ on $(0, 1)$
- ♢ $f(x) = \log \int_0^\infty t^{x-1} e^{-t} dt$ on $x > 0$

Challenges

Prove or disprove the following:

- ♡ $f(x) = \log \operatorname{Tr}(e^{A-xB})$ for symmetric A, B is convex on $x \in \mathbb{R}$
- ♡ $f(X) = A/\det(A)$ is convex in matrix-order
- ♡ The function $1/f$ is concave for

$$f(x) = \sum_{i=1}^{n} (-1)^{n-1} \Big(\sum_{\substack{|S|=i, S \subset [n] \\ j \in S}} x_j \Big)^{-1}, x \in \mathbb{R}^n_{++}$$

- ♡ ***Open problem:*** $x \mapsto \frac{1^x + 2^x + \cdots + (n+1)^x}{1^x + \cdots + n^x}$ is log-convex on $\mathbb{R}$