Convex Optimization

Fundamentals and Applications in Statistical Signal Processing

João Mota

EURASIP/UDRC Summer School 2020

Heriot-Watt University

Optimization Problems

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in \Omega \end{aligned}$$

Optimization Problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \end{array}$$

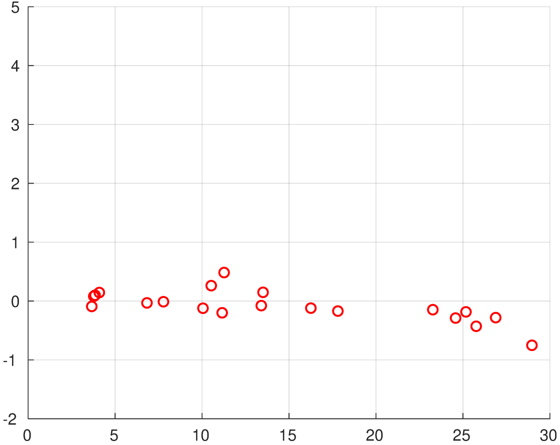
- $x \in \mathbb{R}^n$: optimization variable
- $f : \mathbb{R}^n \to \mathbb{R}$: cost function (or objective)
- $\Omega \subset \mathbb{R}^n$: constraint set

Example: Polynomial Fitting

Given $\{(x_i, y_i)\}_{i=1}^{m} \subset \mathbb{R}^2$, find "best" fitting polynomial of order $k < m$

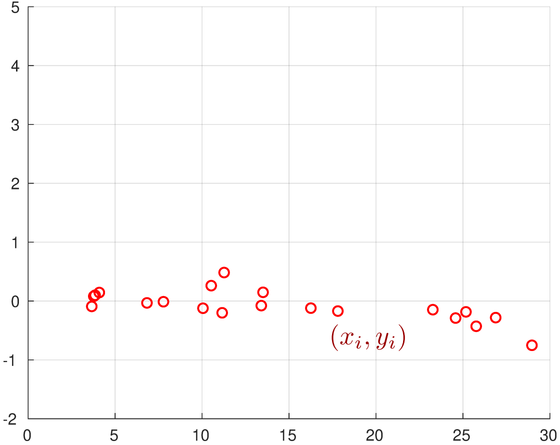
Example: Polynomial Fitting

Given $\{(x_i, y_i)\}_{i=1}^m \subset \mathbb{R}^2$, find "best" fitting polynomial of order $k < m$



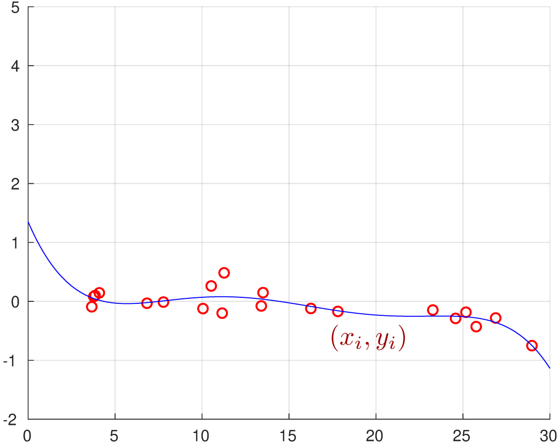
Example: Polynomial Fitting

Given $\{(x_i, y_i)\}_{i=1}^{m} \subset \mathbb{R}^2$, find "best" fitting polynomial of order $k < m$



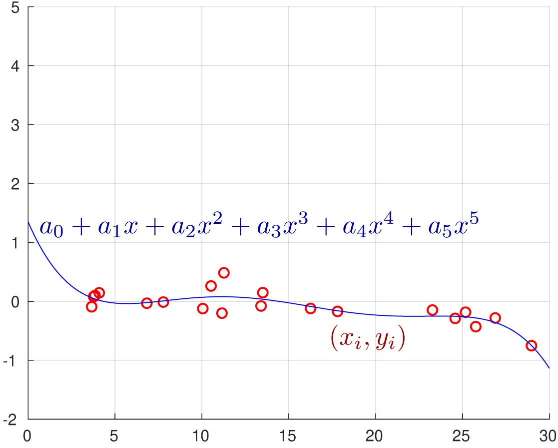
Example: Polynomial Fitting

Given $\left\{(x_i, y_i)\right\}_{i=1}^{m} \subset \mathbb{R}^2$, find "best" fitting polynomial of order $k < m$

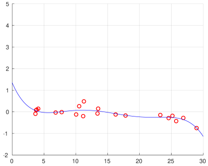


Example: Polynomial Fitting

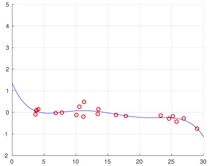
Given $\{(x_i, y_i)\}_{i=1}^m \subset \mathbb{R}^2$, find "best" fitting polynomial of order $k < m$



Example: Polynomial Fitting



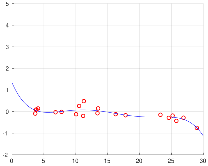
Example: Polynomial Fitting



Polynomial of order $k = 5$:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

Example: Polynomial Fitting

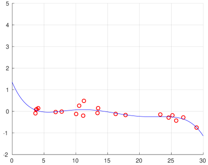


Polynomial of order $k = 5$:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

We need to find a_0, a_1, . . . , a_5 from the data $\{(x_i, y_i)\}_{i=1}^m$

Example: Polynomial Fitting



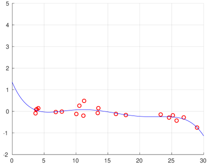
Polynomial of order $k = 5$:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

We need to find a_0, a_1, . . . , a_5 from the data $\{(x_i, y_i)\}_{i=1}^m$

Criterion: minimize the sum of squared errors (*least-squares*)

Example: Polynomial Fitting



Polynomial of order $k = 5$:

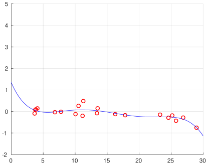
$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

We need to find a_0, a_1, . . . , a_5 from the data $\{(x_i, y_i)\}_{i=1}^m$

Criterion: minimize the sum of squared errors (*least-squares*)

$$\underset{a_0,\ldots,a_5}{\text{minimize}} \quad \sum_{i=1}^{m} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3 - a_4 x_i^4 - a_5 x_i^5 \right)^2$$

Example: Polynomial Fitting



Polynomial of order $k = 5$:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

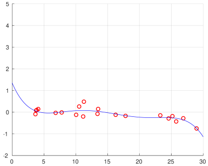
We need to find a_0, a_1, ..., a_5 from the data $\left\{(x_i, y_i)\right\}_{i=1}^m$

Criterion: minimize the sum of squared errors (*least-squares*)

$$\underset{a_0,\ldots,a_5}{\text{minimize}} \quad \sum_{i=1}^{m} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3 - a_4 x_i^4 - a_5 x_i^5 \right)^2$$

variable: $a \in \mathbb{R}^5$

Example: Polynomial Fitting



Polynomial of order $k = 5$:

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

We need to find a_0, a_1, . . . , a_5 from the data $\big\{(x_i, y_i)\big\}_{i=1}^{m}$

Criterion: minimize the sum of squared errors (*least-squares*)

$$\underset{a_0,\ldots,a_5}{\text{minimize}} \quad \sum_{i=1}^{m} \Big(y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3 - a_4 x_i^4 - a_5 x_i^5\Big)^2 = \text{f(a)}$$

variable: $a \in \mathbb{R}^5$

Example: Polynomial Fitting

$$\underset{a_0,\ldots,a_5}{\text{minimize}} \quad f(a) = \sum_{i=1}^{m} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3 - a_4 x_i^4 - a_5 x_i^5 \right)^2$$

Example: Polynomial Fitting

$$\underset{a_0,\ldots,a_5}{\text{minimize}} \quad f(a) = \sum_{i=1}^{m} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3 - a_4 x_i^4 - a_5 x_i^5 \right)^2$$

A more convenient representation: ***vectors and matrices!***

Example: Polynomial Fitting

$$\underset{a_0,\ldots,a_5}{\text{minimize}} \quad f(a) = \sum_{i=1}^{m} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3 - a_4 x_i^4 - a_5 x_i^5 \right)^2$$

A more convenient representation: ***vectors and matrices!***

$$f(a) = \left\| \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_{y} - \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & x_1^5 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & x_2^5 \\ & & & \vdots & & \\ 1 & x_m & x_m^2 & x_m^3 & x_m^4 & x_m^5 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}}_{a} \right\|_2^2$$

Example: Polynomial Fitting

$$\underset{a_0,\ldots,a_5}{\text{minimize}} \quad f(a) = \sum_{i=1}^{m} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3 - a_4 x_i^4 - a_5 x_i^5 \right)^2$$

A more convenient representation: ***vectors and matrices!***

$$f(a) = \left\| \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_{y} - \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & x_1^5 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & x_2^5 \\ & & & \vdots & & \\ 1 & x_m & x_m^2 & x_m^3 & x_m^4 & x_m^5 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}}_{a} \right\|_2^2$$

$$\underset{a \in \mathbb{R}^5}{\text{minimize}} \quad f(a) = \left\| y - Xa \right\|_2^2$$

Example: Polynomial Fitting

$$\underset{a_0,\ldots,a_5}{\text{minimize}} \quad f(a) = \sum_{i=1}^{m} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 - a_3 x_i^3 - a_4 x_i^4 - a_5 x_i^5 \right)^2$$

A more convenient representation: ***vectors and matrices!***

$$f(a) = \left\| \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_{y} - \underbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 & x_1^5 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 & x_2^5 \\ & & & \vdots & & \\ 1 & x_m & x_m^2 & x_m^3 & x_m^4 & x_m^5 \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix}}_{a} \right\|_2^2$$

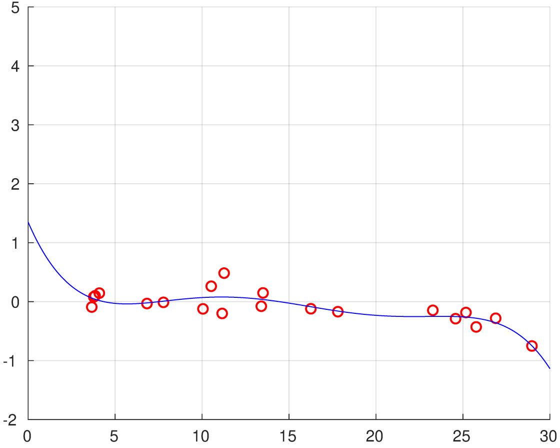
$$\underset{a \in \mathbb{R}^5}{\text{minimize}} \quad f(a) = \left\| y - Xa \right\|_2^2 \qquad \nabla f(a^\star) = 0 \iff X^\top X a^\star = X^\top y$$

Example: Polynomial Fitting

What if there are outliers?

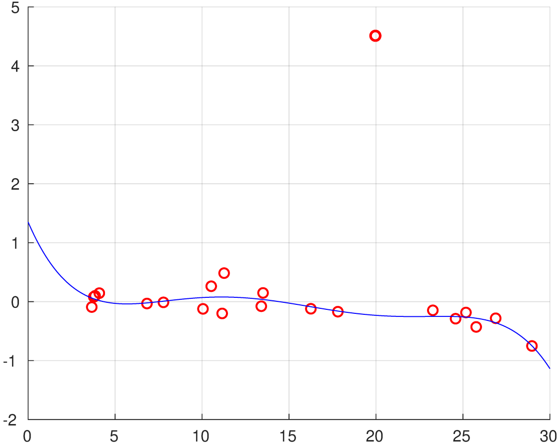
Example: Polynomial Fitting

What if there are outliers?



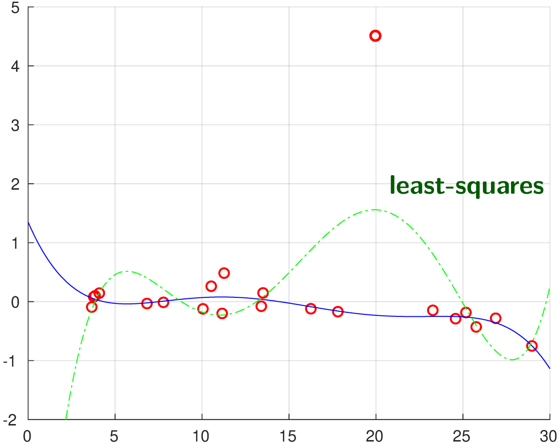
Example: Polynomial Fitting

What if there are outliers?



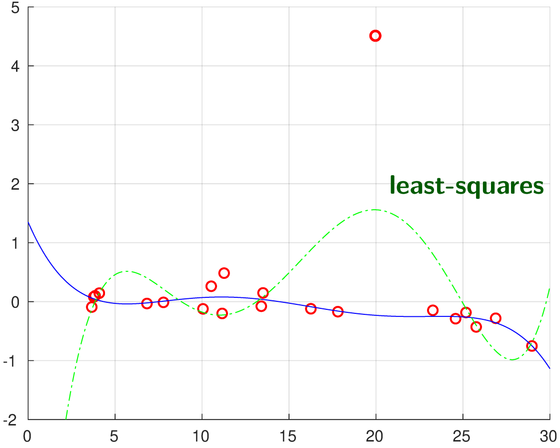
Example: Polynomial Fitting

What if there are outliers?



Example: Polynomial Fitting

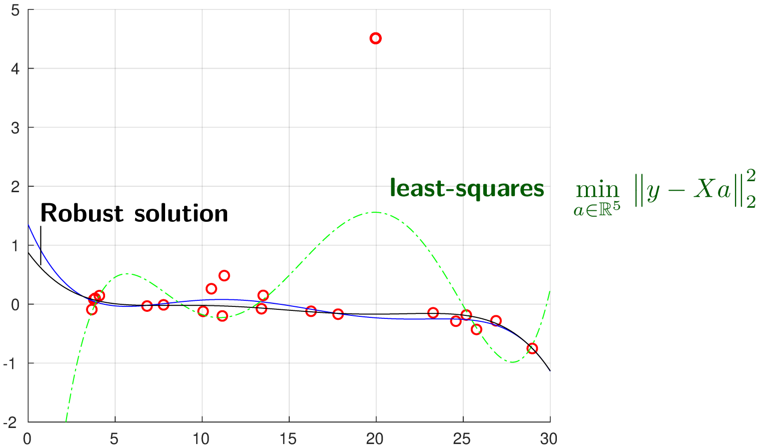
What if there are outliers?



least-squares $\min_{a \in \mathbb{R}^5} \|y - Xa\|_2^2$

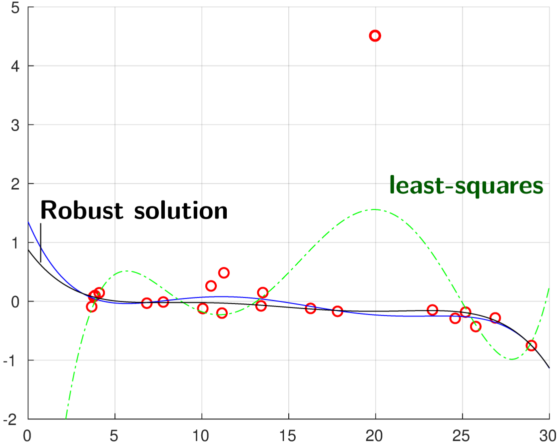
Example: Polynomial Fitting

What if there are outliers?



Example: Polynomial Fitting

What if there are outliers?



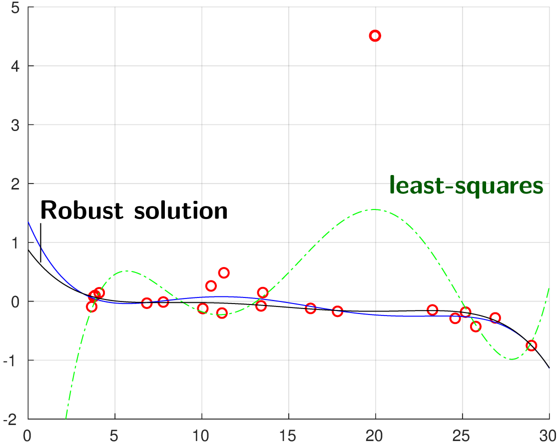
$$\min_{a\in\mathbb{R}^5} \|y - Xa\|_2^2$$

$$\Downarrow$$

$$\min_{a\in\mathbb{R}^5} \|y - Xa\|_1$$

Example: Polynomial Fitting

What if there are outliers?



$$\min_{a \in \mathbb{R}^5} \|y - Xa\|_2^2$$

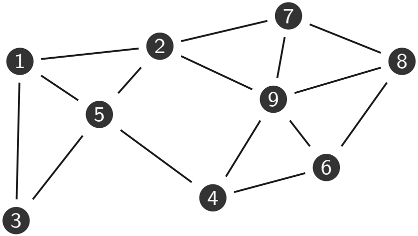
$$\Downarrow$$

$$\min_{a \in \mathbb{R}^5} \|y - Xa\|_1$$

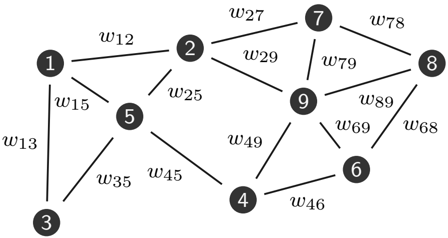
no closed-form

Example: MAXCUT

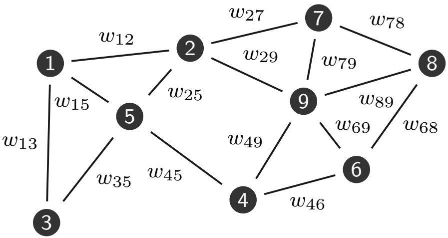
Example: MAXCUT



Example: MAXCUT

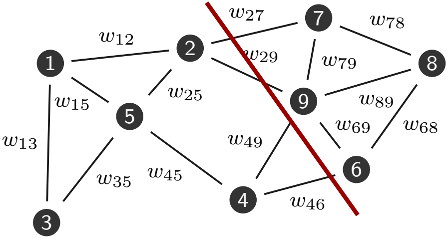


Example: MAXCUT



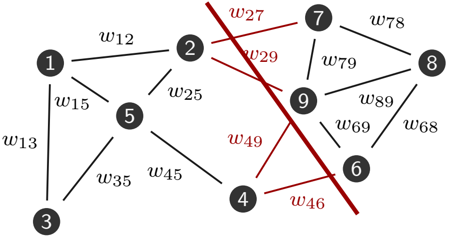
Cut: set of edges whose removal splits the graph into two

Example: MAXCUT



Cut: set of edges whose removal splits the graph into two

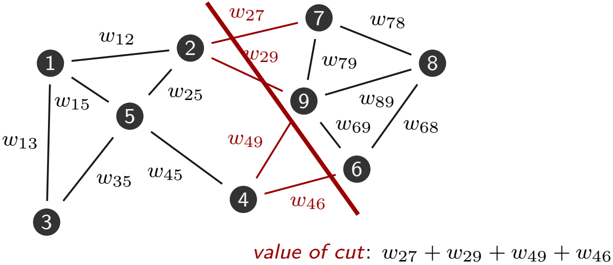
Example: MAXCUT



value of cut: $w_{27} + w_{29} + w_{49} + w_{46}$

Cut: set of edges whose removal splits the graph into two

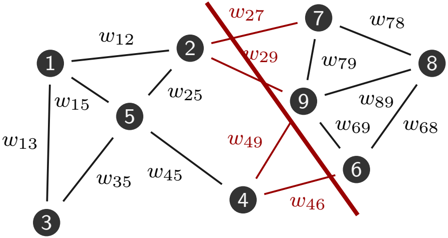
Example: MAXCUT



Cut: set of edges whose removal splits the graph into two

MAXCUT problem: find the cut with maximum weight

Example: MAXCUT



value of cut: $w_{27} + w_{29} + w_{49} + w_{46}$

Cut: set of edges whose removal splits the graph into two

MAXCUT problem: find the cut with maximum weight

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{maximize}} & \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{1 - x_i x_j}{2} \\ \text{subject to} & x_i \in \{-1, 1\}, \qquad i = 1, \ldots, n\,. \end{array}$$

Difficulty of optimization problems

- Closed-form solution ***(easy)***

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \left\| y - Ax \right\|_2^2$$

Difficulty of optimization problems

- Closed-form solution ***(easy)***

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; \left\|y - Ax\right\|_2^2$$

- No closed-form solution, but still solvable ***(easy)***

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; \left\|y - Ax\right\|_1$$

Difficulty of optimization problems

- Closed-form solution ***(easy)***

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|y - Ax\|_2^2$$

- No closed-form solution, but still solvable ***(easy)***

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|y - Ax\|_1$$

- Combinatorial, NP-Hard, requires exhaustive search ***(hard)***

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{maximize}} & \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{1 - x_i x_j}{2} \\ \text{subject to} & x_i \in \{-1, 1\}, \qquad i = 1, \ldots, n. \end{array}$$

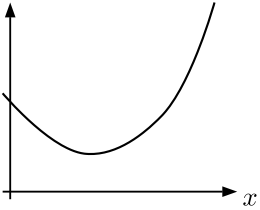
Difficulty of optimization problems

"*In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity.*" [Rockafellar, 93']

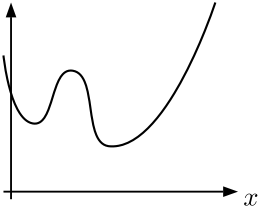
Difficulty of optimization problems

"*In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity.*" [Rockafellar, 93']

convex



nonconvex



Convex problems

Convex problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \end{array}$$

Convex problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \text{ ——— } \textit{convex function} \\ \text{subject to} & x \in \Omega \text{ ——— } \textit{convex set} \end{array}$$

Convex problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \text{ ——— } \textit{convex function} \\ \text{subject to} & x \in \Omega \text{ ——— } \textit{convex set} \end{array}$$

- Every *local* minimum is a *global* minimum

Convex problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \text{ ——— } \textit{convex function} \\ \text{subject to} & x \in \Omega \text{ ——— } \textit{convex set} \end{array}$$

- Every *local* minimum is a *global* minimum
- Solved *efficiently* (polynomial-time algorithms)

Convex problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \text{ ——— } \textit{convex function} \\ \text{subject to} & x \in \Omega \text{ ——— } \textit{convex set} \end{array}$$

- Every *local* minimum is a *global* minimum
- Solved *efficiently* (polynomial-time algorithms)
- Lots of *applications:* machine learning, communications, economics and finance, control systems, electronic circuit design, statistics, etc.

Convex problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \text{ ——— } \textit{convex function} \\ \text{subject to} & x \in \Omega \text{ ——— } \textit{convex set} \end{array}$$

- Every *local* minimum is a *global* minimum
- Solved *efficiently* (polynomial-time algorithms)
- Lots of *applications:* machine learning, communications, economics and finance, control systems, electronic circuit design, statistics, etc.
- Many algorithms for *nonconvex optimization* use convex surrogates

Convex problems

Hierarchical classification (specialized solvers):

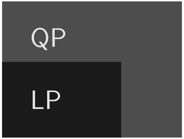
Convex problems

Hierarchical classification (specialized solvers):

LP — linear programming

Convex problems

Hierarchical classification (specialized solvers):

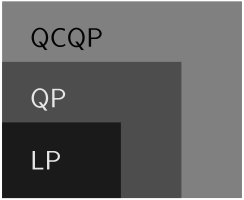


quadratic programming

linear programming

Convex problems

Hierarchical classification (specialized solvers):



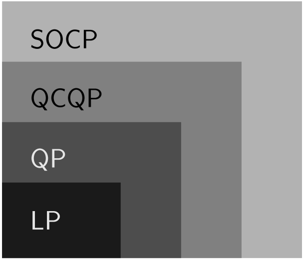
quadratically constrained QP

quadratic programming

linear programming

Convex problems

Hierarchical classification (specialized solvers):



second-order cone programming

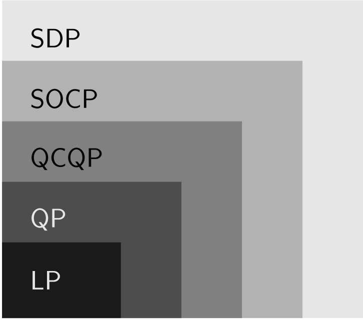
quadratically constrained QP

quadratic programming

linear programming

Convex problems

Hierarchical classification (specialized solvers):



semidefinite programming

second-order cone programming

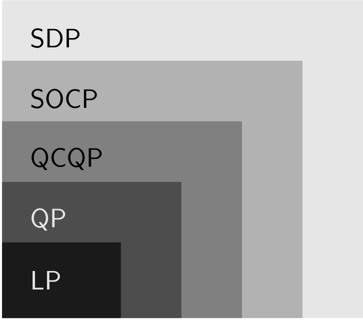
quadratically constrained QP

quadratic programming

linear programming

Convex problems

Hierarchical classification (specialized solvers):



semidefinite programming

second-order cone programming

quadratically constrained QP

quadratic programming

linear programming

Other classifications:

differentiable vs. *nondifferentiable* programming

unconstrained vs. *constrained* programming

Outline

Convex sets

Identifying convex sets

Examples: geometrical sets and filter design constraints

Convex functions

Identifying convex functions

Relation to convex sets

Optimization problems

Convex problems, properties, and problem manipulation

Examples and solvers

Statistical estimation

Maximum likelihood & maximum a posteriori

Nonparametric estimation

Hypothesis testing & optimal detection

Convex sets

Convex sets

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \end{array}$$

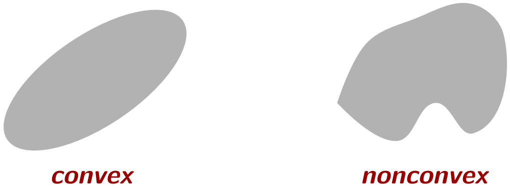
Convex sets

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \end{array}$$

$x \in \Omega$ —— *convex set*

Convex sets

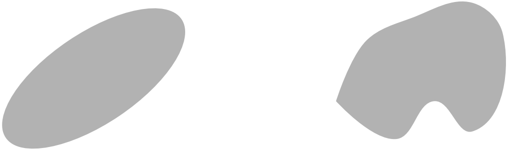
$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \quad \text{——— convex set} \end{array}$$



Convex sets

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \end{array}$$

$x \in \Omega$ — *convex set*



convex ***nonconvex***

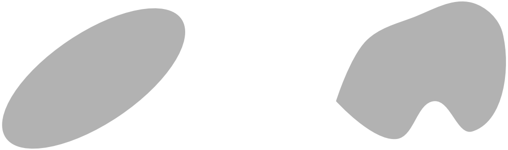
Definition:

$C \subset \mathbb{R}^n$ is *convex* when for any $x, y \in C$

Convex sets

$$\underset{x}{\text{minimize}} \quad f(x)$$

subject to $x \in \Omega$ ——— *convex set*



convex ***nonconvex***

Definition:

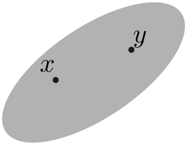
$C \subset \mathbb{R}^n$ is *convex* when for any $x, y \in C$

$$(1-\alpha)x + \alpha y \in C\,, \qquad \text{for all } 0 \le \alpha \le 1.$$

Convex sets

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \end{array}$$

— *convex set*



convex

nonconvex

Definition:

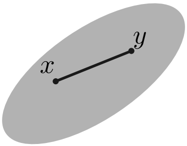
$C \subset \mathbb{R}^n$ is *convex* when for any $x, y \in C$

$$(1-\alpha)x + \alpha y \in C\,, \qquad \text{for all } 0 \le \alpha \le 1.$$

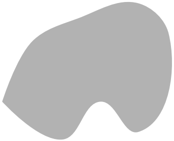
Convex sets

$$\underset{x}{\text{minimize}} \quad f(x)$$

subject to $x \in \Omega$ —— *convex set*



convex



nonconvex

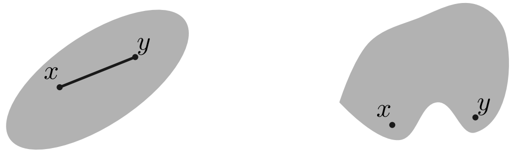
Definition:

$C \subset \mathbb{R}^n$ is *convex* when for any $x, y \in C$

$$(1-\alpha)x + \alpha y \in C\,, \qquad \text{for all } 0 \le \alpha \le 1.$$

Convex sets

$$\underset{x}{\text{minimize}} \quad f(x)$$
$$\text{subject to} \quad x \in \Omega \text{ — convex set}$$



convex ***nonconvex***

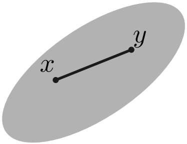
Definition:

$C \subset \mathbb{R}^n$ is *convex* when for any $x, y \in C$

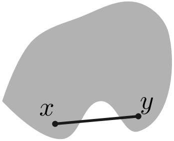
$$(1 - \alpha)x + \alpha y \in C\,, \qquad \text{for all } 0 \le \alpha \le 1.$$

Convex sets

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \text{ —— } \textit{convex set} \end{array}$$



convex



nonconvex

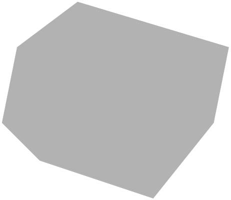
Definition:

$C \subset \mathbb{R}^n$ is *convex* when for any $x, y \in C$

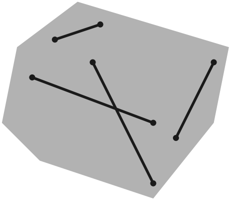
$$(1-\alpha)x + \alpha y \in C\,, \qquad \text{for all } 0 \le \alpha \le 1.$$

Examples of convex sets

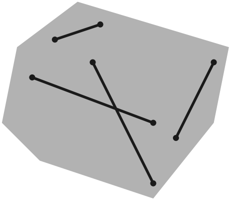
Examples of convex sets

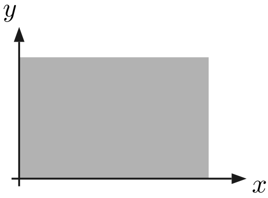


Examples of convex sets



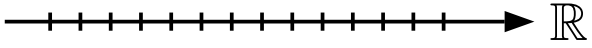
Examples of convex sets





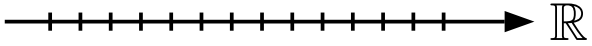
Examples of nonconvex sets

Examples of nonconvex sets

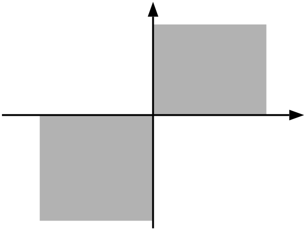


discrete sets

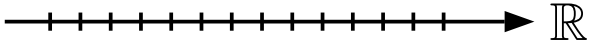
Examples of nonconvex sets



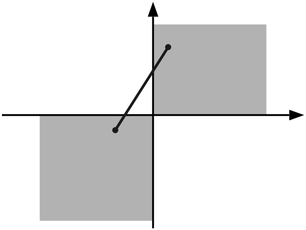
discrete sets



Examples of nonconvex sets



discrete sets



How to identify convex sets?

How to identify convex sets?

vocabulary $+$ **grammar**

How to identify convex sets?

vocabulary + **grammar**

simple sets

How to identify convex sets?

vocabulary $+$ **grammar**

simple sets — *operations preserving convexity*

Simple sets

Simple sets

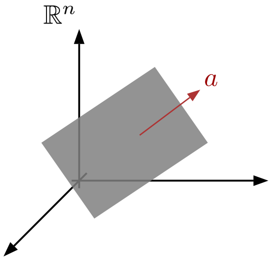
Hyperplanes

$$\mathcal{H}_{a,b} = \left\{ x \in \mathbb{R}^n \,:\, a^\top x = b \right\}$$

Simple sets

Hyperplanes

$$\mathcal{H}_{a,b} = \left\{ x \in \mathbb{R}^n \, : \, a^\top x = b \right\}$$



Simple sets

Halfspaces

Simple sets

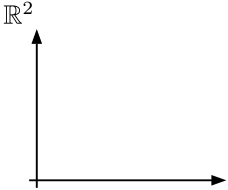
Halfspaces

$$\mathcal{H}_{a,b}^{-} = \left\{ x \in \mathbb{R}^n \,:\, a^\top x \leq b \right\}$$

Simple sets

Halfspaces

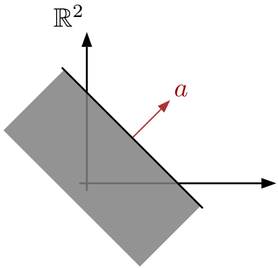
$$\mathcal{H}_{a,b}^{-} = \left\{ x \in \mathbb{R}^n \, : \, a^\top x \leq b \right\}$$



Simple sets

Halfspaces

$$\mathcal{H}_{a,b}^{-} = \left\{ x \in \mathbb{R}^n \,:\, a^\top x \leq b \right\}$$



Simple sets

ℓ_p-Norm Balls

Simple sets

ℓ_p-Norm Balls

$$B_p(c, R) = \Big\{x \in \mathbb{R}^n \,:\, \|x - c\|_p \le R\Big\}$$

Simple sets

ℓ_p-Norm Balls

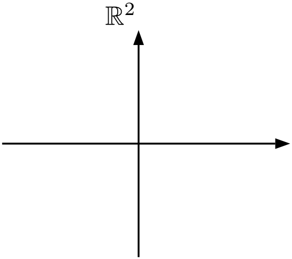
$$B_p(c,R)=\Big\{x\in\mathbb{R}^n\ :\ \|x-c\|_p\le R\Big\}$$

$$\|x\|_p=\begin{cases}\left(\displaystyle\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}} & ,\ 1\le p<\infty\\[2ex] \displaystyle\max_i\ |x_i| & ,\ p=\infty\end{cases}$$

Simple sets

ℓ_p-Norm Balls

$$B_p(c,R) = \left\{ x \in \mathbb{R}^n \,:\, \|x - c\|_p \leq R \right\}$$

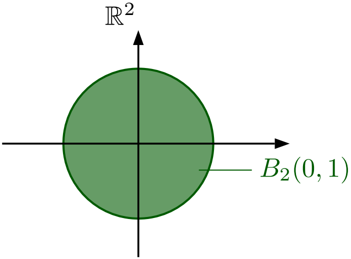


$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} & , \; 1 \leq p < \infty \\ \max_i \; |x_i| & , \; p = \infty \end{cases}$$

Simple sets

ℓ_p-Norm Balls

$$B_p(c, R) = \left\{ x \in \mathbb{R}^n \; : \; \|x - c\|_p \leq R \right\}$$

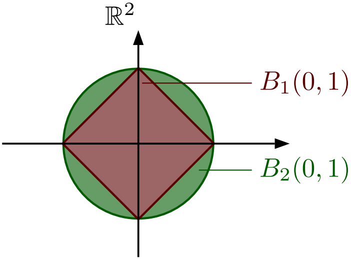


$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} & , \; 1 \leq p < \infty \\ \max_i \; |x_i| & , \; p = \infty \end{cases}$$

Simple sets

$\boldsymbol{\ell_p}$-Norm Balls

$$B_p(c, R) = \left\{ x \in \mathbb{R}^n \,:\, \|x - c\|_p \leq R \right\}$$

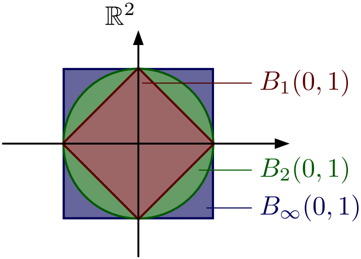


$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} & , \;\; 1 \leq p < \infty \\ \max_i \; |x_i| & , \;\; p = \infty \end{cases}$$

Simple sets

ℓ_p-Norm Balls

$$B_p(c,R) = \left\{ x \in \mathbb{R}^n \,:\, \|x - c\|_p \leq R \right\}$$



$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} & , \ 1 \leq p < \infty \\ \max_i \ |x_i| & , \ p = \infty \end{cases}$$

Simple sets

Positive Semidefinite Matrices

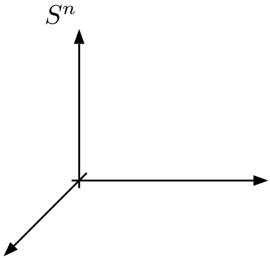
Simple sets

Positive Semidefinite Matrices

$$\mathcal{S}_n^+ = \left\{ X \in S^n \,:\, X \succeq 0_{n \times n} \right\}$$

Positive Semidefinite Matrices

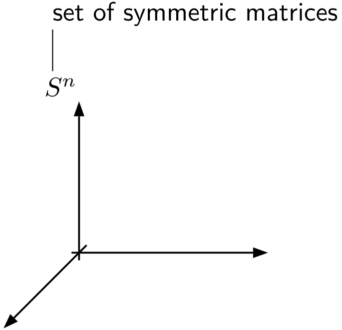
$$\mathcal{S}_n^+ = \left\{ X \in S^n \,:\, X \succeq 0_{n\times n} \right\}$$



Simple sets

Positive Semidefinite Matrices

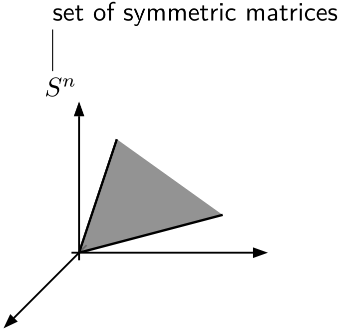
$$\mathcal{S}_n^+ = \left\{ X \in S^n \,:\, X \succeq 0_{n\times n} \right\}$$



Simple sets

Positive Semidefinite Matrices

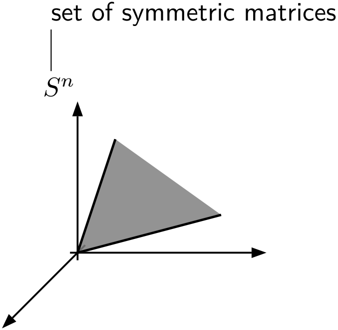
$$\mathcal{S}_n^+ = \left\{ X \in S^n \,:\, X \succeq 0_{n\times n} \right\}$$



Simple sets

Positive Semidefinite Matrices

$$\mathcal{S}_n^+ = \left\{ X \in S^n \,:\, X \succeq 0_{n\times n} \right\}$$



$$X \succeq 0_{n\times n} \quad \Longleftrightarrow \quad \lambda_{\min}(X) \geq 0 \quad \Longleftrightarrow \quad v^\top X v \geq 0\,, \quad \forall v$$

How to identify convex sets?

vocabulary $+$ **grammar**

simple sets — *operations preserving convexity*

How to identify convex sets?

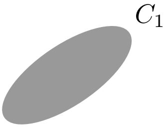
vocabulary + **grammar**

simple sets

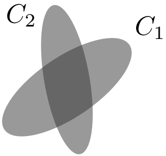
operations preserving convexity

How to identify convex sets?

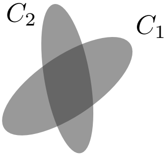
How to identify convex sets?



How to identify convex sets?

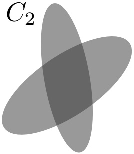


How to identify convex sets?



Intersection

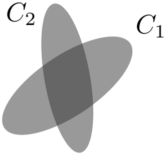
How to identify convex sets?



Intersection

$C_1, C_2, \ldots, C_m$: convex

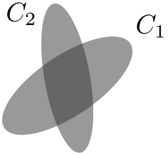
How to identify convex sets?

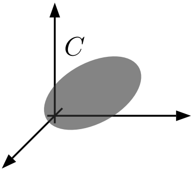


Intersection

$C_1, C_2, \ldots, C_m$: convex $\quad\Longrightarrow\quad$ $C_1 \cap C_2 \cap \cdots \cap C_m$: convex

How to identify convex sets?

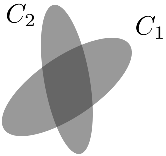


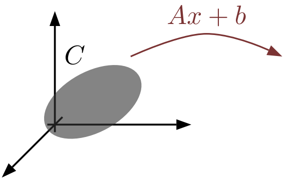


Intersection

$C_1, C_2, \ldots, C_m$: convex $\quad \Longrightarrow \quad C_1 \cap C_2 \cap \cdots \cap C_m$: convex

How to identify convex sets?

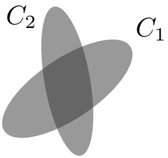


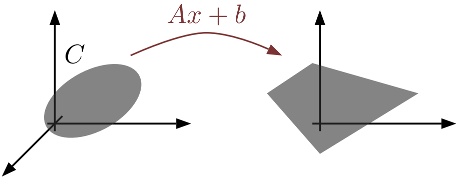


Intersection

$$C_1, C_2, \ldots, C_m \text{ : convex} \quad \Longrightarrow \quad C_1 \cap C_2 \cap \cdots \cap C_m \text{ : convex}$$

How to identify convex sets?

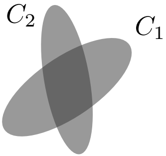


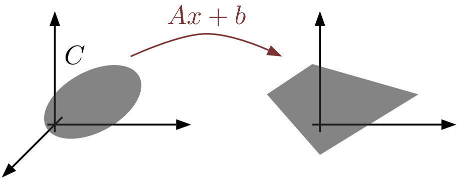


Intersection

$$C_1, C_2, \ldots, C_m \,:\, \text{convex} \qquad \Longrightarrow \qquad C_1 \cap C_2 \cap \cdots \cap C_m \,:\, \text{convex}$$

How to identify convex sets?



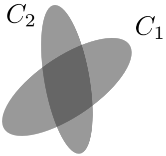


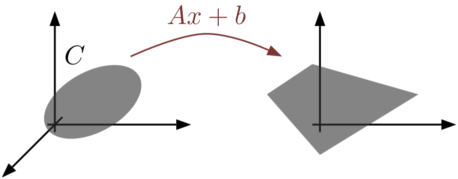
Intersection

$$C_1, C_2, \ldots, C_m \ : \ \text{convex} \qquad \Longrightarrow \qquad C_1 \cap C_2 \cap \cdots \cap C_m \ : \ \text{convex}$$

Affine operations

How to identify convex sets?





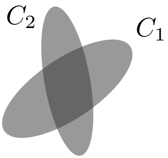
Intersection

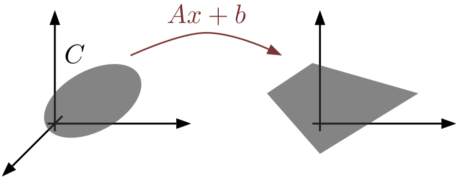
$$C_1, C_2, \ldots, C_m \ : \ \text{convex} \qquad \Longrightarrow \qquad C_1 \cap C_2 \cap \cdots \cap C_m \ : \ \text{convex}$$

Affine operations

$$C \ : \ \text{convex} \qquad \Longrightarrow \qquad \left\{ Ax + b \ : \ x \in C \right\} \ : \ \text{convex}$$

How to identify convex sets?





Intersection

$$C_1, C_2, \ldots, C_m \,:\, \text{convex} \qquad \Longrightarrow \qquad C_1 \cap C_2 \cap \cdots \cap C_m \,:\, \text{convex}$$

Affine operations

$$C \,:\, \text{convex} \qquad \Longrightarrow \qquad \Big\{ Ax + b \,:\, x \in C \Big\} \,:\, \text{convex}$$

$$C \,:\, \text{convex} \qquad \Longleftarrow \qquad \Big\{ Ax + b \,:\, x \in C \Big\} \,:\, \text{convex}$$

Example

Example

Polyhedrons

Example

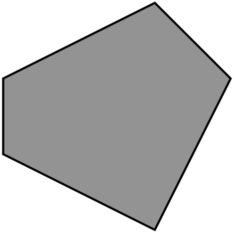
Polyhedrons

$$\mathcal{P} = \left\{ x \in \mathbb{R}^n \, : \, a_i^\top x \le b_i \, , \; i = 1, \ldots, m \right\}$$

Example

Polyhedrons

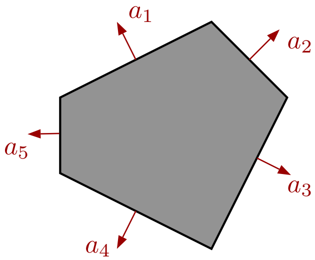
$$\mathcal{P} = \left\{ x \in \mathbb{R}^n \, : \, a_i^\top x \le b_i \, , \; i = 1, \ldots, m \right\}$$



Example

Polyhedrons

$$\mathcal{P} = \left\{ x \in \mathbb{R}^n \,:\, a_i^\top x \le b_i\,,\ i = 1, \ldots, m \right\}$$

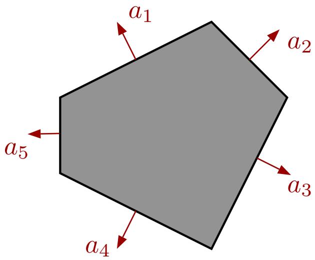


Example

Polyhedrons

$$\mathcal{P} = \left\{ x \in \mathbb{R}^n \,:\, a_i^\top x \leq b_i\,,\ i = 1, \ldots, m \right\}$$

$$= \bigcap_{i=1}^{m} \mathcal{H}^-_{a_i, b_i}$$

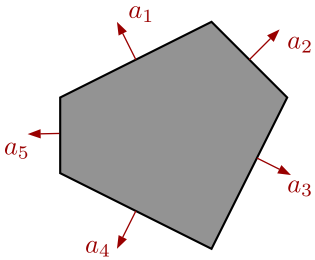


Example

Polyhedrons

$$\mathcal{P} = \left\{ x \in \mathbb{R}^n \, : \, a_i^\top x \le b_i \,, \; i = 1, \ldots, m \right\}$$

$$= \bigcap_{i=1}^{m} \underbrace{\mathcal{H}^-_{a_i, b_i}}_{\text{convex}}$$

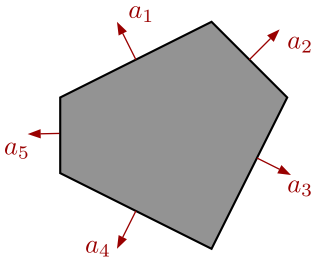


Example

Polyhedrons

$$\mathcal{P} = \left\{ x \in \mathbb{R}^n \,:\, a_i^\top x \le b_i\,,\ i = 1, \ldots, m \right\}$$

$$= \underbrace{\bigcap_{i=1}^{m} \underbrace{\mathcal{H}^-_{a_i, b_i}}_{\text{convex}}}_{\text{convex}}$$



Example

Ellipsoids

Example

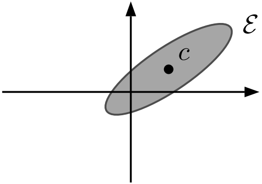
Ellipsoids $\big(A \succ 0\big)$

$$\mathcal{E} = \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\Big\}$$

Example

Ellipsoids $\left(A \succ 0\right)$

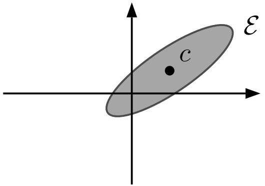
$$\mathcal{E} = \left\{ x \,:\, (x - c)^\top A^{-1} (x - c) \leq 1 \right\}$$



Example

Ellipsoids $\big(A \succ 0\big)$

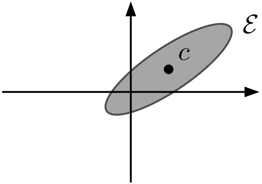
$$\mathcal{E} = \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\Big\}$$
$$= \Big\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\Big\}$$



Example

Ellipsoids $\big(A \succ 0\big)$

$$\mathcal{E} = \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\Big\}$$
$$= \Big\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\Big\}$$

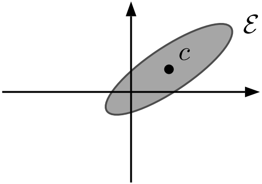


$$A \overset{\text{EVD}}{=} Q\Sigma Q^\top$$

Example

Ellipsoids $\big(A \succ 0\big)$

$$\mathcal{E} = \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\Big\}$$
$$= \Big\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\Big\}$$

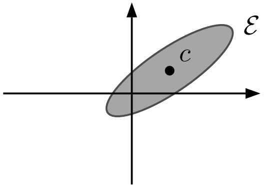


$$A \overset{\text{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top$$

Example

Ellipsoids $\big(A \succ 0\big)$

$$\mathcal{E} = \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \leq 1\Big\}$$
$$= \Big\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \leq 1\Big\}$$



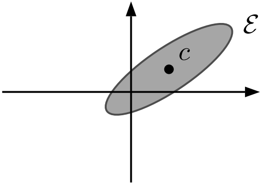
$$A \overset{\text{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = \big(Q\Sigma^{\frac{1}{2}}Q^\top\big)\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)$$

Example

Ellipsoids $\big(A \succ 0\big)$

$$\mathcal{E} = \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \leq 1\Big\}$$

$$= \Big\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \leq 1\Big\}$$

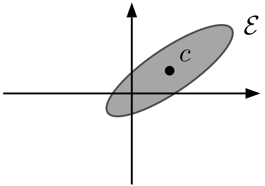


$$A \overset{\mathsf{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = \underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\, A^{\frac{1}{2}}}\underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\, A^{\frac{1}{2}}}$$

Example

Ellipsoids $\big(A \succ 0\big)$

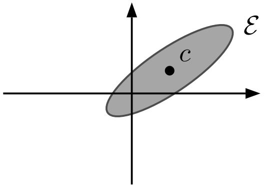
$$\mathcal{E} = \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\Big\}$$
$$= \Big\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\Big\}$$
$$= \Big\{x \,:\, \big\|A^{-\frac{1}{2}}(x-c)\big\|_2^2 \le 1\Big\}$$



$$A \overset{\text{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = \underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\, A^{\frac{1}{2}}}\underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\, A^{\frac{1}{2}}}$$

Example

Ellipsoids $\big(A \succ 0\big)$

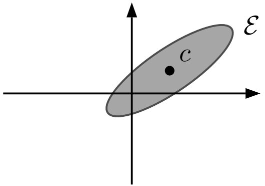


$$\begin{aligned}
\mathcal{E} &= \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\Big\} \\
&= \Big\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\Big\} \\
&= \Big\{x \,:\, \big\|A^{-\frac{1}{2}}(x-c)\big\|_2^2 \le 1\Big\} \\
&= \Big\{A^{\frac{1}{2}}y + c \,:\, \big\|y\big\|_2^2 \le 1\Big\}
\end{aligned}$$

$$A \overset{\text{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = \underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\,A^{\frac{1}{2}}}\underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\,A^{\frac{1}{2}}}$$

Example

Ellipsoids $\big(A \succ 0\big)$



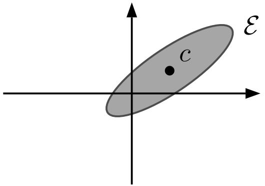
$$
\begin{aligned}
\mathcal{E} &= \left\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\right\} \\
&= \left\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\right\} \\
&= \left\{x \,:\, \left\|A^{-\frac{1}{2}}(x-c)\right\|_2^2 \le 1\right\} \\
&= \left\{A^{\frac{1}{2}} y + c \,:\, \left\|y\right\|_2^2 \le 1\right\} \\
&= A^{\frac{1}{2}} B_2(0,1) + c
\end{aligned}
$$

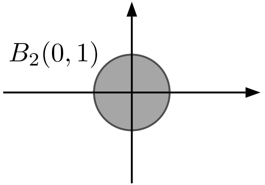
$$
A \overset{\textsf{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = \underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\, A^{\frac{1}{2}}}\underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\, A^{\frac{1}{2}}}
$$

Example

Ellipsoids $\big(A \succ 0\big)$

$$\begin{aligned}
\mathcal{E} &= \Big\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\Big\} \\
&= \Big\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\Big\} \\
&= \Big\{x \,:\, \big\|A^{-\frac{1}{2}}(x-c)\big\|_2^2 \le 1\Big\} \\
&= \Big\{A^{\frac{1}{2}} y + c \,:\, \big\|y\big\|_2^2 \le 1\Big\} \\
&= A^{\frac{1}{2}} B_2(0,1) + c
\end{aligned}$$



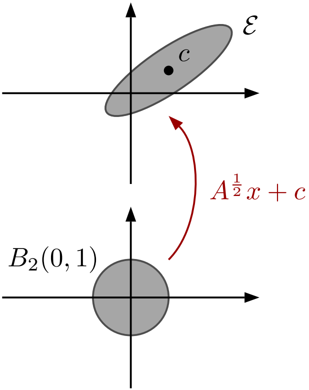


$$A \overset{\textsf{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = \underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\,A^{\frac{1}{2}}}\underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\,A^{\frac{1}{2}}}$$

Example

Ellipsoids $\left(A \succ 0\right)$

$$\mathcal{E} = \left\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\right\}$$
$$= \left\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\right\}$$
$$= \left\{x \,:\, \left\|A^{-\frac{1}{2}}(x-c)\right\|_2^2 \le 1\right\}$$
$$= \left\{A^{\frac{1}{2}}y + c \,:\, \left\|y\right\|_2^2 \le 1\right\}$$
$$= A^{\frac{1}{2}} B_2(0,1) + c$$

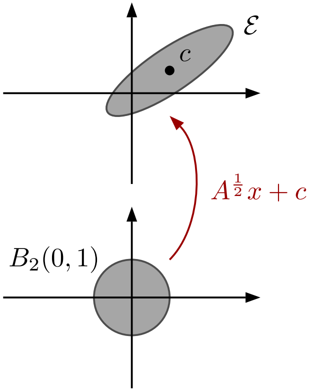


$$A \overset{\text{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = \underbrace{\left(Q\Sigma^{\frac{1}{2}}Q^\top\right)}_{=:\, A^{\frac{1}{2}}}\underbrace{\left(Q\Sigma^{\frac{1}{2}}Q^\top\right)}_{=:\, A^{\frac{1}{2}}}$$

Example

Ellipsoids $\big(A \succ 0\big)$

$$\mathcal{E} = \left\{x \,:\, (x-c)^\top A^{-1}(x-c) \le 1\right\}$$
$$= \left\{x \,:\, (x-c)^\top A^{-\frac{1}{2}} A^{-\frac{1}{2}}(x-c) \le 1\right\}$$
$$= \left\{x \,:\, \left\|A^{-\frac{1}{2}}(x-c)\right\|_2^2 \le 1\right\}$$
$$= \left\{A^{\frac{1}{2}} y + c \,:\, \|y\|_2^2 \le 1\right\}$$
$$= A^{\frac{1}{2}} B_2(0,1) + c \ :\ \textit{convex}$$



$$A \overset{\text{EVD}}{=} Q\Sigma Q^\top = Q\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = \underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\,A^{\frac{1}{2}}}\underbrace{\big(Q\Sigma^{\frac{1}{2}}Q^\top\big)}_{=:\,A^{\frac{1}{2}}}$$

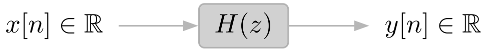
Example

Example

Filter design constraints

Example

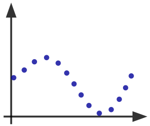
Filter design constraints



Example

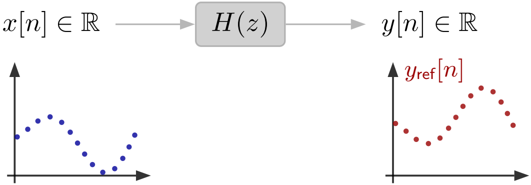
Filter design constraints

$x[n] \in \mathbb{R}$ → $H(z)$ → $y[n] \in \mathbb{R}$



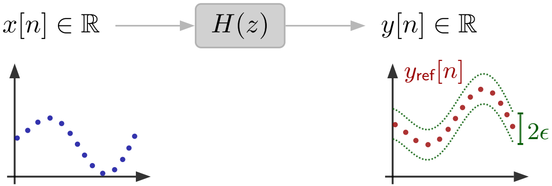
Example

Filter design constraints



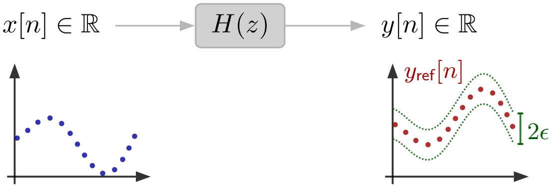
Example

Filter design constraints



Example

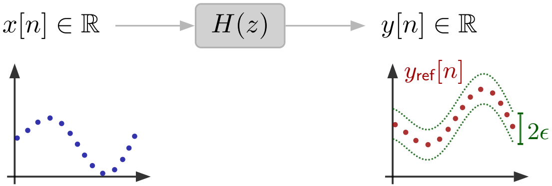
Filter design constraints



Goal: design $H(z)$ such that $\max_n \left|y[n] - y_{\text{ref}}[n]\right| \leq \epsilon$ for a fixed $x[n]$

Example

Filter design constraints

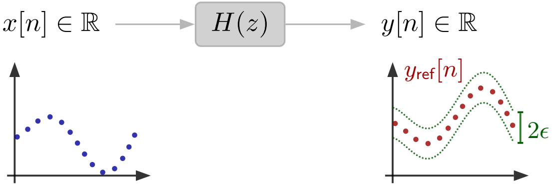


Goal: design $H(z)$ such that $\max_n \left|y[n] - y_{\text{ref}}[n]\right| \le \epsilon$ for a fixed $x[n]$

Assume finite impulse response (FIR):

Example

Filter design constraints



Goal: design $H(z)$ such that $\max_n \left| y[n] - y_{\text{ref}}[n] \right| \leq \epsilon$ for a fixed $x[n]$

Assume finite impulse response (FIR):

$$y[n] = h_0 \, x[n] + h_1 \, x[n-1] + \cdots + h_d \, x[n-d] \,, \quad n = 1, \ldots, N$$

Example

$$y[n] = h_0\, x[n] + h_1\, x[n-1] + \cdots + h_d\, x[n-d]\,, \quad n = 1, \ldots, N$$

Example

$$y[n] = h_0\, x[n] + h_1\, x[n-1] + \cdots + h_d\, x[n-d]\,, \quad n = 1, \ldots, N$$

Matrix form:

$$\underbrace{\begin{bmatrix} y[1] \\ y[2] \\ y[3] \\ \vdots \\ y[N] \end{bmatrix}}_{y \in \mathbb{R}^N} = \underbrace{\begin{bmatrix} x[1] & 0 & 0 & \cdots & 0 \\ x[2] & x[1] & 0 & \cdots & 0 \\ x[3] & x[2] & x[1] & \cdots & 0 \\ & & & & \\ x[N] & x[N-1] & x[N-2] & \cdots & x[N-d] \end{bmatrix}}_{X \in \mathbb{R}^{N \times d}} \underbrace{\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_d \end{bmatrix}}_{h \in \mathbb{R}^d}$$

Example

$$y[n] = h_0\, x[n] + h_1\, x[n-1] + \cdots + h_d\, x[n-d]\,, \quad n = 1, \ldots, N$$

Matrix form:

$$\underbrace{\begin{bmatrix} y[1] \\ y[2] \\ y[3] \\ \vdots \\ y[N] \end{bmatrix}}_{y \in \mathbb{R}^N} = \underbrace{\begin{bmatrix} x[1] & 0 & 0 & \cdots & 0 \\ x[2] & x[1] & 0 & \cdots & 0 \\ x[3] & x[2] & x[1] & \cdots & 0 \\ & & & & \\ x[N] & x[N-1] & x[N-2] & \cdots & x[N-d] \end{bmatrix}}_{X \in \mathbb{R}^{N \times d}} \underbrace{\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_d \end{bmatrix}}_{h \in \mathbb{R}^d}$$

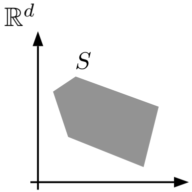
Constraint: $S = \left\{ h \in \mathbb{R}^d \,:\, \|y_{\mathsf{ref}} - Xh\|_\infty \leq \epsilon \right\}$

Example

Constraint: $S = \left\{ h \in \mathbb{R}^d \,:\, \|y_{\mathsf{ref}} - Xh\|_\infty \leq \epsilon \right\}$

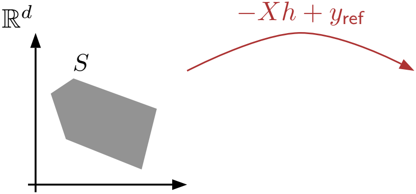
Example

Constraint: $S = \left\{ h \in \mathbb{R}^d \; : \; \|y_{\mathsf{ref}} - Xh\|_\infty \leq \epsilon \right\}$



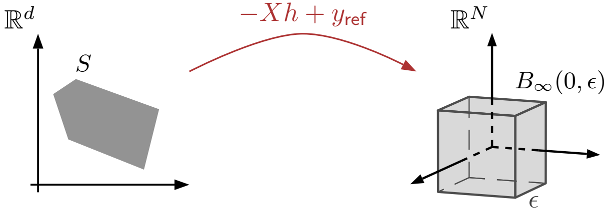
Example

Constraint: $S = \left\{ h \in \mathbb{R}^d \,:\, \|y_{\mathsf{ref}} - Xh\|_\infty \le \epsilon \right\}$



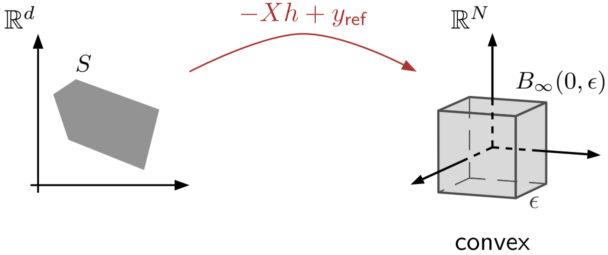
Example

Constraint: $S = \{h \in \mathbb{R}^d \,:\, \|y_{\mathsf{ref}} - Xh\|_\infty \leq \epsilon\}$



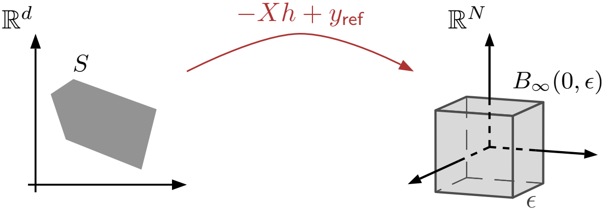
Example

Constraint: $S = \left\{ h \in \mathbb{R}^d \,:\, \|y_{\mathsf{ref}} - Xh\|_\infty \le \epsilon \right\}$



Example

Constraint: $S = \left\{ h \in \mathbb{R}^d \,:\, \|y_{\text{ref}} - Xh\|_\infty \le \epsilon \right\}$



convex $\Longleftarrow$ convex

Outline

Convex sets

Identifying convex sets

Examples: geometrical sets and filter design constraints

Convex functions

Identifying convex functions

Relation to convex sets

Optimization problems

Convex problems, properties, and problem manipulation

Examples and solvers

Statistical estimation

Maximum likelihood & maximum a posteriori

Nonparametric estimation

Hypothesis testing & optimal detection

Convex functions

Convex functions

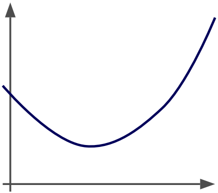
$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in \Omega \end{array}$$

Convex functions

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \quad \text{—— } \textit{convex function} \\ \text{subject to} & x \in \Omega \end{array}$$

Convex functions

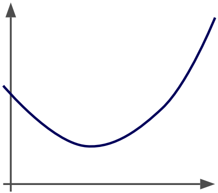
$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{—— } \textit{convex function}$$
$$\text{subject to} \quad x \in \Omega$$



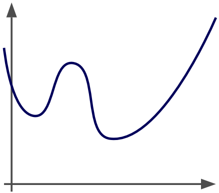
convex

Convex functions

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \quad \text{—— } \textit{convex function} \\ \text{subject to} & x \in \Omega \end{array}$$



convex

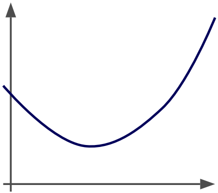


nonconvex

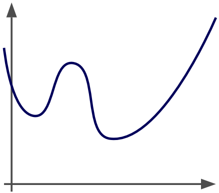
Convex functions

minimize$_x$ $f(x)$ —— *convex function*

subject to $x \in \Omega$



convex



nonconvex

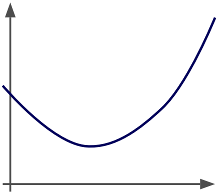
Definition:

f : dom $f \subseteq \mathbb{R}^n \to \mathbb{R}$ is *convex* when for any $x, y \in$ dom f,

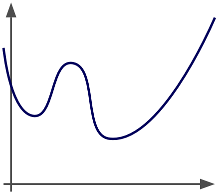
Convex functions

minimize$_x$ $f(x)$ —— *convex function*

subject to $x \in \Omega$



convex



nonconvex

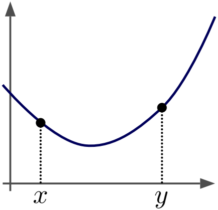
Definition:

f : dom $f \subseteq \mathbb{R}^n \to \mathbb{R}$ is *convex* when for any $x, y \in$ dom f,

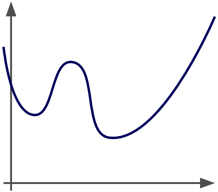
$$f\big((1-\alpha)x + \alpha y\big) \le (1-\alpha)f(x) + \alpha f(y)\,, \qquad \text{for all } 0 \le \alpha \le 1.$$

Convex functions

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{——— } \textit{convex function}$$
$$\text{subject to} \quad x \in \Omega$$



convex



nonconvex

Definition:

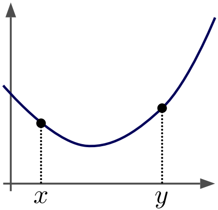
$f \;:\; \text{dom}\, f \subseteq \mathbb{R}^n \to \mathbb{R}$ is *convex* when for any $x, y \in \text{dom}\, f$,

$$f\big((1-\alpha)x + \alpha y\big) \le (1-\alpha) f(x) + \alpha f(y)\,, \qquad \text{for all } 0 \le \alpha \le 1.$$

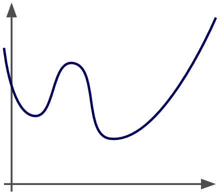
Convex functions

minimize$_x$ $f(x)$ ——— *convex function*

subject to $x \in \Omega$



convex



nonconvex

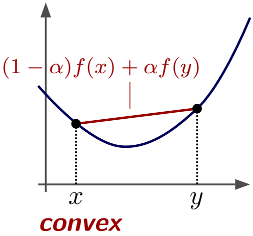
Definition:

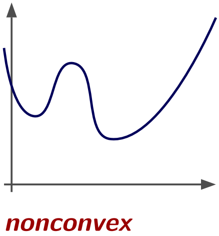
f : dom $f \subseteq \mathbb{R}^n \to \mathbb{R}$ is *convex* when for any $x, y \in$ dom f,

$$f\big((1-\alpha)x + \alpha y\big) \leq (1-\alpha) f(x) + \alpha f(y), \qquad \text{for all } 0 \leq \alpha \leq 1.$$

Convex functions

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{——— } \textit{convex function}$$
$$\text{subject to} \quad x \in \Omega$$





Definition:

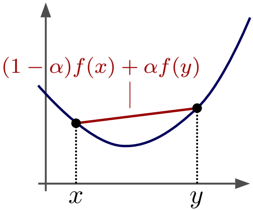
f : dom $f \subseteq \mathbb{R}^n \to \mathbb{R}$ is *convex* when for any $x, y \in$ dom f,

$$f\big((1-\alpha)x + \alpha y\big) \le (1-\alpha)f(x) + \alpha f(y), \qquad \text{for all } 0 \le \alpha \le 1.$$

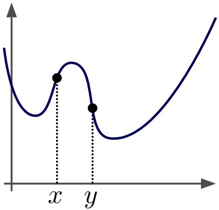
Convex functions

minimize$_x$ $f(x)$ —— *convex function*

subject to $x \in \Omega$



convex



nonconvex

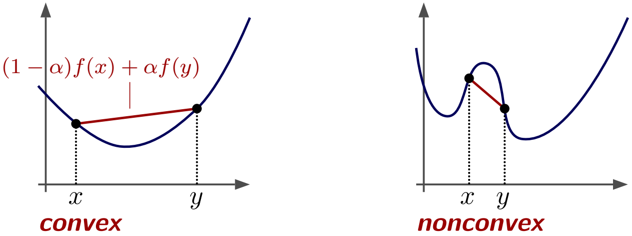
Definition:

f : dom $f \subseteq \mathbb{R}^n \to \mathbb{R}$ is *convex* when for any $x, y \in$ dom f,

$$f\big((1-\alpha)x + \alpha y\big) \le (1-\alpha)f(x) + \alpha f(y), \qquad \text{for all } 0 \le \alpha \le 1.$$

Convex functions

$$\underset{x}{\text{minimize}} \quad f(x) \quad \text{—— convex function}$$
$$\text{subject to} \quad x \in \Omega$$



Definition:

f : dom $f \subseteq \mathbb{R}^n \to \mathbb{R}$ is *convex* when for any $x, y \in$ dom f,

$$f\big((1-\alpha)x+\alpha y\big) \leq (1-\alpha)f(x)+\alpha f(y), \qquad \text{for all } 0 \leq \alpha \leq 1.$$

How to identify convex functions?

vocabulary $+$ **grammar**

How to identify convex functions?

vocabulary	+	**grammar**
definition		*operations preserving convexity*
differentiability conds.		
1D convexity		

Convexity under differentiability

Convexity under differentiability

$$f\big((1-\alpha)x + \alpha y\big) \leq (1-\alpha)f(x) + \alpha f(y)\,, \qquad \forall_{x,y\in \mathsf{dom}\, f}\,,\;\; \alpha \in [0,1]$$

Convexity under differentiability

$$f\big((1-\alpha)x + \alpha y\big) \leq (1-\alpha)f(x) + \alpha f(y)\,, \qquad \forall_{x,y\in\mathsf{dom}\,f}\,,\ \ \alpha \in [0,1]$$

Equivalent statements

Convexity under differentiability

$$f\big((1-\alpha)x + \alpha y\big) \leq (1-\alpha)f(x) + \alpha f(y)\,, \qquad \forall_{x,y \in \mathsf{dom}\, f}\,,\;\; \alpha \in [0,1]$$

Equivalent statements

- When f is differentiable,

Convexity under differentiability

$$f\big((1-\alpha)x+\alpha y\big) \le (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \textsf{dom}\, f}\,,\ \ \alpha\in[0,1]$$

Equivalent statements

- When f is differentiable,

$$f(y) \ge f(x)+\nabla f(x)^\top (y-x)\,, \quad \forall_{x,y\in \textsf{dom}\, f}$$

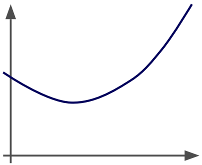
Convexity under differentiability

$$f\big((1-\alpha)x + \alpha y\big) \leq (1-\alpha)f(x) + \alpha f(y)\,, \qquad \forall_{x,y \in \mathsf{dom}\, f}\,,\ \ \alpha \in [0,1]$$

Equivalent statements

- When f is differentiable,

$$f(y) \geq f(x) + \nabla f(x)^\top (y-x)\,, \quad \forall_{x,y \in \mathsf{dom}\, f}$$



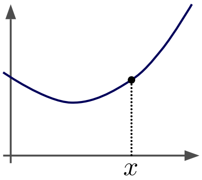
Convexity under differentiability

$$f\big((1-\alpha)x + \alpha y\big) \leq (1-\alpha) f(x) + \alpha f(y)\,, \qquad \forall_{x,y \in \mathsf{dom}\, f}\,,\;\; \alpha \in [0,1]$$

Equivalent statements

- When f is differentiable,

$$f(y) \geq f(x) + \nabla f(x)^{\top} (y - x)\,, \quad \forall_{x,y \in \mathsf{dom}\, f}$$



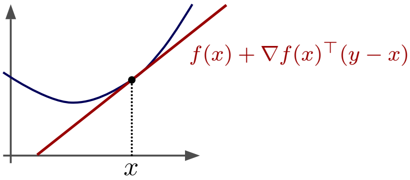
Convexity under differentiability

$$f\big((1-\alpha)x+\alpha y\big) \leq (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \mathsf{dom}\, f}\,,\ \ \alpha\in[0,1]$$

Equivalent statements

- When f is differentiable,

$$f(y) \geq f(x) + \nabla f(x)^\top (y-x)\,, \quad \forall_{x,y\in \mathsf{dom}\, f}$$



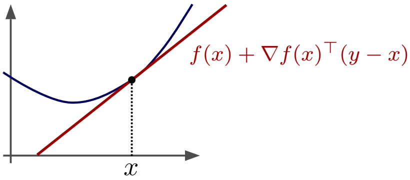
Convexity under differentiability

$$f\big((1-\alpha)x + \alpha y\big) \le (1-\alpha) f(x) + \alpha f(y)\,, \qquad \forall_{x,y \in \mathsf{dom}\, f}\,, \;\; \alpha \in [0,1]$$

Equivalent statements

- When f is differentiable,

$$f(y) \ge f(x) + \nabla f(x)^\top (y - x)\,, \quad \forall_{x,y \in \mathsf{dom}\, f}$$



- When f is twice-differentiable,

$$\nabla^2 f(x) \succeq 0\,, \quad \forall_{x \in \mathsf{dom}\, f}$$

Examples

Examples

Norms $f(x) = \|x\|$.

Examples

Norms $f(x) = \|x\|$.

Since for any x and y, and $0 \leq \alpha \leq 1$,

$$\|(1-\alpha)x + \alpha y\|$$

Examples

Norms $f(x) = \|x\|$.

Since for any x and y, and $0 \leq \alpha \leq 1$,

$$\|(1-\alpha)x + \alpha y\| \overset{\text{tr. ineq.}}{\leq} \|(1-\alpha)x\| + \|\alpha y\|$$

Examples

Norms $f(x) = \|x\|$.

Since for any x and y, and $0 \le \alpha \le 1$,

$$\|(1-\alpha)x + \alpha y\| \overset{\text{tr. ineq.}}{\le} \|(1-\alpha)x\| + \|\alpha y\| = (1-\alpha)\|x\| + \alpha\|y\|,$$

Examples

Norms $f(x) = \|x\|$.

Since for any x and y, and $0 \leq \alpha \leq 1$,

$$\|(1-\alpha)x + \alpha y\| \overset{\text{tr. ineq.}}{\leq} \|(1-\alpha)x\| + \|\alpha y\| = (1-\alpha)\|x\| + \alpha\|y\|,$$

all norms are convex.

Examples

Norms $f(x) = \|x\|$.

Since for any x and y, and $0 \leq \alpha \leq 1$,

$$\|(1-\alpha)x + \alpha y\| \overset{\text{tr. ineq.}}{\leq} \|(1-\alpha)x\| + \|\alpha y\| = (1-\alpha)\|x\| + \alpha\|y\|,$$

all norms are convex.

Exponential

Examples

Norms $\quad f(x) = \|x\|$.

Since for any x and y, and $0 \leq \alpha \leq 1$,

$$\|(1-\alpha)x + \alpha y\| \overset{\text{tr. ineq.}}{\leq} \|(1-\alpha)x\| + \|\alpha y\| = (1-\alpha)\|x\| + \alpha\|y\|\,,$$

all norms are convex.

Exponential $\quad f(x) = \exp(ax), \quad a \in \mathbb{R}$.

Examples

Norms $f(x) = \|x\|$.

Since for any x and y, and $0 \leq \alpha \leq 1$,

$$\|(1-\alpha)x + \alpha y\| \overset{\text{tr. ineq.}}{\leq} \|(1-\alpha)x\| + \|\alpha y\| = (1-\alpha)\|x\| + \alpha\|y\|,$$

all norms are convex.

Exponential $f(x) = \exp(ax), \quad a \in \mathbb{R}$.

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x)$$

Examples

Norms $\quad f(x) = \|x\|$.

Since for any x and y, and $0 \leq \alpha \leq 1$,

$$\|(1-\alpha)x + \alpha y\| \overset{\text{tr. ineq.}}{\leq} \|(1-\alpha)x\| + \|\alpha y\| = (1-\alpha)\|x\| + \alpha\|y\| ,$$

all norms are convex.

Exponential $\quad f(x) = \exp(ax), \quad a \in \mathbb{R}$.

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x) = a^2 \exp(ax) \geq 0$$

Examples

Norms $f(x) = \|x\|$.

Since for any x and y, and $0 \leq \alpha \leq 1$,

$$\|(1-\alpha)x + \alpha y\| \overset{\text{tr. ineq.}}{\leq} \|(1-\alpha)x\| + \|\alpha y\| = (1-\alpha)\|x\| + \alpha\|y\|,$$

all norms are convex.

Exponential $f(x) = \exp(ax), \quad a \in \mathbb{R}$.

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} f(x) = a^2 \exp(ax) \geq 0 \qquad \Longrightarrow \qquad f \text{ : convex}$$

Examples

Quadratic function $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \quad (A \succeq 0)$

Examples

Quadratic function $\quad f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \qquad (A \succeq 0)$

$$\nabla^2 f(x)$$

Examples

Quadratic function $\quad f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \qquad (A \succeq 0)$

$$\nabla^2 f(x) = A \succeq 0$$

Examples

Quadratic function $\quad f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \qquad (A \succeq 0)$

$$\nabla^2 f(x) = A \succeq 0 \qquad \Longrightarrow \qquad f : \text{convex}$$

Examples

Quadratic function $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \quad (A \succeq 0)$

$$\nabla^2 f(x) = A \succeq 0 \qquad \Longrightarrow \qquad f \text{ : convex}$$

Particular cases:

Examples

Quadratic function $\quad f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \quad (A \succeq 0)$

$$\nabla^2 f(x) = A \succeq 0 \qquad \Longrightarrow \qquad f \text{ : convex}$$

Particular cases:

$$A = I_n,\ b = 0_n,\ c = 0$$

Examples

Quadratic function $f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \quad (A \succeq 0)$

$$\nabla^2 f(x) = A \succeq 0 \qquad \Longrightarrow \qquad f \text{ : convex}$$

Particular cases:

$$A = I_n,\ b = 0_n,\ c = 0 \qquad \Longrightarrow \qquad \|x\|_2^2 \text{ : convex}$$

Examples

Quadratic function $\quad f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \qquad (A \succeq 0)$

$$\nabla^2 f(x) = A \succeq 0 \qquad \Longrightarrow \qquad f \text{ : convex}$$

Particular cases:

$$A = I_n,\ b = 0_n,\ c = 0 \qquad \Longrightarrow \qquad \|x\|_2^2 \text{ : convex}$$

$$A = 0_{n\times n}$$

Examples

Quadratic function $\quad f(x) = \frac{1}{2}x^\top Ax + b^\top x + c \qquad (A \succeq 0)$

$$\nabla^2 f(x) = A \succeq 0 \qquad \Longrightarrow \qquad f : \text{convex}$$

Particular cases:

$$A = I_n,\ b = 0_n,\ c = 0 \qquad \Longrightarrow \qquad \|x\|_2^2 : \text{ convex}$$

$$A = 0_{n\times n} \qquad \Longrightarrow \qquad b^\top x + c : \text{ convex}$$

Examples

For $X, W \in \mathbb{R}^{m \times n}$, $\quad X \mapsto \mathsf{tr}\big(W^\top X\big)$

Examples

For $X, W \in \mathbb{R}^{m\times n}$, $\quad X \mapsto \mathrm{tr}\big(W^\top X\big) \quad$ is *convex* (an inner product)

Examples

For $X, W \in \mathbb{R}^{m \times n}$, $\quad X \mapsto \mathrm{tr}\big(W^\top X\big)$ is *convex* (an inner product)

$$\mathrm{tr}\big(W^\top X\big) = \sum_{k=1}^{n} \big(W^\top X\big)_{kk}$$

Examples

For $X, W \in \mathbb{R}^{m\times n}$, $\quad X \mapsto \mathrm{tr}\big(W^\top X\big)$ is *convex* (an inner product)

$$\mathrm{tr}\big(W^\top X\big) = \sum_{k=1}^{n} \big(W^\top X\big)_{kk} = \sum_{k=1}^{n}\sum_{i=1}^{m} W_{ik}X_{ik}$$

Examples

For $X, W \in \mathbb{R}^{m\times n}$, $X \mapsto \mathrm{tr}\big(W^\top X\big)$ is *convex* (an inner product)

$$\mathrm{tr}\big(W^\top X\big) = \sum_{k=1}^{n} \big(W^\top X\big)_{kk} = \sum_{k=1}^{n}\sum_{i=1}^{m} W_{ik}X_{ik} = \begin{bmatrix} W_{11} \\ \vdots \\ W_{m1} \\ \vdots \\ W_{1n} \\ \vdots \\ W_{mn} \end{bmatrix}^\top \begin{bmatrix} X_{11} \\ \vdots \\ X_{m1} \\ \vdots \\ X_{1n} \\ \vdots \\ X_{mn} \end{bmatrix}$$

Examples

For $X, W \in \mathbb{R}^{m \times n}$, $\quad X \mapsto \mathrm{tr}\big(W^\top X\big)$ is *convex* (an inner product)

$$\mathrm{tr}\big(W^\top X\big) = \sum_{k=1}^{n} \big(W^\top X\big)_{kk} = \sum_{k=1}^{n} \sum_{i=1}^{m} W_{ik} X_{ik} = \begin{bmatrix} W_{11} \\ \vdots \\ W_{m1} \\ \vdots \\ W_{1n} \\ \vdots \\ W_{mn} \end{bmatrix}^\top \begin{bmatrix} X_{11} \\ \vdots \\ X_{m1} \\ \vdots \\ X_{1n} \\ \vdots \\ X_{mn} \end{bmatrix}$$

$$= \mathrm{vec}(W)^\top \mathrm{vec}(X)$$

Equivalent definitions of convexity

Equivalent definitions of convexity

$$f\big((1-\alpha)x + \alpha y\big) \leq (1-\alpha)f(x) + \alpha f(y)\,, \qquad \forall_{x,y \in \mathbf{dom}\, f}\,, \;\; \alpha \in [0,1]$$

Equivalent definitions of convexity

$$f\big((1-\alpha)x+\alpha y\big) \leq (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \mathsf{dom}\, f}\,, \;\; \alpha\in[0,1]$$

Equivalent statements

Equivalent definitions of convexity

$$f\big((1-\alpha)x + \alpha y\big) \le (1-\alpha)f(x) + \alpha f(y)\,, \qquad \forall_{x,y \in \mathbf{dom}\, f}\,, \;\; \alpha \in [0,1]$$

Equivalent statements

- $g(t) = f(x+ty) \quad (g : \mathbb{R} \to \mathbb{R})$ is convex for all $x + ty \in \mathbf{dom}\, f$

Equivalent definitions of convexity

$$f\big((1-\alpha)x+\alpha y\big) \le (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \mathsf{dom}\, f}\,,\;\; \alpha\in[0,1]$$

Equivalent statements

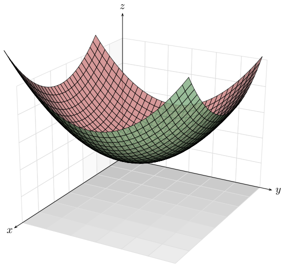
- $g(t) = f(x+ty) \quad (g:\mathbb{R}\to\mathbb{R})$ is convex for all $x+ty \in$ dom f
- Sublevel sets $S_\alpha := \{x\,:\,f(x)\le\alpha\}$ are convex for all $\alpha\in\mathbb{R}$

Equivalent definitions of convexity

$$f\big((1-\alpha)x+\alpha y\big) \le (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \mathsf{dom}\, f}\,,\;\; \alpha\in[0,1]$$

Equivalent statements

- $g(t) = f(x+ty) \quad (g:\mathbb{R}\to\mathbb{R})$ is convex for all $x+ty\in$ dom f
- Sublevel sets $S_\alpha := \{x \,:\, f(x)\le\alpha\}$ are convex for all $\alpha\in\mathbb{R}$

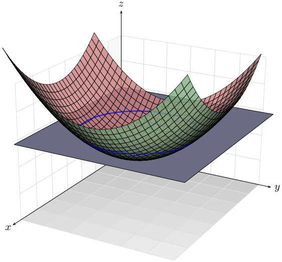


Equivalent definitions of convexity

$$f\big((1-\alpha)x+\alpha y\big) \le (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \mathsf{dom}\, f}\,,\;\; \alpha\in[0,1]$$

Equivalent statements

- $g(t) = f(x+ty) \quad (g:\mathbb{R}\to\mathbb{R})$ is convex for all $x+ty \in$ dom f
- Sublevel sets $S_\alpha := \{x \,:\, f(x)\le\alpha\}$ are convex for all $\alpha\in\mathbb{R}$

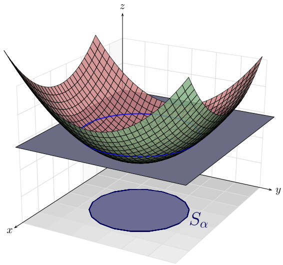


Equivalent definitions of convexity

$$f\big((1-\alpha)x+\alpha y\big) \le (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \mathsf{dom}\, f}\,,\ \ \alpha\in[0,1]$$

Equivalent statements

- $g(t) = f(x+ty) \quad (g : \mathbb{R} \to \mathbb{R})$ is convex for all $x+ty \in$ dom f
- Sublevel sets $S_\alpha := \{x \,:\, f(x) \le \alpha\}$ are convex for all $\alpha \in \mathbb{R}$

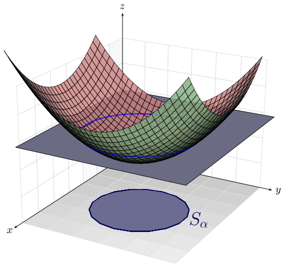


Equivalent definitions of convexity

$$f\big((1-\alpha)x+\alpha y\big) \le (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \mathrm{dom}\, f}\,,\;\; \alpha\in[0,1]$$

Equivalent statements

- $g(t) = f(x+ty) \quad (g:\mathbb{R}\to\mathbb{R})$ is convex for all $x+ty \in \mathrm{dom}\, f$
- Sublevel sets $S_\alpha := \{x \,:\, f(x)\le\alpha\}$ are convex for all $\alpha\in\mathbb{R}$
- Epigraph $\mathrm{epi}\, f := \{(x,t) \,:\, f(x)\le t\}$ is convex

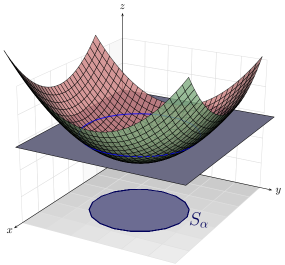


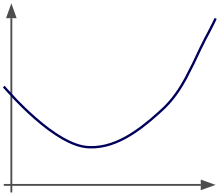
Equivalent definitions of convexity

$$f\big((1-\alpha)x+\alpha y\big) \le (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in\mathsf{dom}\,f}\,,\;\; \alpha\in[0,1]$$

Equivalent statements

- $g(t) = f(x+ty) \quad (g:\mathbb{R}\to\mathbb{R})$ is convex for all $x+ty \in$ dom f
- Sublevel sets $S_\alpha := \{x \,:\, f(x)\le\alpha\}$ are convex for all $\alpha\in\mathbb{R}$
- Epigraph epi $f := \{(x,t) \,:\, f(x)\le t\}$ is convex



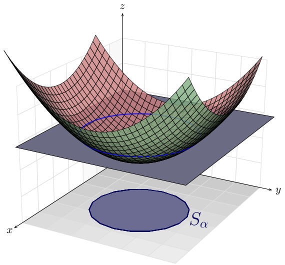


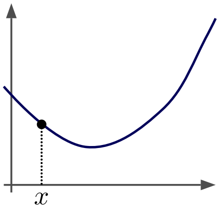
Equivalent definitions of convexity

$$f\big((1-\alpha)x+\alpha y\big) \le (1-\alpha)f(x)+\alpha f(y)\,, \qquad \forall_{x,y\in \mathsf{dom}\, f}\,, \;\; \alpha\in[0,1]$$

Equivalent statements

- $g(t) = f(x+ty) \quad (g:\mathbb{R}\to\mathbb{R})$ is convex for all $x+ty \in \mathsf{dom}\, f$
- Sublevel sets $S_\alpha := \{x \,:\, f(x)\le\alpha\}$ are convex for all $\alpha\in\mathbb{R}$
- Epigraph $\mathsf{epi}\, f := \{(x,t) \,:\, f(x)\le t\}$ is convex



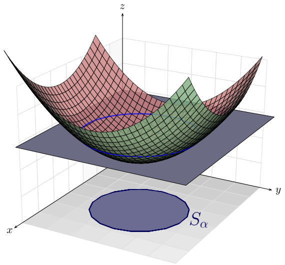


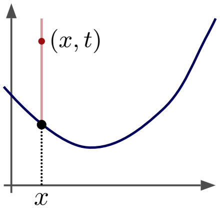
Equivalent definitions of convexity

$$f\big((1-\alpha)x + \alpha y\big) \le (1-\alpha) f(x) + \alpha f(y)\,, \qquad \forall_{x,y \in \mathsf{dom}\, f}\,, \;\; \alpha \in [0,1]$$

Equivalent statements

- $g(t) = f(x + ty) \quad (g : \mathbb{R} \to \mathbb{R})$ is convex for all $x + ty \in \mathsf{dom}\, f$
- Sublevel sets $S_\alpha := \{x \,:\, f(x) \le \alpha\}$ are convex for all $\alpha \in \mathbb{R}$
- Epigraph $\mathsf{epi}\, f := \{(x,t) \,:\, f(x) \le t\}$ is convex



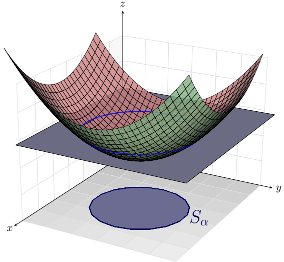


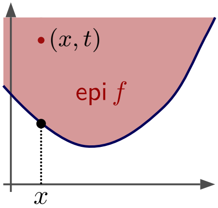
Equivalent definitions of convexity

$$f\big((1-\alpha)x + \alpha y\big) \le (1-\alpha)f(x) + \alpha f(y)\,, \qquad \forall_{x,y\in \text{dom}\, f}\,,\ \ \alpha \in [0,1]$$

Equivalent statements

- $g(t) = f(x+ty) \quad (g : \mathbb{R} \to \mathbb{R})$ is convex for all $x + ty \in \text{dom}\, f$
- Sublevel sets $S_\alpha := \{x \,:\, f(x) \le \alpha\}$ are convex for all $\alpha \in \mathbb{R}$
- Epigraph epi $f := \{(x,t) \,:\, f(x) \le t\}$ is convex





How to identify convex functions?

vocabulary + **grammar**

definition

differentiability conds.

1D convexity

operations preserving convexity

How to identify convex functions?

vocabulary + **grammar**

definition

differentiability conds.

1D convexity

operations preserving convexity

Operations preserving convexity

Operations preserving convexity

Nonnegative weighted sums

Operations preserving convexity

Nonnegative weighted sums

$$\begin{cases} f_1, \dots, f_m \ : \text{convex} \\ w_1, \dots, w_m \geq 0 \end{cases}$$

Operations preserving convexity

Nonnegative weighted sums

$$\begin{cases} f_1, \ldots, f_m \ : \text{convex} \\ w_1, \ldots, w_m \geq 0 \end{cases} \qquad \Longrightarrow \qquad w_1 f_1 + \cdots + w_m f_m \ : \text{convex}$$

Operations preserving convexity

Nonnegative weighted sums

$$\left\{ \begin{array}{l} f_1, \ldots, f_m \ : \ \text{convex} \\ w_1, \ldots, w_m \geq 0 \end{array} \right. \qquad \Longrightarrow \qquad w_1 f_1 + \cdots + w_m f_m \ : \ \text{convex}$$

Precomposition with affine maps

Operations preserving convexity

Nonnegative weighted sums

$$\begin{cases} f_1, \ldots, f_m \ : \text{ convex} \\ w_1, \ldots, w_m \geq 0 \end{cases} \quad \Longrightarrow \quad w_1 f_1 + \cdots + w_m f_m \ : \text{ convex}$$

Precomposition with affine maps

$$f \ : \text{ convex}$$

Operations preserving convexity

Nonnegative weighted sums

$$\left\{\begin{array}{l} f_1, \ldots, f_m \text{ : convex} \\ w_1, \ldots, w_m \geq 0 \end{array}\right. \quad \Longrightarrow \quad w_1 f_1 + \cdots + w_m f_m \text{ : convex}$$

Precomposition with affine maps

$$f \text{ : convex} \quad \Longrightarrow \quad g(x) = f(Ax + b) \text{ : convex}$$

Operations preserving convexity

Nonnegative weighted sums

$$\begin{cases} f_1, \ldots, f_m \ : \text{convex} \\ w_1, \ldots, w_m \geq 0 \end{cases} \quad \Longrightarrow \quad w_1 f_1 + \cdots + w_m f_m \ : \text{convex}$$

Precomposition with affine maps

$$f \ : \text{convex} \quad \Longrightarrow \quad g(x) = f(Ax + b) \ : \text{convex}$$

Pointwise maximum/supremum

Operations preserving convexity

Nonnegative weighted sums

$$\left\{\begin{array}{l} f_1, \ldots, f_m \ : \text{convex} \\ w_1, \ldots, w_m \geq 0 \end{array}\right. \qquad \Longrightarrow \qquad w_1 f_1 + \cdots + w_m f_m \ : \text{convex}$$

Precomposition with affine maps

$$f \ : \text{convex} \qquad \Longrightarrow \qquad g(x) = f(Ax + b) \ : \text{convex}$$

Pointwise maximum/supremum

$$f_i \ : \text{convex}, \quad \text{for } i \in I$$

Operations preserving convexity

Nonnegative weighted sums

$$\left\{ \begin{array}{l} f_1, \ldots, f_m \ : \ \text{convex} \\ w_1, \ldots, w_m \geq 0 \end{array} \right. \qquad \Longrightarrow \qquad w_1 f_1 + \cdots + w_m f_m \ : \ \text{convex}$$

Precomposition with affine maps

$$f \ : \ \text{convex} \qquad \Longrightarrow \qquad g(x) = f(Ax + b) \ : \ \text{convex}$$

Pointwise maximum/supremum

$$f_i \ : \ \text{convex}, \quad \text{for } i \in I \qquad \Longrightarrow \qquad \sup_{i \in I} f_i \ : \ \text{convex}$$

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

$$\begin{aligned} \lambda_{\max}(A) \ = \ & \max_{v} \quad v^\top A v \\ & \text{s.t.} \quad \|v\|_2 = 1 \end{aligned}$$

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

$$\lambda_{\max}(A) \;=\; \begin{array}{ll} \max\limits_{v} & v^\top A v \\ \text{s.t.} & \|v\|_2 = 1 \end{array} \;=\; \max_{v \in V} \; f_v(A) \,,$$

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

$$\lambda_{\max}(A) \;=\; \begin{array}{ll} \max\limits_{v} & v^\top A v \\ \text{s.t.} & \|v\|_2 = 1 \end{array} \;=\; \max_{v \in V} \; f_v(A)\,,$$

where

$$V = \{v \,:\, \|v\|_2 = 1\}$$

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

$$\lambda_{\max}(A) = \begin{array}{ll} \max\limits_{v} & v^\top A v \\ \text{s.t.} & \|v\|_2 = 1 \end{array} = \max_{v \in V} f_v(A) \,,$$

where

$$V = \{v : \|v\|_2 = 1\}$$

$$f_v(A) = v^\top A v$$

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

$$\lambda_{\max}(A) = \begin{array}{ll} \max\limits_{v} & v^\top A v \\ \text{s.t.} & \|v\|_2 = 1 \end{array} = \max_{v \in V} f_v(A) \,,$$

where

$$V = \{v \,:\, \|v\|_2 = 1\}$$

$$f_v(A) = v^\top A v = \mathrm{tr}\big(v^\top A v\big)$$

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

$$\lambda_{\max}(A) = \begin{array}{ll} \max\limits_{v} & v^\top A v \\ \text{s.t.} & \|v\|_2 = 1 \end{array} = \max_{v \in V} f_v(A) \,,$$

where

$$V = \{v \,:\, \|v\|_2 = 1\}$$

$$f_v(A) = v^\top A v = \mathrm{tr}\big(v^\top A v\big) = \mathrm{tr}\big(v v^\top A\big)$$

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

$$\lambda_{\max}(A) = \begin{array}{ll} \max\limits_{v} & v^\top A v \\ \text{s.t.} & \|v\|_2 = 1 \end{array} = \max_{v \in V} f_v(A) \,,$$

where

$$V = \{v \,:\, \|v\|_2 = 1\}$$

$$f_v(A) = v^\top A v = \text{tr}\big(v^\top A v\big) = \text{tr}\big(v v^\top A\big) = \text{tr}\big((v v^\top)^\top A\big)$$

Operations preserving convexity

The set I is arbitrary: can even be uncountable

Example: Largest eigenvalue of a matrix $\lambda_{\max}(A)$ is convex

$$\lambda_{\max}(A) = \begin{array}{ll} \max\limits_{v} & v^\top A v \\ \text{s.t.} & \|v\|_2 = 1 \end{array} = \max_{v \in V} f_v(A) \ ,$$

where

$$V = \{v : \|v\|_2 = 1\}$$

$$f_v(A) = v^\top A v = \text{tr}\big(v^\top A v\big) = \text{tr}\big(v v^\top A\big) = \text{tr}\big((v v^\top)^\top A\big) \ : \quad \text{convex}$$

Examples

Examples

- $f(x) = \left\|y - Ax\right\|_2^2$

Examples

- $f(x) = \left\|y - Ax\right\|_2^2 = g(h(x))$

Examples

- $f(x) = \left\|y - Ax\right\|_2^2 = g(h(x))$

$$h(x) = -Ax + y$$

Examples

- $f(x) = \|y - Ax\|_2^2 = g(h(x))$

$$h(x) = -Ax + y \qquad\qquad \textit{affine}$$

Examples

- $f(x) = \left\|y - Ax\right\|_2^2 = g(h(x))$

$$h(x) = -Ax + y \qquad \textit{affine}$$

$$g(z) = \|z\|_2^2$$

Examples

- $f(x) = \left\|y - Ax\right\|_2^2 = g(h(x))$

$$h(x) = -Ax + y \qquad \textit{affine}$$

$$g(z) = \|z\|_2^2 \qquad \textit{convex}$$

Examples

- $f(x) = \left\|y - Ax\right\|_2^2 = g(h(x))$: *convex*

$$h(x) = -Ax + y \qquad \textit{affine}$$

$$g(z) = \|z\|_2^2 \qquad \textit{convex}$$

Examples

- $f(x) = \|y - Ax\|_2^2 = g(h(x))$: *convex*

$$h(x) = -Ax + y \qquad \textit{affine}$$

$$g(z) = \|z\|_2^2 \qquad \textit{convex}$$

- $f(x) = \|y - Ax\|_1$

Examples

- $f(x) = \left\|y - Ax\right\|_2^2 = g(h(x)) \quad : \quad$ *convex*

$$h(x) = -Ax + y \qquad \textit{affine}$$

$$g(z) = \|z\|_2^2 \qquad \textit{convex}$$

- $f(x) = \left\|y - Ax\right\|_1 = g(h(x)) : \quad$ *convex*

$$h(x) = -Ax + y \qquad \textit{affine}$$

$$g(z) = \|z\|_1 \qquad \textit{convex}$$

Outline

Convex sets

Identifying convex sets

Examples: geometrical sets and filter design constraints

Convex functions

Identifying convex functions

Relation to convex sets

Optimization problems

Convex problems, properties, and problem manipulation

Examples and solvers

Statistical estimation

Maximum likelihood & maximum a posteriori

Nonparametric estimation

Hypothesis testing & optimal detection

Convex optimization problems

Convex optimization problems

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0\,, \quad i = 1, \ldots, m \\ & h_i(x) = 0\,, \quad i = 1, \ldots, p \end{aligned}$$

Convex optimization problems

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \le 0\,, \quad i = 1, \ldots, m \\ & h_i(x) = 0\,, \quad i = 1, \ldots, p \end{aligned}$$

convex

Convex optimization problems

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) && \textit{convex} \\ \text{subject to} \quad & g_i(x) \leq 0\,, \quad i = 1, \dots, m && \textit{convex} \\ & h_i(x) = 0\,, \quad i = 1, \dots, p \end{aligned}$$

Convex optimization problems

$$
\begin{array}{lll}
\underset{x}{\text{minimize}} & f(x) & \textit{convex} \\
\text{subject to} & g_i(x) \leq 0\,, \quad i = 1, \ldots, m & \textit{convex} \\
& h_i(x) = 0\,, \quad i = 1, \ldots, p & \textit{affine}
\end{array}
$$

Convex optimization problems

$$\begin{array}{lll} \underset{x}{\text{minimize}} & f(x) & \textit{convex} \\ \text{subject to} & g_i(x) \leq 0\,, \quad i = 1, \ldots, m & \textit{convex} \\ & h_i(x) = 0\,, \quad i = 1, \ldots, p & \textit{affine} \end{array}$$

Some notation

Convex optimization problems

$$
\begin{array}{lll}
\underset{x}{\text{minimize}} & f(x) & \textit{convex} \\
\text{subject to} & g_i(x) \leq 0\,, \quad i = 1, \ldots, m & \textit{convex} \\
 & h_i(x) = 0\,, \quad i = 1, \ldots, p & \textit{affine}
\end{array}
$$

Some notation

Optimal value: $$p^\star = \begin{array}{ll} \inf_x & f(x) \\ \text{s.t.} & x \in \Omega \end{array}$$

Convex optimization problems

$$\begin{array}{lll} \underset{x}{\text{minimize}} & f(x) & \text{---------- } \textit{convex} \\ \text{subject to} & g_i(x) \leq 0\,, \quad i = 1, \ldots, m & \textit{convex} \\ & h_i(x) = 0\,, \quad i = 1, \ldots, p & \textit{affine} \end{array}$$

Some notation

Optimal value: $$p^\star = \begin{array}{ll} \inf\limits_{x} & f(x) \\ \text{s.t.} & x \in \Omega \end{array} \quad \in [-\infty, +\infty]$$

Convex optimization problems

$$
\begin{array}{lll}
\underset{x}{\text{minimize}} & f(x) & \textit{convex} \\
\text{subject to} & g_i(x) \leq 0\,, \quad i = 1, \ldots, m & \textit{convex} \\
 & h_i(x) = 0\,, \quad i = 1, \ldots, p & \textit{affine}
\end{array}
$$

Some notation

unbounded

Optimal value: $$p^\star = \begin{array}{ll} \inf\limits_{x} & f(x) \\ \text{s.t.} & x \in \Omega \end{array} \quad \in [-\infty, +\infty]$$

Convex optimization problems

$$\begin{array}{lll} \underset{x}{\text{minimize}} & f(x) & \text{—— } \textit{convex} \\ \text{subject to} & g_i(x) \le 0\,, \quad i = 1, \dots, m & \textit{convex} \\ & h_i(x) = 0\,, \quad i = 1, \dots, p & \textit{affine} \end{array}$$

Some notation

unbounded

Optimal value: $$p^\star = \inf_{x} \; f(x) \quad \in [-\infty, +\infty]$$
$$\text{s.t.} \quad x \in \Omega$$

infeasible

Convex optimization problems

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) && \textit{convex} \\ \text{subject to} \quad & g_i(x) \leq 0\,, \quad i = 1, \ldots, m && \textit{convex} \\ & h_i(x) = 0\,, \quad i = 1, \ldots, p && \textit{affine} \end{aligned}$$

Some notation

Optimal value: $$p^\star = \inf_x \; f(x) \quad \text{s.t.} \quad x \in \Omega \qquad \in [-\infty, +\infty]$$

unbounded ($-\infty$), *infeasible* ($+\infty$)

A minimizer: $$x^\star \in \operatorname*{argmin}_x \; f(x) \quad \text{s.t.} \quad x \in \Omega$$

Convex optimization problems

Convex optimization problems

Theorem

In convex problems, a *local* minimizer is always a *global* minimizer.

Convex optimization problems

Theorem

In convex problems, a *local* minimizer is always a *global* minimizer.

Proof (for unconstrained problems w/ differentiable objective):

Convex optimization problems

Theorem

In convex problems, a *local* minimizer is always a *global* minimizer.

Proof (for unconstrained problems w/ differentiable objective):

Recall that for any $x, y \in \mathrm{dom}\, f$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

Convex optimization problems

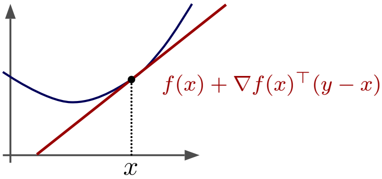
Theorem

In convex problems, a *local* minimizer is always a *global* minimizer.

Proof (for unconstrained problems w/ differentiable objective):

Recall that for any $x, y \in \mathbf{dom}\, f$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$



Convex optimization problems

Theorem

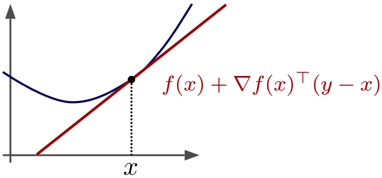
In convex problems, a *local* minimizer is always a *global* minimizer.

Proof (for unconstrained problems w/ differentiable objective):

Recall that for any $x, y \in \mathbf{dom}\, f$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

$x^\star$: local minimizer $\Rightarrow \nabla f(x^\star) = 0$



Convex optimization problems

Theorem

In convex problems, a *local* minimizer is always a *global* minimizer.

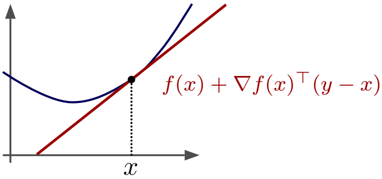
Proof (for unconstrained problems w/ differentiable objective):

Recall that for any $x, y \in \mathbf{dom}\, f$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

$x^\star$: local minimizer $\Rightarrow \nabla f(x^\star) = 0$

Therefore, $f(y) \geq f(x^\star)$, for all y.



Convex optimization problems

Theorem

In convex problems, a *local* minimizer is always a *global* minimizer.

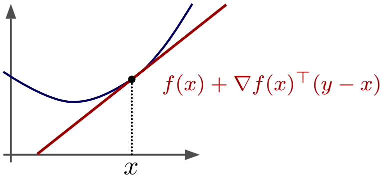
Proof (for unconstrained problems w/ differentiable objective):

Recall that for any $x, y \in \mathbf{dom}\, f$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x)$$

$x^\star$: local minimizer $\Rightarrow \nabla f(x^\star) = 0$

Therefore, $f(y) \geq f(x^\star)$, for all y.



$\square$

Equivalence between optimization problems

Equivalence between optimization problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in X \end{array} \qquad \text{(P1)}$$

$$\begin{array}{ll} \underset{y}{\text{minimize}} & g(y) \\ \text{subject to} & y \in Y \end{array} \qquad \text{(P2)}$$

Equivalence between optimization problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in X \end{array} \tag{P1}$$

$$\begin{array}{ll} \underset{y}{\text{minimize}} & g(y) \\ \text{subject to} & y \in Y \end{array} \tag{P2}$$

(P1) and (P2) are *equivalent* when

Equivalence between optimization problems

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x \in X \end{array} \tag{P1}$$

$$\begin{array}{ll} \underset{y}{\text{minimize}} & g(y) \\ \text{subject to} & y \in Y \end{array} \tag{P2}$$

(P1) and (P2) are *equivalent* when

- Given a solution $x^\star$ of (P1) we can obtain a solution $y^\star$ of (P2)
- Given a solution $y^\star$ of (P2) we can obtain a solution $x^\star$ of (P1)

Examples

Examples

$$\underset{x}{\text{maximize}} \quad f(x)$$

Examples

$$\underset{x}{\text{maximize}} \quad f(x) \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad -f(x)$$

Examples

$$\underset{x}{\text{maximize}} \quad f(x) \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad -f(x)$$

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x > 0 \end{array}$$

Examples

$$\underset{x}{\text{maximize}} \quad f(x) \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad -f(x)$$

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x > 0 \end{array} \qquad \Longleftrightarrow \qquad \underset{y}{\text{minimize}} \quad f\left(e^y\right)$$

Examples

$$\underset{x}{\text{maximize}} \quad f(x) \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad -f(x)$$

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x > 0 \end{array} \qquad \Longleftrightarrow \qquad \underset{y}{\text{minimize}} \quad f\left(e^y\right)$$

$$\underset{x}{\text{minimize}} \quad f(x)$$

Examples

$$\underset{x}{\text{maximize}} \quad f(x) \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad -f(x)$$

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & x > 0 \end{array} \qquad \Longleftrightarrow \qquad \underset{y}{\text{minimize}} \quad f\left(e^y\right)$$

$$\underset{x}{\text{minimize}} \quad f(x) \qquad \Longleftrightarrow \qquad \begin{array}{ll} \underset{x,t}{\text{minimize}} & t \\ \text{subject to} & f(x) \leq t \end{array}$$

Examples

Examples

$$\underset{x}{\text{minimize}} \quad a\, f(x) + b$$

Examples

$$\underset{x}{\text{minimize}} \quad \underbrace{a}_{>0} f(x) + b$$

Examples

$$\underset{x}{\text{minimize}} \quad \underbrace{a}_{>0} f(x) + b \quad \Longleftrightarrow \quad \underset{x}{\text{minimize}} \quad f(x)$$

Examples

$$\underset{x}{\text{minimize}} \quad \underbrace{a}_{>0} f(x) + b \quad \Longleftrightarrow \quad \underset{x}{\text{minimize}} \quad f(x)$$

$$\underset{x}{\text{minimize}} \quad f(x)$$

Examples

$$\underset{x}{\text{minimize}} \quad a\,f(x) + b \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad f(x)$$

$a > 0$

$$\underset{x}{\text{minimize}} \quad f(x)$$

f convex

Examples

$$\underset{x}{\text{minimize}} \quad \underbrace{a}_{>0} f(x) + b \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad f(x)$$

$$\underset{x}{\text{minimize}} \quad \underbrace{f(x)}_{\text{convex}} \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad g\big(f(x)\big)$$

Examples

$$\underset{x}{\text{minimize}} \quad a\, f(x) + b \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad f(x)$$

($a > 0$)

$$\underset{x}{\text{minimize}} \quad f(x) \qquad \Longleftrightarrow \qquad \underset{x}{\text{minimize}} \quad g\big(f(x)\big)$$

($f(x)$: convex)

if $\text{im}\, f \subseteq \text{dom}\, g$

and $g \circ f$: convex

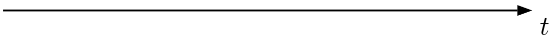
Air Traffic Control

Air Traffic Control

- n airplanes land in order $1, 2, \ldots, n$
- t_i: arrival time of airplane i
- Airplane i has to land in interval $[m_i, M_i]$

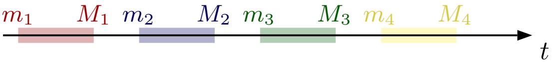
Air Traffic Control

- n airplanes land in order $1, 2, \ldots, n$
- t_i: arrival time of airplane i
- Airplane i has to land in interval $[m_i,\, M_i]$



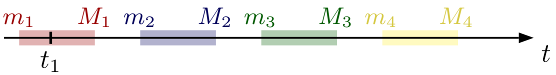
Air Traffic Control

- n airplanes land in order $1, 2, \ldots, n$
- t_i: arrival time of airplane i
- Airplane i has to land in interval $[m_i, M_i]$



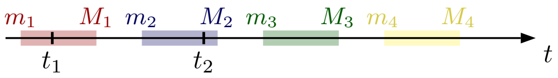
Air Traffic Control

- n airplanes land in order $1, 2, \ldots, n$
- t_i: arrival time of airplane i
- Airplane i has to land in interval $[m_i, M_i]$



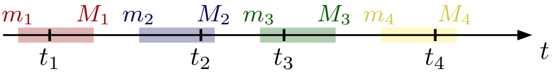
Air Traffic Control

- n airplanes land in order $1, 2, \ldots, n$
- t_i: arrival time of airplane i
- Airplane i has to land in interval $[m_i, M_i]$



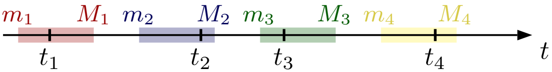
Air Traffic Control

- n airplanes land in order $1, 2, \ldots, n$
- t_i: arrival time of airplane i
- Airplane i has to land in interval $[m_i, M_i]$



Air Traffic Control

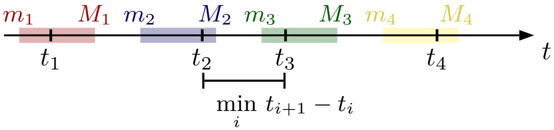
- n airplanes land in order $1, 2, \ldots, n$
- t_i: arrival time of airplane i
- Airplane i has to land in interval $[m_i, M_i]$



Goal: compute $t_1, \ldots, t_n$ that maximize $\min_i \; t_{i+1} - t_i$

Air Traffic Control

- n airplanes land in order $1, 2, \ldots, n$
- t_i: arrival time of airplane i
- Airplane i has to land in interval $[m_i,\, M_i]$



Goal: compute t_1, ..., t_n that maximize $\min_i\ t_{i+1} - t_i$

Air Traffic Control

$$\begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{maximize}} & \min\Big\{t_2 - t_1\,,\ t_3 - t_2\,,\ \ldots\,,\ t_n - t_{n-1}\Big\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1, \ldots, n \end{array}$$

Air Traffic Control

$$\begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{maximize}} & \min\left\{t_2 - t_1\,,\ t_3 - t_2\,,\ \ldots\,,\ t_n - t_{n-1}\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

$$\iff \quad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & -\min\left\{t_2 - t_1\,,\ t_3 - t_2\,,\ \ldots\,,\ t_n - t_{n-1}\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

Air Traffic Control

$$\begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{maximize}} & \min\left\{t_2 - t_1\,,\ t_3 - t_2\,,\ \ldots\,,\ t_n - t_{n-1}\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

$$\iff \quad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & -\min\left\{t_2 - t_1\,,\ t_3 - t_2\,,\ \ldots\,,\ t_n - t_{n-1}\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

$$\iff \quad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \max\left\{t_1 - t_2\,,\ t_2 - t_3\,,\ \ldots\,,\ t_{n-1} - t_n\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

Air Traffic Control

$$\begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{maximize}} & \min\left\{t_2 - t_1\,,\ t_3 - t_2\,,\ \ldots\,,\ t_n - t_{n-1}\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

$$\iff \quad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & -\min\left\{t_2 - t_1\,,\ t_3 - t_2\,,\ \ldots\,,\ t_n - t_{n-1}\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

$$\iff \quad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \max\left\{t_1 - t_2\,,\ t_2 - t_3\,,\ \ldots\,,\ t_{n-1} - t_n\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

Convex

Air Traffic Control

$$\begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \max\left\{t_1 - t_2\,,\ t_2 - t_3\,,\ \ldots\,,\ t_{n-1} - t_n\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

Air Traffic Control

$$\begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \max\left\{t_1 - t_2\,,\ t_2 - t_3\,,\ \ldots\,,\ t_{n-1} - t_n\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

$$\iff \quad \begin{array}{ll} \underset{t_1,\ldots,t_n,s}{\text{minimize}} & s \\ \text{subject to} & \max\left\{t_1 - t_2\,,\ t_2 - t_3\,,\ \ldots\,,\ t_{n-1} - t_n\right\} \le s \\ & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

Air Traffic Control

$$\begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \max\left\{t_1 - t_2\,,\ t_2 - t_3\,,\ \ldots\,,\ t_{n-1} - t_n\right\} \\ \text{subject to} & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

$$\iff \quad \begin{array}{ll} \underset{t_1,\ldots,t_n,s}{\text{minimize}} & s \\ \text{subject to} & \max\left\{t_1 - t_2\,,\ t_2 - t_3\,,\ \ldots\,,\ t_{n-1} - t_n\right\} \le s \\ & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

$$\iff \quad \begin{array}{ll} \underset{t_1,\ldots,t_n,s}{\text{minimize}} & s \\ \text{subject to} & t_1 - t_2 \le s \\ & \quad\vdots \\ & t_{n-1} - t_n \le s \\ & m_i \le t_i \le M_i\,, \quad i = 1,\ldots,n \end{array}$$

Air Traffic Control

This is a linear program (LP):

Air Traffic Control

This is a linear program (LP):

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^\top x \\ \text{subject to} & Ax \le b \end{array}$$

Air Traffic Control

This is a linear program (LP):

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & c^\top x \\ \text{subject to} \quad & Ax \leq b \end{aligned}$$

with $x = (t_1, \ldots, t_n, s) \in \mathbb{R}^{n+1}, \; c = (0, \cdots, 0, 1)$

Air Traffic Control

This is a linear program (LP):

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^\top x \\ \text{subject to} & Ax \le b \end{array}$$

with $x = (t_1, \ldots, t_n, s) \in \mathbb{R}^{n+1}$, $c = (0, \cdots, 0, 1)$, and

$$A = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & -1 \\ 0 & 1 & -1 & \cdots & 0 & 0 & -1 \\ & & & \ddots & & & \\ 0 & 0 & 0 & \cdots & 1 & -1 & -1 \\ -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & & & \ddots & & & \\ 0 & 0 & 0 & \cdots & 0 & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -m_1 \\ M_1 \\ \vdots \\ -m_n \\ M_n \end{bmatrix}$$

Air Traffic Control

Can be solved, e.g., with MATLAB's linprog solver

Air Traffic Control

Can be solved, e.g., with MATLAB's linprog solver

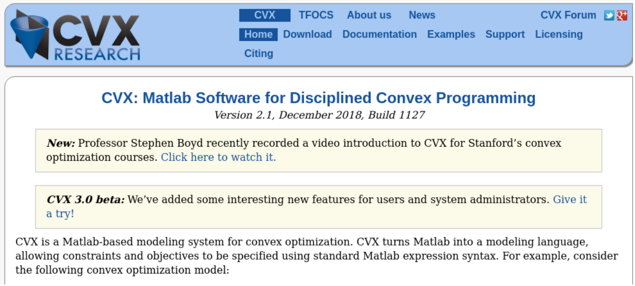
But we have to *explicitly construct* A, b, and c . . .

Air Traffic Control

Can be solved, e.g., with MATLAB's linprog solver

But we have to *explicitly construct* A, b, and c . . .

CVX (cvxr.com/cvx) manipulates and solves *convex* problems



Air Traffic Control

```
cvx_begin
    variables t1 t2 t3 t4 t5;
    maximize( min([t2-t1, t3-t2, t4-t3, t5-t4]) );
    subject to
        1 <= t1 <= 2;
        3 <= t2 <= 4;
        5 <= t3 <= 6;
        7 <= t4 <= 8;
        9 <= t5 <= 10;
cvx_end
```

Air Traffic Control

```
cvx_begin
    variables t1 t2 t3 t4 t5;
    maximize( min([t2-t1, t3-t2, t4-t3, t5-t4]) );
    subject to
        1 <= t1 <= 2;
        3 <= t2 <= 4;
        5 <= t3 <= 6;
        7 <= t4 <= 8;
        9 <= t5 <= 10;
cvx_end
```

$$(t_1^\star\,,\ t_2^\star\,,\ t_3^\star\,,\ t_4^\star\,,\ t_5^\star) = (1\,,\ 3.25\,,\ 5.5\,,\ 7.75\,,\ 10)$$

Portfolio Optimization

Portfolio Optimization

- £T to invest in n assets

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad \qquad \Sigma = \mathbb{E}\big[(r - \mu)(r - \mu)^\top\big]$$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \dots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \dots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad \qquad \Sigma = \mathbb{E}\big[(r-\mu)(r-\mu)^\top\big]$$

- An investment $x = (x_1, \dots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad \Sigma = \mathbb{E}\big[(r-\mu)(r-\mu)^\top\big]$$

- An investment $x = (x_1, \ldots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

$$\mathbb{E}\big[s(x)\big]$$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad \Sigma = \mathbb{E}\big[(r - \mu)(r - \mu)^\top\big]$$

- An investment $x = (x_1, \ldots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

$$\mathbb{E}\big[s(x)\big] = \mathbb{E}[r_1]\, x_1 + \cdots + \mathbb{E}[r_n]\, x_n$$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \dots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \dots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad \Sigma = \mathbb{E}\big[(r-\mu)(r-\mu)^\top\big]$$

- An investment $x = (x_1, \dots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

$$\mathbb{E}\big[s(x)\big] = \mathbb{E}[r_1]\, x_1 + \cdots + \mathbb{E}[r_n]\, x_n = \mu_1\, x_1 + \cdots + \mu_n\, x_n$$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad \Sigma = \mathbb{E}\big[(r-\mu)(r-\mu)^\top\big]$$

- An investment $x = (x_1, \ldots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

$$\mathbb{E}\big[s(x)\big] = \mathbb{E}[r_1]\, x_1 + \cdots + \mathbb{E}[r_n]\, x_n = \mu_1\, x_1 + \cdots + \mu_n\, x_n = \mu^\top x$$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad\qquad \Sigma = \mathbb{E}\big[(r-\mu)(r-\mu)^\top\big]$$

- An investment $x = (x_1, \ldots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

$$\mathbb{E}\big[s(x)\big] = \mathbb{E}[r_1]\, x_1 + \cdots + \mathbb{E}[r_n]\, x_n = \mu_1\, x_1 + \cdots + \mu_n\, x_n = \mu^\top x$$

$\mathrm{Var}\big(s(x)\big)$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad \Sigma = \mathbb{E}\big[(r-\mu)(r-\mu)^\top\big]$$

- An investment $x = (x_1, \ldots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

$$\mathbb{E}\big[s(x)\big] = \mathbb{E}[r_1]\, x_1 + \cdots + \mathbb{E}[r_n]\, x_n = \mu_1\, x_1 + \cdots + \mu_n\, x_n = \mu^\top x$$

$$\mathsf{Var}\big(s(x)\big) = \mathbb{E}\Big[\big(s(x) - \mathbb{E}[s(x)]\big)^2\Big]$$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad\qquad \Sigma = \mathbb{E}\big[(r-\mu)(r-\mu)^\top\big]$$

- An investment $x = (x_1, \ldots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

$$\mathbb{E}\big[s(x)\big] = \mathbb{E}[r_1]\, x_1 + \cdots + \mathbb{E}[r_n]\, x_n = \mu_1\, x_1 + \cdots + \mu_n\, x_n = \mu^\top x$$

$$\mathsf{Var}\big(s(x)\big) = \mathbb{E}\Big[\big(s(x) - \mathbb{E}[s(x)]\big)^2\Big] = \mathbb{E}\Big[\big((r-\mu)^\top x\big)^2\Big]$$

Portfolio Optimization

- £T to invest in n assets
- r_i: return of asset i, for $i = 1, \ldots, n$ (random variables)
- First two moments of the random vector $r = (r_1, \ldots, r_n)$ are known:

$$\mu = \mathbb{E}[r] \qquad \Sigma = \mathbb{E}\big[(r-\mu)(r-\mu)^\top\big]$$

- An investment $x = (x_1, \ldots, x_n)$ has return

$$s(x) = r^\top x = r_1 x_1 + \cdots + r_n x_n$$

$$\mathbb{E}\big[s(x)\big] = \mathbb{E}[r_1]\, x_1 + \cdots + \mathbb{E}[r_n]\, x_n = \mu_1\, x_1 + \cdots + \mu_n\, x_n = \mu^\top x$$

$$\mathsf{Var}\big(s(x)\big) = \mathbb{E}\Big[\big(s(x) - \mathbb{E}[s(x)]\big)^2\Big] = \mathbb{E}\Big[\big((r-\mu)^\top x\big)^2\Big] = x^\top \Sigma\, x$$

Portfolio Optimization

Different investment strategies

Portfolio Optimization

Different investment strategies

- Minimize variance while guaranteeing minimum expected return $s_{\min}$:

Portfolio Optimization

Different investment strategies

- Minimize variance while guaranteeing minimum expected return $s_{\min}$:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \mathsf{Var}\big(s(x)\big) \\ \text{subject to} \quad & x \geq 0_n \\ & 1_n^\top x = T \\ & \mathbb{E}\big[s(x)\big] \geq s_{\min} \end{aligned}$$

Portfolio Optimization

Different investment strategies

- Minimize variance while guaranteeing minimum expected return $s_{\min}$:

$$
\begin{array}{ll} \underset{x}{\text{minimize}} & \mathsf{Var}\big(s(x)\big) \\ \text{subject to} & x \geq 0_n \\ & 1_n^\top x = T \\ & \mathbb{E}\big[s(x)\big] \geq s_{\min} \end{array}
\quad \Longleftrightarrow \quad
\begin{array}{ll} \underset{x}{\text{minimize}} & x^\top \Sigma x \\ \text{subject to} & x \geq 0_n \\ & 1_n^\top x = T \\ & \mu^\top x \geq s_{\min} \end{array}
$$

Portfolio Optimization

Different investment strategies

- Minimize variance while guaranteeing minimum expected return $s_{\min}$:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \text{Var}\big(s(x)\big) \\ \text{subject to} & x \geq 0_n \\ & 1_n^\top x = T \\ & \mathbb{E}\big[s(x)\big] \geq s_{\min} \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \underset{x}{\text{minimize}} & x^\top \Sigma x \\ \text{subject to} & x \geq 0_n \\ & 1_n^\top x = T \\ & \mu^\top x \geq s_{\min} \end{array}$$

Convex Quadratic Program (QP)

Portfolio Optimization

- Maximize expected return with risk penalty $\beta \geq 0$:

Portfolio Optimization

- Maximize expected return with risk penalty $\beta \geq 0$:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \mathbb{E}\big[s(x)\big] - \beta\,\mathsf{Var}(s(x)) \\ \text{subject to} \quad & x \geq 0_n \\ & 1_n^\top x = T \end{aligned}$$

Portfolio Optimization

- Maximize expected return with risk penalty $\beta \geq 0$:

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \mathbb{E}\big[s(x)\big] - \beta\,\mathsf{Var}(s(x)) \\ \text{subject to} \quad & x \geq 0_n \\ & 1_n^\top x = T \end{aligned}$$

$$\iff \qquad \begin{aligned} \underset{x}{\text{minimize}} \quad & -\mu^\top x + \beta\, x^\top \Sigma\, x \\ \text{subject to} \quad & x \geq 0_n \\ & 1_n^\top x = T \end{aligned}$$

Portfolio Optimization

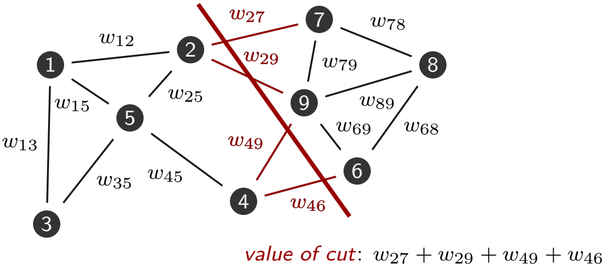
- Maximize expected return with risk penalty $\beta \geq 0$:

$$\begin{array}{ll} \underset{x}{\text{maximize}} & \mathbb{E}\big[s(x)\big] - \beta\,\mathsf{Var}(s(x)) \\ \text{subject to} & x \geq 0_n \\ & 1_n^\top x = T \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{x}{\text{minimize}} & -\mu^\top x + \beta\, x^\top \Sigma\, x \\ \text{subject to} & x \geq 0_n \\ & 1_n^\top x = T \end{array}$$

Convex QP

MAXCUT



Cut: set of edges whose removal splits the graph into two

MAXCUT problem: find the cut with maximum weight

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{maximize}} & \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \frac{1 - x_i x_j}{2} \\ \text{subject to} & x_i \in \{-1, 1\}, \qquad i = 1, \ldots, n. \end{array}$$

$$
\begin{aligned}
p^\star = \quad & \max_{x} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,\frac{1-x_i x_j}{2} \\
& \text{s.t.} \quad x_i \in \{-1, 1\}\,, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
p^\star = \; & \max_x \quad \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n w_{ij}\,\frac{1 - x_i x_j}{2} \\
& \text{s.t.} \quad x_i \in \{-1, 1\}\,, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
= - \; & \min_x \quad \frac{1}{4}\left[\sum_{i=1}^n\sum_{j=1}^n w_{ij} x_i x_j - \sum_{i=1}^n\sum_{j=1}^n w_{ij}\right] \\
& \text{s.t.} \quad x_i^2 = 1\,, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
p^\star = \ \max_x \quad & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\frac{1 - x_i x_j}{2} \\
\text{s.t.} \quad & x_i \in \{-1, 1\}, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
= -\ \min_x \quad & \frac{1}{4}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_i x_j - \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right] \\
\text{s.t.} \quad & x_i^2 = 1, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
= -\ \min_x \quad & \frac{1}{4}\left(x^\top W x - 1_n^\top W 1_n\right) \\
\text{s.t.} \quad & x_i^2 = 1, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
p^\star = \ & \max_x \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\,\frac{1 - x_i x_j}{2} \\
& \text{s.t.} \quad x_i \in \{-1, 1\}\,, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
= - & \min_x \quad \frac{1}{4}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_i x_j - \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right] \\
& \text{s.t.} \quad x_i^2 = 1\,, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{aligned}
= - & \min_x \quad \frac{1}{4}\left(x^\top W x - 1_n^\top W 1_n\right) \\
& \text{s.t.} \quad x_i^2 = 1\,, \qquad i = 1, \ldots, n
\end{aligned}
$$

$W \in \mathbb{R}^{n \times n}$: weighted adjacency matrix

$$
\begin{aligned}
-p^\star = \quad \min_{x} \quad & \frac{1}{4}\Big(x^\top W x - 1_n^\top W 1_n\Big) \\
\text{s.t.} \quad & x_i^2 = 1\,, \qquad i = 1, \ldots, n
\end{aligned}
$$

$$
\begin{array}{rll}
-p^\star = & \min\limits_{x} & \dfrac{1}{4}\Big(x^\top W x - 1_n^\top W 1_n\Big) \\
& \text{s.t.} & x_i^2 = 1\,, \qquad i = 1, \ldots, n
\end{array}
$$

- $x^\top W x \;=\; \mathrm{tr}\big(x^\top W x\big) \;=\; \mathrm{tr}\big(W x x^\top\big)$

$$
\begin{aligned}
-p^\star = \quad \min_x \quad & \frac{1}{4}\Big(x^\top W x - 1_n^\top W 1_n\Big) \\
\text{s.t.} \quad & x_i^2 = 1\,, \qquad i = 1, \ldots, n
\end{aligned}
$$

- $x^\top W x \;=\; \mathsf{tr}\big(x^\top W x\big) \;=\; \mathsf{tr}\big(W x x^\top\big)$
- For any $X \in S^n$ and $x \in \mathbb{R}^n$, $\quad X = xx^\top \quad \Longleftrightarrow \quad \begin{cases} X \succeq 0 \\ \mathsf{rank}(X) = 1 \end{cases}$

$$\begin{array}{rll} -p^\star = & \min\limits_{x} & \frac{1}{4}\Big(x^\top W x - 1_n^\top W 1_n\Big) \\ & \text{s.t.} & x_i^2 = 1\,, \qquad i = 1, \ldots, n \end{array}$$

- $x^\top W x \;=\; \mathsf{tr}\big(x^\top W x\big) \;=\; \mathsf{tr}\big(W x x^\top\big)$
- For any $X \in S^n$ and $x \in \mathbb{R}^n$, $\quad X = x x^\top \quad \iff \quad \begin{cases} X \succeq 0 \\ \mathsf{rank}(X) = 1 \end{cases}$

$$\begin{array}{rll} -p^\star = & \min\limits_{X \in S^n} & \frac{1}{4}\Big(\mathsf{tr}\big(W X\big) - 1_n^\top W 1_n\Big) \\ & \text{s.t.} & X_{ii} = 1\,, \qquad i = 1, \ldots, n \\ & & X \succeq 0 \\ & & \mathsf{rank}(X) = 1 \end{array}$$

$$
\begin{aligned}
-p^\star = \quad \min_{x} \quad & \frac{1}{4}\Big(x^\top W x - 1_n^\top W 1_n\Big) \\
\text{s.t.} \quad & x_i^2 = 1\,, \qquad i = 1,\ldots,n
\end{aligned}
$$

- $x^\top W x \;=\; \mathrm{tr}\big(x^\top W x\big) \;=\; \mathrm{tr}\big(W x x^\top\big)$
- For any $X \in S^n$ and $x \in \mathbb{R}^n$, $\quad X = xx^\top \quad \Longleftrightarrow \quad \begin{cases} X \succeq 0 \\ \mathrm{rank}(X) = 1 \end{cases}$

$$
\begin{aligned}
-p^\star = \quad \min_{X \in S^n} \quad & \frac{1}{4}\Big(\mathrm{tr}\big(WX\big) - 1_n^\top W 1_n\Big) \\
\text{s.t.} \quad & X_{ii} = 1\,, \qquad i = 1,\ldots,n \\
& X \succeq 0 \\
& \mathrm{rank}(X) = 1 \quad \text{——— } \textit{nonconvex}
\end{aligned}
$$

Relax . . .

$$
\begin{aligned}
-p^\star = \quad \min_{X \in S^n} \quad & \frac{1}{4}\Big(\mathsf{tr}\big(WX\big) - 1_n^\top W 1_n\Big) \\
\text{s.t.} \quad & X_{ii} = 1\,, \qquad i = 1, \ldots, n \\
& X \succeq 0 \\
& \mathsf{rank}(X) = 1
\end{aligned}
$$

Relax . . .

$$
\begin{aligned}
-p^\star = \quad & \min_{X \in S^n} && \frac{1}{4}\Big(\mathsf{tr}(WX) - 1_n^\top W 1_n\Big) \\
& \text{s.t.} && X_{ii} = 1\,, \qquad i = 1, \dots, n \\
& && X \succeq 0 \\
& && \mathsf{rank}(X) = 1
\end{aligned}
$$

$$
\begin{aligned}
\geq \quad & \min_{X \in S^n} && \frac{1}{4}\Big(\mathsf{tr}(WX) - 1_n^\top W 1_n\Big) \\
& \text{s.t.} && X_{ii} = 1\,, \qquad i = 1, \dots, n \\
& && X \succeq 0
\end{aligned}
$$

Relax . . .

$$
\begin{aligned}
-p^\star = \quad & \min_{X \in S^n} && \frac{1}{4}\Big(\mathsf{tr}\big(WX\big) - 1_n^\top W 1_n\Big) \\
& \text{s.t.} && X_{ii} = 1\,, \qquad i = 1, \ldots, n \\
& && X \succeq 0 \\
& && \textcolor{darkred}{\mathsf{rank}(X) = 1}
\end{aligned}
$$

$$
\begin{aligned}
\geq \quad & \min_{X \in S^n} && \frac{1}{4}\Big(\mathsf{tr}\big(WX\big) - 1_n^\top W 1_n\Big) \\
& \text{s.t.} && X_{ii} = 1\,, \qquad i = 1, \ldots, n \\
& && X \succeq 0
\end{aligned}
$$

$$=: -d^\star$$

Relax ...

$$
\begin{aligned}
-p^\star = \quad & \min_{X \in S^n} && \frac{1}{4}\Big(\mathsf{tr}\big(WX\big) - 1_n^\top W 1_n\Big) \\
& \text{s.t.} && X_{ii} = 1\,, \qquad i = 1, \ldots, n \\
& && X \succeq 0 \\
& && \color{darkred}{\mathsf{rank}(X) = 1} \\
\\
\geq \quad & \min_{X \in S^n} && \frac{1}{4}\Big(\mathsf{tr}\big(WX\big) - 1_n^\top W 1_n\Big) \\
& \text{s.t.} && X_{ii} = 1\,, \qquad i = 1, \ldots, n \\
& && X \succeq 0 \\
\\
=: -d^\star &
\end{aligned}
$$

Convex Semi-Definite Program (SDP)

Procedure: (Goemans & Williamson, 95')

Procedure: (Goemans & Williamson, 95')

- Solve SDP, obtaining $X^\star$

Procedure: (Goemans & Williamson, 95')

- Solve SDP, obtaining $X^\star$

- Factorize $X^\star = V^\top V$, where $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$

Procedure: (Goemans & Williamson, 95')

- Solve SDP, obtaining $X^\star$
- Factorize $X^\star = V^\top V$, where $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$
- Draw r uniformly at random from unit-sphere $\left\{x \in \mathbb{R}^n \,:\, \|x\|_2 = 1\right\}$

Procedure: (Goemans & Williamson, 95')

- Solve SDP, obtaining $X^\star$

- Factorize $X^\star = V^\top V$, where $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$

- Draw r uniformly at random from unit-sphere $\left\{x \in \mathbb{R}^n \,:\, \|x\|_2 = 1\right\}$

- The following sets define a cut C on the vertices of the graph:

$$S = \left\{i \,:\, r^\top v_i \geq 0\right\} \qquad S^c = \left\{i \,:\, r^\top v_i < 0\right\}$$

Procedure: (Goemans & Williamson, 95')

- Solve SDP, obtaining $X^\star$
- Factorize $X^\star = V^\top V$, where $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$
- Draw r uniformly at random from unit-sphere $\left\{x \in \mathbb{R}^n \,:\, \|x\|_2 = 1\right\}$
- The following sets define a cut C on the vertices of the graph:

$$S = \left\{i \,:\, r^\top v_i \geq 0\right\} \qquad S^c = \left\{i \,:\, r^\top v_i < 0\right\}$$

It can be shown that

$$d^\star \,\geq\, p^\star \,\geq\, \mathbb{E}[C] \,\geq\, 0.87856\, d^\star$$

Solving Optimization Problems

Solving Optimization Problems

CVX & other solvers are great for

Solving Optimization Problems

CVX & other solvers are great for

- *prototyping*
- *small-scale* problems

Solving Optimization Problems

CVX & other solvers are great for

- *prototyping*
- *small-scale* problems

Large-scale problems & real-time solutions require tailored solvers

Example:

Example:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

Example:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are generated randomly.

Example:

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \|x\|_1 \\ \text{subject to} & Ax = b \end{array}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are generated randomly.

For $n = 5000$ and $m = 500$,

Example:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \|x\|_1 \\ \text{subject to} \quad & Ax = b \end{aligned}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are generated randomly.

For $n = 5000$ and $m = 500$,

- CVX: 56.16 s
- SPGL1: 0.82 s (tailored solver)

Optimization Algorithms

Optimization Algorithms

- **Unconstrained & differentiable**

  Gradient descent (and Nesterov's acceleration scheme)

  Block coordinate descent

  Newton (and approximations)

Optimization Algorithms

- **Unconstrained & differentiable**

  Gradient descent (and Nesterov's acceleration scheme)

  Block coordinate descent

  Newton (and approximations)

- **Constrained & differentiable**

  Projection methods (projected gradient descent, Frank-Wolfe, . . .)

  Interior-point algorithms (classes LP, QP, SOCP, SDP)

Optimization Algorithms

- **Unconstrained & differentiable**

  Gradient descent (and Nesterov's acceleration scheme)

  Block coordinate descent

  Newton (and approximations)

- **Constrained & differentiable**

  Projection methods (projected gradient descent, Frank-Wolfe, . . .)

  Interior-point algorithms (classes LP, QP, SOCP, SDP)

- **Non-differentiable**

  Subgradient descent

  Proximal methods (proximal gradient descent, ADMM, primal-dual)

Optimization Algorithms

- **Unconstrained & differentiable**

  Gradient descent (and Nesterov's acceleration scheme)

  Block coordinate descent

  Newton (and approximations)

- **Constrained & differentiable**

  Projection methods (projected gradient descent, Frank-Wolfe, . . .)

  Interior-point algorithms (classes LP, QP, SOCP, SDP)

- **Non-differentiable**

  Subgradient descent

  Proximal methods (proximal gradient descent, ADMM, primal-dual)

And now, *neural networks!*

Nesterov's acceleration scheme

Nesterov's acceleration scheme

$$\underset{x}{\text{minimize}} \quad f(x)$$

Nesterov's acceleration scheme

$$\underset{x}{\text{minimize}} \quad f(x)$$

- $f : \mathbb{R}^n \to \mathbb{R}$: differentiable

Nesterov's acceleration scheme

$$\underset{x}{\text{minimize}} \quad f(x)$$

- $f : \mathbb{R}^n \to \mathbb{R}$: differentiable
- ∇f : Lipschitz-continuous

Nesterov's acceleration scheme

$$\underset{x}{\text{minimize}} \quad f(x)$$

- $f : \mathbb{R}^n \to \mathbb{R}$: differentiable
- ∇f: Lipschitz-continuous, i.e., there exists $L \geq 0$ such that

$$\|\nabla f(y) - \nabla f(x)\|_2 \leq L\|y - x\|_2 \,, \qquad \text{for all } x, y \in \mathbf{dom}\, f$$

Nesterov's acceleration scheme

$$\underset{x}{\text{minimize}} \quad f(x)$$

- $f \,:\, \mathbb{R}^n \to \mathbb{R}$: differentiable
- ∇f : Lipschitz-continuous, i.e., there exists $L \geq 0$ such that

$$\|\nabla f(y) - \nabla f(x)\|_2 \leq L\|y - x\|_2\,, \qquad \text{for all } x, y \in \text{dom}\, f$$

- $f^\star := \min_x \; f(x) > -\infty$

Nesterov's acceleration scheme

Gradient descent:

Nesterov's acceleration scheme

Gradient descent: starting at arbitrary $x^0 \in \mathbb{R}^n$,

Nesterov's acceleration scheme

Gradient descent: starting at arbitrary $x^0 \in \mathbb{R}^n$,

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\big(x^k\big)$$

Nesterov's acceleration scheme

Gradient descent: starting at arbitrary $x^0 \in \mathbb{R}^n$,

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\big(x^k\big)$$

$$f(x^k) - f^\star \le \frac{L}{2k}\big\|x^0 - x^\star\big\|_2^2$$

Nesterov's acceleration scheme

Gradient descent: starting at arbitrary $x^0 \in \mathbb{R}^n$,

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\big(x^k\big)$$

$$f(x^k) - f^\star \leq \frac{L}{2k}\big\|x^0 - x^\star\big\|_2^2$$

Nesterov's method: starting at arbitrary $y^0 \in \mathbb{R}^n$,

Nesterov's acceleration scheme

Gradient descent: starting at arbitrary $x^0 \in \mathbb{R}^n$,

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\big(x^k\big)$$

$$f(x^k) - f^\star \le \frac{L}{2k}\big\|x^0 - x^\star\big\|_2^2$$

Nesterov's method: starting at arbitrary $y^0 \in \mathbb{R}^n$,

$$x^{k+1} = y^k - \frac{1}{L}\nabla f\big(y^k\big)$$

$$y^{k+1} = x^k - \frac{k-1}{k+2}\big(x^{k+1} - x^k\big)$$

Nesterov's acceleration scheme

Gradient descent: starting at arbitrary $x^0 \in \mathbb{R}^n$,

$$x^{k+1} = x^k - \frac{1}{L}\nabla f\big(x^k\big)$$

$$f(x^k) - f^\star \le \frac{L}{2k}\big\|x^0 - x^\star\big\|_2^2$$

Nesterov's method: starting at arbitrary $y^0 \in \mathbb{R}^n$,

$$x^{k+1} = y^k - \frac{1}{L}\nabla f\big(y^k\big)$$

$$y^{k+1} = x^k - \frac{k-1}{k+2}\big(x^{k+1} - x^k\big)$$

$$f(x^k) - f^\star \le \frac{2L}{(k+1)^2}\big\|x^0 - x^\star\big\|_2^2$$

Example

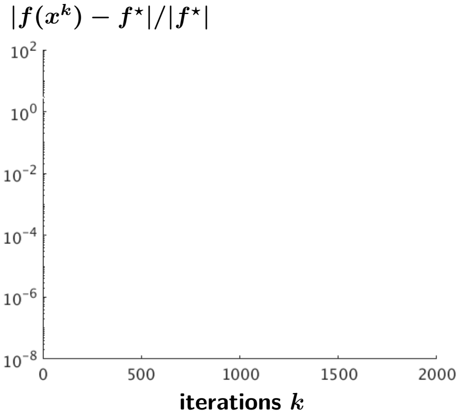
$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \log \sum_{i=1}^{m} \exp\left(a_i^\top x + b\right)$$

randomly generated data with $m = 500$, $n = 50$

Example

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \log \sum_{i=1}^{m} \exp\left(a_i^\top x + b\right)$$

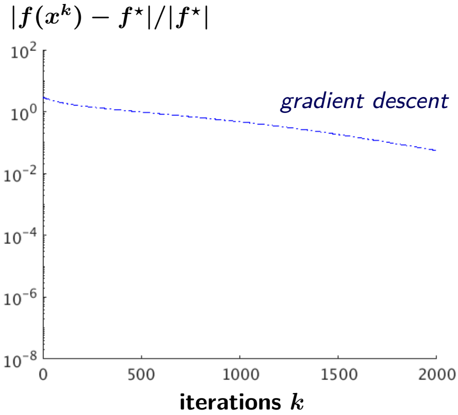
randomly generated data with $m = 500$, $n = 50$



Example

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \log \sum_{i=1}^{m} \exp\left(a_i^\top x + b\right)$$

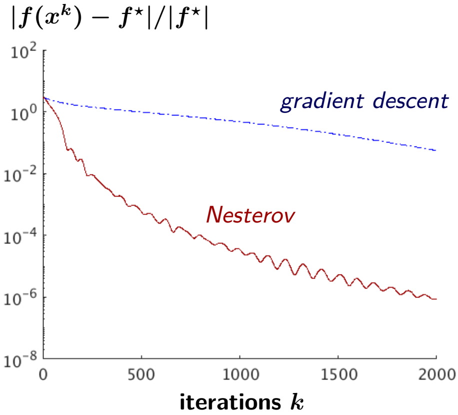
randomly generated data with $m = 500$, $n = 50$



Example

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \log \sum_{i=1}^{m} \exp\left(a_i^\top x + b\right)$$

randomly generated data with $m = 500$, $n = 50$



Outline

Convex sets

Identifying convex sets

Examples: geometrical sets and filter design constraints

Convex functions

Identifying convex functions

Relation to convex sets

Optimization problems

Convex problems, properties, and problem manipulation

Examples and solvers

Statistical estimation

Maximum likelihood & maximum a posteriori

Nonparametric estimation

Hypothesis testing & optimal detection

Maximum likelihood

Maximum likelihood

$Y \in \mathbb{R}^m$: random vector with density $f_Y(y\,;\,x)$

Maximum likelihood

$Y \in \mathbb{R}^m$: random vector with density $f_Y(y\,;\,x)$

$x \in \mathbb{R}^n$: parameter to estimate

Maximum likelihood

$Y \in \mathbb{R}^m$: random vector with density $f_Y(y\,;\,x)$

$x \in \mathbb{R}^n$: parameter to estimate (we may know that $x \in C \subset \mathbb{R}^n$)

Maximum likelihood

$Y \in \mathbb{R}^m$: random vector with density $f_Y(y\,;x)$

$x \in \mathbb{R}^n$: parameter to estimate (we may know that $x \in C \subset \mathbb{R}^n$)

Maximum Likelihood (ML) estimate: Given realization y of Y,

Maximum likelihood

$Y \in \mathbb{R}^m$: random vector with density $f_Y(y\,;\,x)$

$x \in \mathbb{R}^n$: parameter to estimate (we may know that $x \in C \subset \mathbb{R}^n$)

Maximum Likelihood (ML) estimate: Given realization y of Y,

$$\begin{array}{ll} \underset{x}{\text{maximize}} & f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array}$$

Maximum likelihood

$Y \in \mathbb{R}^m$: random vector with density $f_Y(y\,;\,x)$

$x \in \mathbb{R}^n$: parameter to estimate (we may know that $x \in C \subset \mathbb{R}^n$)

Maximum Likelihood (ML) estimate: Given realization y of Y,

$$\begin{array}{ll} \underset{x}{\text{maximize}} & f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \underset{x}{\text{maximize}} & \log f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array}$$

Maximum likelihood

$Y \in \mathbb{R}^m$: random vector with density $f_Y(y\,;\,x)$

$x \in \mathbb{R}^n$: parameter to estimate (we may know that $x \in C \subset \mathbb{R}^n$)

Maximum Likelihood (ML) estimate: Given realization y of Y,

$$\begin{array}{ll} \underset{x}{\text{maximize}} & f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \underset{x}{\text{maximize}} & \log\, f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array}$$

$$\Longleftrightarrow \quad \begin{array}{ll} \underset{x}{\text{minimize}} & -\log\, f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array}$$

Maximum likelihood

$Y \in \mathbb{R}^m$: random vector with density $f_Y(y\,;\,x)$

$x \in \mathbb{R}^n$: parameter to estimate (we may know that $x \in C \subset \mathbb{R}^n$)

Maximum Likelihood (ML) estimate: Given realization y of Y,

$$\begin{array}{ll} \underset{x}{\text{maximize}} & f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \underset{x}{\text{maximize}} & \log f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array}$$

$$\Longleftrightarrow \quad \begin{array}{ll} \underset{x}{\text{minimize}} & -\log f_Y(y\,;\,x) \\ \text{subject to} & x \in C \end{array}$$

negative log-likelihood

Maximum likelihood

Example: linear measurement model

Maximum likelihood

Example: linear measurement model

- Parameter to estimate: $x \in \mathbb{R}^n$

Maximum likelihood

Example: linear measurement model

- Parameter to estimate: $x \in \mathbb{R}^n$
- Observations: $Y_i = a_i^\top x + V_i, \quad i = 1, \ldots, m,$ for known $a_i \in \mathbb{R}^n$

Maximum likelihood

Example: linear measurement model

- Parameter to estimate: $x \in \mathbb{R}^n$
- Observations: $Y_i = a_i^\top x + V_i, \quad i = 1, \ldots, m,$ for known $a_i \in \mathbb{R}^n$

  V_i: iid copies of random variable V with density $f_V(v)$

Maximum likelihood

Example: linear measurement model

- Parameter to estimate: $x \in \mathbb{R}^n$
- Observations: $Y_i = a_i^\top x + V_i, \quad i = 1, \ldots, m,$ for known $a_i \in \mathbb{R}^n$

  V_i: iid copies of random variable V with density $f_V(v)$

$\Longrightarrow \quad Y_i$ has density $f_V(y_i - a_i^\top x)$

Maximum likelihood

Example: linear measurement model

- Parameter to estimate: $x \in \mathbb{R}^n$
- Observations: $Y_i = a_i^\top x + V_i, \quad i = 1, \ldots, m,$ for known $a_i \in \mathbb{R}^n$

  V_i: iid copies of random variable V with density $f_V(v)$

$\Longrightarrow \quad Y_i$ has density $f_V(y_i - a_i^\top x)$ and Y_i's are independent

Maximum likelihood

Example: linear measurement model

- Parameter to estimate: $x \in \mathbb{R}^n$
- Observations: $Y_i = a_i^\top x + V_i, \quad i = 1, \ldots, m,$ for known $a_i \in \mathbb{R}^n$

  V_i: iid copies of random variable V with density $f_V(v)$

$\Longrightarrow \quad Y_i$ has density $f_V(y_i - a_i^\top x)$ and Y_i's are independent

$$\Longrightarrow \quad f_{Y_1 \cdots Y_m}(y_1, \ldots, y_m) = \prod_{i=1}^m f_{Y_i}(y_i) = \prod_{i=1}^m f_V(y_i - a_i^\top x)$$

Maximum likelihood

Example: linear measurement model

- Parameter to estimate: $x \in \mathbb{R}^n$
- Observations: $Y_i = a_i^\top x + V_i, \quad i = 1, \ldots, m,$ for known $a_i \in \mathbb{R}^n$

  V_i: iid copies of random variable V with density $f_V(v)$

$\Longrightarrow \quad Y_i$ has density $f_V(y_i - a_i^\top x)$ and Y_i's are independent

$$\Longrightarrow \quad f_{Y_1 \cdots Y_m}(y_1, \ldots, y_m) = \prod_{i=1}^m f_{Y_i}(y_i) = \prod_{i=1}^m f_V(y_i - a_i^\top x)$$

The a_i's will denote the rows of $A = \begin{bmatrix} — & a_1^\top & — \\ & \vdots & \\ — & a_m^\top & — \end{bmatrix} \in \mathbb{R}^{m \times n}$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \ \in \ \arg\min_x \ -\log \prod_{i=1}^{m} f_V(y_i - a_i^\top x)$$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \;\in\; \arg\min_x \; -\log \prod_{i=1}^{m} f_V(y_i - a_i^\top x) \;=\; \arg\min_x \; -\sum_{i=1}^{m} \log\, f_V(y_i - a_i^\top x)$$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \;\in\; \arg\min_{x} \; -\log \prod_{i=1}^{m} f_V(y_i - a_i^\top x) \;=\; \arg\min_{x} \; -\sum_{i=1}^{m} \log \, f_V(y_i - a_i^\top x)$$

Common noise densities:

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x -\log \prod_{i=1}^m f_V(y_i - a_i^\top x) = \arg\min_x -\sum_{i=1}^m \log f_V(y_i - a_i^\top x)$$

Common noise densities:

Gaussian:

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \;\in\; \arg\min_x \; -\log \prod_{i=1}^m f_V(y_i - a_i^\top x) \;=\; \arg\min_x \; -\sum_{i=1}^m \log\, f_V(y_i - a_i^\top x)$$

Common noise densities:

Gaussian: $V \sim \mathcal{N}(0, \sigma^2)$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \;\in\; \arg\min_x \; -\log \prod_{i=1}^{m} f_V(y_i - a_i^\top x) \;=\; \arg\min_x \; -\sum_{i=1}^{m} \log\, f_V(y_i - a_i^\top x)$$

Common noise densities:

Gaussian: $V \sim \mathcal{N}(0, \sigma^2)$ $\qquad f_V(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{v^2}{2\sigma^2}\right)$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \;\in\; \arg\min_x \; -\log \prod_{i=1}^m f_V(y_i - a_i^\top x) \;=\; \arg\min_x \; -\sum_{i=1}^m \log\, f_V(y_i - a_i^\top x)$$

Common noise densities:

Gaussian: $V \sim \mathcal{N}(0, \sigma^2)$ $\qquad f_V(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{v^2}{2\sigma^2}\right)$

$$\widehat{x}_{\mathsf{ML}} \;\in\; \arg\min_x \; -\sum_{i=1}^m \log\, \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - a_i^\top x)^2}{2\sigma^2}\right)$$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x \; -\log \prod_{i=1}^m f_V(y_i - a_i^\top x) \;=\; \arg\min_x \; -\sum_{i=1}^m \log\, f_V(y_i - a_i^\top x)$$

Common noise densities:

Gaussian: $V \sim \mathcal{N}(0, \sigma^2)$ $\qquad f_V(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{v^2}{2\sigma^2}\right)$

$$\begin{aligned}\widehat{x}_{\mathsf{ML}} &\in \arg\min_x \; -\sum_{i=1}^m \log\, \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - a_i^\top x)^2}{2\sigma^2}\right) \\ &= \arg\min_x \; \frac{1}{2\sigma^2} \sum_{i=1}^m \left(y_i - a_i^\top x\right)^2\end{aligned}$$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \;\in\; \arg\min_{x} \; -\log \prod_{i=1}^{m} f_V(y_i - a_i^\top x) \;=\; \arg\min_{x} \; -\sum_{i=1}^{m} \log\, f_V(y_i - a_i^\top x)$$

Common noise densities:

Gaussian: $V \sim \mathcal{N}(0, \sigma^2)$ $\qquad f_V(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{v^2}{2\sigma^2}\right)$

$$\begin{aligned}
\widehat{x}_{\mathsf{ML}} \;&\in\; \arg\min_{x} \; -\sum_{i=1}^{m} \log\, \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - a_i^\top x)^2}{2\sigma^2}\right) \\
&=\; \arg\min_{x} \; \frac{1}{2\sigma^2} \sum_{i=1}^{m} \left(y_i - a_i^\top x\right)^2 \\
&=\; \arg\min_{x} \; \frac{1}{2\sigma^2} \left\|y - Ax\right\|_2^2
\end{aligned}$$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x \; -\log \prod_{i=1}^{m} f_V(y_i - a_i^\top x) \;=\; \arg\min_x \; -\sum_{i=1}^{m} \log \, f_V(y_i - a_i^\top x)$$

Common noise densities:

Gaussian: $V \sim \mathcal{N}(0, \sigma^2)$ $\qquad f_V(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{v^2}{2\sigma^2}\right)$

$$\begin{aligned}
\widehat{x}_{\mathsf{ML}} &\in \arg\min_x \; -\sum_{i=1}^{m} \log \, \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - a_i^\top x)^2}{2\sigma^2}\right) \\
&= \arg\min_x \; \frac{1}{2\sigma^2} \sum_{i=1}^{m} \left(y_i - a_i^\top x\right)^2 \\
&= \arg\min_x \; \frac{1}{2\sigma^2} \left\| y - Ax \right\|_2^2
\end{aligned}$$

Convex

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x -\sum_{i=1}^m \log f_V(y_i - a_i^\top x)$$

Laplacian:

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_{x} -\sum_{i=1}^{m} \log f_V(y_i - a_i^\top x)$$

Laplacian: $V \sim \mathcal{L}(0,\, a)$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_{x} -\sum_{i=1}^{m} \log f_V(y_i - a_i^\top x)$$

Laplacian: $V \sim \mathcal{L}(0,\, a)$ $\qquad f_V(v) = \frac{1}{\sqrt{2a}} \exp\left(-\frac{|v|}{a}\right)$ $\qquad (a > 0)$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x \; -\sum_{i=1}^m \log \, f_V(y_i - a_i^\top x)$$

Laplacian: $\quad V \sim \mathcal{L}(0,\, a) \qquad f_V(v) = \frac{1}{\sqrt{2a}} \exp\left(-\frac{|v|}{a}\right) \qquad (a > 0)$

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x \; -\sum_{i=1}^m \log \, \frac{1}{\sqrt{2a}} \exp\left(-\frac{|y_i - a_i^\top x|}{a}\right)$$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_{x} \; -\sum_{i=1}^{m} \log\, f_V(y_i - a_i^\top x)$$

Laplacian: $V \sim \mathcal{L}(0,\, a) \qquad f_V(v) = \frac{1}{\sqrt{2a}} \exp\left(-\frac{|v|}{a}\right) \qquad (a > 0)$

$$\begin{aligned} \widehat{x}_{\mathsf{ML}} &\in \arg\min_{x} \; -\sum_{i=1}^{m} \log\, \frac{1}{\sqrt{2a}} \exp\left(-\frac{|y_i - a_i^\top x|}{a}\right) \\ &= \arg\min_{x} \; \frac{1}{a} \sum_{i=1}^{m} \left|y_i - a_i^\top x\right| \end{aligned}$$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x \; -\sum_{i=1}^m \log f_V(y_i - a_i^\top x)$$

Laplacian: $V \sim \mathcal{L}(0,\, a)$ $\qquad f_V(v) = \frac{1}{\sqrt{2a}} \exp\left(-\frac{|v|}{a}\right)$ $\qquad (a > 0)$

$$\begin{aligned}
\widehat{x}_{\mathsf{ML}} &\in \arg\min_x \; -\sum_{i=1}^m \log \frac{1}{\sqrt{2a}} \exp\left(-\frac{|y_i - a_i^\top x|}{a}\right) \\
&= \arg\min_x \; \frac{1}{a} \sum_{i=1}^m \left|y_i - a_i^\top x\right| \\
&= \arg\min_x \; \frac{1}{a} \left\|y - Ax\right\|_1
\end{aligned}$$

Maximum likelihood

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x \; - \sum_{i=1}^{m} \log \, f_V(y_i - a_i^\top x)$$

Laplacian: $V \sim \mathcal{L}(0,\, a)$ $\qquad f_V(v) = \frac{1}{\sqrt{2a}} \exp\left(-\frac{|v|}{a}\right)$ $\qquad (a > 0)$

$$\begin{aligned}
\widehat{x}_{\mathsf{ML}} \;&\in\; \arg\min_x \; - \sum_{i=1}^{m} \log \, \frac{1}{\sqrt{2a}} \exp\left(-\frac{|y_i - a_i^\top x|}{a}\right) \\
&=\; \arg\min_x \; \frac{1}{a} \sum_{i=1}^{m} \left|y_i - a_i^\top x\right| \\
&=\; \arg\min_x \; \frac{1}{a} \left\|y - Ax\right\|_1
\end{aligned}$$

Convex

Maximum likelihood

Uniform:

Maximum likelihood

Uniform: $V \sim \mathcal{U}(-c, c)$

Maximum likelihood

Uniform: $V \sim \mathcal{U}(-c, c)$ $$f_V(v) = \begin{cases} \frac{1}{2c} & , \ |v| \leq c \\ 0 & , \ \text{oth.} \end{cases}$$

Maximum likelihood

Uniform: $V \sim \mathcal{U}(-c, c)$ $\qquad f_V(v) = \begin{cases} \frac{1}{2c} & , \ |v| \leq c \\ 0 & , \text{ oth.} \end{cases}$

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_x \ -\sum_{i=1}^m \log \ f_V(y_i - a_i^\top x)$$

Maximum likelihood

Uniform: $V \sim \mathcal{U}(-c, c)$ $\qquad f_V(v) = \begin{cases} \frac{1}{2c} & , \ |v| \leq c \\ 0 & , \ \text{oth.} \end{cases}$

$$\widehat{x}_{\mathsf{ML}} \in \arg\min_{x} \ -\sum_{i=1}^{m} \log \ f_V(y_i - a_i^\top x)$$

$$= \arg\min_{x} \quad -\textstyle\sum_{i=1}^{m} \log \ \frac{1}{2c}$$

$$\text{s.t.} \qquad |y_i - a_i^\top x| \leq c \,, \quad i = 1, \ldots, m \,.$$

Maximum likelihood

Uniform: $V \sim \mathcal{U}(-c, c)$ $\qquad f_V(v) = \begin{cases} \frac{1}{2c} & , \; |v| \leq c \\ 0 & , \; \text{oth.} \end{cases}$

$$\widehat{x}_{\mathsf{ML}} \in \underset{x}{\arg\min} \; -\sum_{i=1}^{m} \log \, f_V(y_i - a_i^\top x)$$

$$= \underset{x}{\arg\min} \quad -\textstyle\sum_{i=1}^{m} \log \, \frac{1}{2c}$$

$$\text{s.t.} \qquad |y_i - a_i^\top x| \leq c \,, \quad i = 1, \ldots, m \,.$$

$$= \text{find} \quad x$$

$$\text{s.t.} \quad \|y - Ax\|_\infty \leq c$$

Maximum likelihood

Uniform: $V \sim \mathcal{U}(-c, c)$ $\qquad f_V(v) = \begin{cases} \frac{1}{2c} & , \ |v| \leq c \\ 0 & , \ \text{oth.} \end{cases}$

$$\begin{aligned}
\widehat{x}_{\mathsf{ML}} \ &\in \ \underset{x}{\arg\min} \ -\sum_{i=1}^{m} \log \, f_V(y_i - a_i^\top x) \\
&= \ \underset{x}{\arg\min} \quad -\textstyle\sum_{i=1}^{m} \log \, \frac{1}{2c} \\
& \qquad\quad \text{s.t.} \qquad |y_i - a_i^\top x| \leq c \, , \quad i = 1, \ldots, m \, . \\
&= \ \text{find} \quad x \\
& \qquad \text{s.t.} \quad \|y - Ax\|_\infty \leq c
\end{aligned}$$

feasibility problem (convex)

Example with a discrete RV

Example with a discrete RV

$Y \in \mathbb{N}$: # of traffic accidents in a given period

Example with a discrete RV

$Y \in \mathbb{N}$: $\#$ of traffic accidents in a given period

$$Y \sim \text{Poisson}(\mu)$$

Example with a discrete RV

$Y \in \mathbb{N}$: # of traffic accidents in a given period

$$Y \sim \text{Poisson}(\mu) \qquad\qquad \mathbb{P}(Y = k) = \frac{e^{-\mu}\mu^k}{k\,!}\,, \quad k = 0, 1, \ldots$$

Example with a discrete RV

$Y \in \mathbb{N}$: $\#$ of traffic accidents in a given period

$$Y \sim \text{Poisson}(\mu) \qquad\qquad \mathbb{P}(Y = k) = \frac{e^{-\mu}\mu^k}{k\,!}, \quad k = 0, 1, \ldots$$

Assumption: μ depends on vector of explanatory variables $U \in \mathbb{R}^n$ as

Example with a discrete RV

$Y \in \mathbb{N}$: # of traffic accidents in a given period

$$Y \sim \text{Poisson}(\mu) \qquad\qquad \mathbb{P}(Y = k) = \frac{e^{-\mu}\mu^k}{k\,!}\,, \quad k = 0, 1, \ldots$$

Assumption: μ depends on vector of explanatory variables $U \in \mathbb{R}^n$ as

$$\mu = a^\top U + b\,, \qquad\qquad a \in \mathbb{R}^n\,,\ b \in \mathbb{R}$$

Example with a discrete RV

$Y \in \mathbb{N}$: # of traffic accidents in a given period

$$Y \sim \text{Poisson}(\mu) \qquad\qquad \mathbb{P}(Y = k) = \frac{e^{-\mu}\mu^k}{k\,!}\,, \quad k = 0, 1, \ldots$$

Assumption: μ depends on vector of explanatory variables $U \in \mathbb{R}^n$ as

$$\mu = a^\top U + b\,, \qquad\qquad a \in \mathbb{R}^n\,,\ b \in \mathbb{R}$$

e.g., $U_1 =$ traffic flow during the period, $U_2 =$ rainfall, ...

Example with a discrete RV

$Y \in \mathbb{N}$: $\#$ of traffic accidents in a given period

$$Y \sim \text{Poisson}(\mu) \qquad\qquad \mathbb{P}(Y = k) = \frac{e^{-\mu}\mu^k}{k\,!}\,, \quad k = 0, 1, \ldots$$

Assumption: μ depends on vector of explanatory variables $U \in \mathbb{R}^n$ as

$$\mu = a^\top U + b\,, \qquad\qquad a \in \mathbb{R}^n\,,\ b \in \mathbb{R}$$

e.g., $U_1 =$ traffic flow during the period, $U_2 =$ rainfall, $\ldots$

Goal:

Given m independent observations $\left\{\left(U^{(i)}, Y^{(i)}\right)\right\}_{i=1}^{m}$, estimate a and b.

Example with a discrete RV

Joint probability mass function: $p_{YU}(y, u\,;\, a, b) = \frac{e^{-(a^\top u+b)}\,(a^\top u+b)^y}{y\,!}$

Example with a discrete RV

Joint probability mass function: $p_{YU}(y, u\,;\, a, b) = \frac{e^{-(a^\top u+b)}\,(a^\top u+b)^y}{y\,!}$

ML estimator:

Example with a discrete RV

Joint probability mass function: $p_{YU}(y, u\,;\, a, b) = \frac{e^{-(a^\top u+b)}\,(a^\top u+b)^y}{y\,!}$

ML estimator:

$$\big(a_{\mathsf{ML}}, b_{\mathsf{ML}}\big) \;\in\; \underset{a,b}{\arg\min} \;\; -\sum_{i=1}^{m} \log\, p_{YU}(y_i, u_i\,;\, a, b)$$

Example with a discrete RV

Joint probability mass function: $p_{YU}(y, u\,;\, a, b) = \frac{e^{-(a^\top u+b)}\,(a^\top u+b)^y}{y\,!}$

ML estimator:

$$\begin{aligned}
\big(a_{\mathsf{ML}}, b_{\mathsf{ML}}\big) \;&\in\; \operatorname*{arg\,min}_{a,b} \;\; -\sum_{i=1}^{m} \log\, p_{YU}(y_i, u_i\,;\, a, b) \\
&=\; \operatorname*{arg\,min}_{a,b} \;\; -\sum_{i=1}^{m} \log\, \frac{e^{-(a^\top u^{(i)}+b)}\,(a^\top u^{(i)}+b)^{y^{(i)}}}{y^{(i)}\,!}
\end{aligned}$$

Example with a discrete RV

Joint probability mass function: $p_{YU}(y, u\,;\, a, b) = \frac{e^{-(a^\top u + b)}\,(a^\top u + b)^y}{y\,!}$

ML estimator:

$$
\begin{aligned}
\big(a_{\mathsf{ML}}, b_{\mathsf{ML}}\big) \;&\in\; \operatorname*{arg\,min}_{a,b} \;\; -\sum_{i=1}^m \log\, p_{YU}(y_i, u_i\,;\, a, b) \\
&=\; \operatorname*{arg\,min}_{a,b} \;\; -\sum_{i=1}^m \log\, \frac{e^{-(a^\top u^{(i)} + b)}\,(a^\top u^{(i)} + b)^{y^{(i)}}}{y^{(i)}\,!} \\
&=\; \operatorname*{arg\,min}_{a,b} \;\; \sum_{i=1}^m a^\top u^{(i)} + b \;-\; y^{(i)} \log\big(a^\top u^{(i)} + b\big)
\end{aligned}
$$

Example with a discrete RV

Joint probability mass function: $p_{YU}(y, u\,;\,a, b) = \frac{e^{-(a^\top u + b)}\,(a^\top u + b)^y}{y\,!}$

ML estimator:

$$\begin{aligned}
\big(a_{\mathsf{ML}}, b_{\mathsf{ML}}\big) \;&\in\; \operatorname*{arg\,min}_{a,b} \; -\sum_{i=1}^{m} \log\, p_{YU}(y_i, u_i\,;\,a, b) \\
&=\; \operatorname*{arg\,min}_{a,b} \; -\sum_{i=1}^{m} \log\, \frac{e^{-(a^\top u^{(i)} + b)}\,(a^\top u^{(i)} + b)^{y^{(i)}}}{y^{(i)}\,!} \\
&=\; \operatorname*{arg\,min}_{a,b} \; \sum_{i=1}^{m} \underbrace{a^\top u^{(i)} + b}_{\text{affine}} \;-\; y^{(i)} \log\big(a^\top u^{(i)} + b\big)
\end{aligned}$$

Example with a discrete RV

Joint probability mass function: $p_{YU}(y, u\,;\, a, b) = \frac{e^{-(a^\top u+b)}\,(a^\top u+b)^y}{y\,!}$

ML estimator:

$$\begin{aligned}
\big(a_{\mathsf{ML}}, b_{\mathsf{ML}}\big) \;&\in\; \operatorname*{arg\,min}_{a,b} \;\; -\sum_{i=1}^m \log\, p_{YU}(y_i, u_i\,;\, a, b) \\
&=\; \operatorname*{arg\,min}_{a,b} \;\; -\sum_{i=1}^m \log\, \frac{e^{-(a^\top u^{(i)}+b)}\,(a^\top u^{(i)}+b)^{y^{(i)}}}{y^{(i)}\,!} \\
&=\; \operatorname*{arg\,min}_{a,b} \;\; \sum_{i=1}^m \underbrace{a^\top u^{(i)} + b}_{\text{affine}} \; \underbrace{-\; y^{(i)} \log\big(a^\top u^{(i)} + b\big)}_{\text{convex}}
\end{aligned}$$

Example with a discrete RV

Joint probability mass function: $p_{YU}(y, u\,;\, a, b) = \frac{e^{-(a^\top u+b)}\,(a^\top u+b)^y}{y\,!}$

ML estimator:

$$\begin{aligned}
\big(a_{\mathsf{ML}}, b_{\mathsf{ML}}\big) \;&\in\; \underset{a,b}{\arg\min}\; -\sum_{i=1}^m \log\, p_{YU}(y_i, u_i\,;\, a, b) \\
&=\; \underset{a,b}{\arg\min}\; -\sum_{i=1}^m \log\, \frac{e^{-(a^\top u^{(i)}+b)}\,(a^\top u^{(i)}+b)^{y^{(i)}}}{y^{(i)}\,!} \\
&=\; \underset{a,b}{\arg\min}\; \sum_{i=1}^m \underbrace{a^\top u^{(i)} + b}_{\text{affine}}\; \underbrace{-\; y^{(i)} \log\big(a^\top u^{(i)} + b\big)}_{\text{convex}}
\end{aligned}$$

Convex

Maximum a posteriori (MAP)

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate *no longer a parameter*

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate *no longer a parameter*

prior knowledge: $f_X(x)$

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate *no longer a parameter*

prior knowledge: $f_X(x)$

MAP estimator:

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate *no longer a parameter*

prior knowledge: $f_X(x)$

MAP estimator:

$$\widehat{x}_{\text{MAP}} \in \arg\max_x \; f_{X|Y}(x|y)$$

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate *no longer a parameter*

prior knowledge: $f_X(x)$

MAP estimator:

$$\widehat{x}_{\text{MAP}} \in \arg\max_x \; f_{X|Y}(x|y) \overset{\text{Bayes}}{=} \arg\max_x \; \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$$

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate *no longer a parameter*

prior knowledge: $f_X(x)$

MAP estimator:

$$\widehat{x}_{\mathsf{MAP}} \in \arg\max_x \; f_{X|Y}(x|y) \overset{\text{Bayes}}{=} \arg\max_x \; \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$$

$$= \arg\max_x \; f_{Y|X}(y|x) f_X(x)$$

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate *no longer a parameter*

prior knowledge: $f_X(x)$

MAP estimator:

$$\widehat{x}_{\mathsf{MAP}} \in \arg\max_x \ f_{X|Y}(x|y) \overset{\text{Bayes}}{=} \arg\max_x \ \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$$

$$= \arg\max_x \ f_{Y|X}(y|x) f_X(x)$$

$$= \arg\max_x \ \log\left(f_{Y|X}(y|x) f_X(x)\right)$$

Maximum a posteriori (MAP)

$Y \in \mathbb{R}^m$: random vector w/ density $f_Y(y)$

$X \in \mathbb{R}^n$: random vector to estimate *no longer a parameter*

prior knowledge: $f_X(x)$

MAP estimator:

$$\widehat{x}_{\mathsf{MAP}} \in \arg\max_x \; f_{X|Y}(x|y) \overset{\text{Bayes}}{=} \arg\max_x \; \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)}$$

$$= \arg\max_x \; f_{Y|X}(y|x) f_X(x)$$

$$= \arg\max_x \; \log\big(f_{Y|X}(y|x) f_X(x)\big)$$

$$= \arg\min_x \; -\log f_{Y|X}(y|x) - \log f_X(x)$$

Maximum a posteriori (MAP)

Example: linear measurement model

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i, \;\; i = 1, \ldots, m$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i, \;\; i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c,\, c)$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i,\ \ i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c,\, c)$

  $f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x)$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i, \;\; i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c,\, c)$

$$f_{Y_1 \cdots Y_m|X}(y_1, \ldots, y_m|x) \;=\; \prod_{i=1}^{m} f_{Y_i|X}(y_i|x)$$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i,\ \ i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c,\, c)$

$$f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \,=\, \prod_{i=1}^m f_{Y_i|X}(y_i|x) \,=\, \prod_{i=1}^m \mathcal{U}\big(a_i^\top x - c\,,\, a_i^\top x + c\big)$$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i,\;\; i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c,\, c)$

$$f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \;=\; \prod_{i=1}^{m} f_{Y_i|X}(y_i|x) \;=\; \prod_{i=1}^{m} \mathcal{U}\big(a_i^\top x - c\,,\, a_i^\top x + c\big)$$

- Random vector to estimate in $\mathbb{R}^n$: $X \sim \mathcal{N}\big(\overline{x}, \Sigma\big)$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i,\ \ i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c,\, c)$

$$f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \;=\; \prod_{i=1}^{m} f_{Y_i|X}(y_i|x) \;=\; \prod_{i=1}^{m} \mathcal{U}\big(a_i^\top x - c\,,\, a_i^\top x + c\big)$$

- Random vector to estimate in $\mathbb{R}^n$: $X \sim \mathcal{N}\big(\overline{x}, \Sigma\big)$ $\big($known $\overline{x}, \Sigma\big)$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i, \;\; i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c,\, c)$

$$f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \;=\; \prod_{i=1}^{m} f_{Y_i|X}(y_i|x) \;=\; \prod_{i=1}^{m} \mathcal{U}\big(a_i^\top x - c\,,\, a_i^\top x + c\big)$$

- Random vector to estimate in $\mathbb{R}^n$: $X \sim \mathcal{N}\big(\overline{x}, \Sigma\big)$ $\big($known $\overline{x}, \Sigma\big)$

$$f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\Big(-\frac{1}{2} (x - \overline{x})^\top \Sigma^{-1} (x - \overline{x}) \Big)$$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i, \;\; i = 1, \ldots, m \;,\; \&\; V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c, \, c)$

$$f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \;=\; \prod_{i=1}^{m} f_{Y_i|X}(y_i|x) \;=\; \prod_{i=1}^{m} \mathcal{U}\big(a_i^\top x - c \,,\, a_i^\top x + c\big)$$

- Random vector to estimate in $\mathbb{R}^n$: $X \sim \mathcal{N}\big(\overline{x}, \Sigma\big) \qquad \big($known $\overline{x}, \Sigma\big)$

$$f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\Big(-\frac{1}{2} (x - \overline{x})^\top \Sigma^{-1} (x - \overline{x}) \Big)$$

$$\widehat{x}_{\mathsf{MAP}} \;\in\; \arg\min_{x} \quad - \log \, f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \;-\; \log \, f_X(x)$$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i, \;\; i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c, c)$

$$f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \; = \; \prod_{i=1}^{m} f_{Y_i | X}(y_i | x) \; = \; \prod_{i=1}^{m} \mathcal{U}\big(a_i^\top x - c \, , \, a_i^\top x + c\big)$$

- Random vector to estimate in $\mathbb{R}^n$: $X \sim \mathcal{N}\big(\overline{x}, \Sigma\big)$ $\qquad \big($known $\overline{x}, \Sigma\big)$

$$f_X(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x - \overline{x})^\top \Sigma^{-1} (x - \overline{x}) \Big)$$

$$\begin{aligned} \widehat{x}_{\mathsf{MAP}} \;\; &\in \;\; \arg\min_x \quad -\log \, f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \, - \, \log \, f_X(x) \\ &= \quad \arg\min_x \quad \tfrac{1}{2}(x - \overline{x})^\top \Sigma^{-1} (x - \overline{x}) \\ &\qquad\quad \text{s.t.} \qquad \|Ax - y\|_\infty \le c \end{aligned}$$

Maximum a posteriori (MAP)

Example: linear measurement model

- Observations: $Y_i = a_i^\top X + V_i,\;\; i = 1, \ldots, m$, & $V_i \overset{\text{iid}}{=} V \sim \mathcal{U}(-c,\, c)$

$$f_{Y_1 \cdots Y_m | X}(y_1, \ldots, y_m | x) \;=\; \prod_{i=1}^{m} f_{Y_i|X}(y_i|x) \;=\; \prod_{i=1}^{m} \mathcal{U}\big(a_i^\top x - c\,,\, a_i^\top x + c\big)$$

- Random vector to estimate in $\mathbb{R}^n$: $\;X \sim \mathcal{N}\big(\overline{x}, \Sigma\big)$ $\qquad$ $\big($known $\overline{x}, \Sigma\big)$

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x - \overline{x})^\top \Sigma^{-1}(x - \overline{x})\Big)$$

$$\widehat{x}_{\mathsf{MAP}} \;\in\; \underset{x}{\arg\min} \quad -\log\, f_{Y_1 \cdots Y_m|X}(y_1, \ldots, y_m|x) \,-\, \log\, f_X(x)$$

$$= \quad \underset{x}{\arg\min} \quad \tfrac{1}{2}(x - \overline{x})^\top \Sigma^{-1}(x - \overline{x})$$

$$\text{s.t.} \qquad \|Ax - y\|_\infty \leq c$$

Convex

Maximum a posteriori (MAP)

LASSO:

Maximum a posteriori (MAP)

LASSO:

- Observations: $Y_i = a_i^\top X + V_i,\ i = 1, \ldots, m,$ with $V_i \stackrel{\text{iid}}{=} V \sim \mathcal{N}(0,\, \sigma^2)$

Maximum a posteriori (MAP)

LASSO:

- Observations: $Y_i = a_i^\top X + V_i,\ i = 1, \ldots, m,$ with $V_i \overset{\text{iid}}{=} V \sim \mathcal{N}(0,\, \sigma^2)$

- Random vector to estimate $X \in \mathbb{R}^n$ has iid Laplacian entries $\mathcal{L}(0, b)$

Maximum a posteriori (MAP)

LASSO:

- Observations: $Y_i = a_i^\top X + V_i,\ i = 1, \ldots, m,$ with $V_i \overset{\text{iid}}{=} V \sim \mathcal{N}(0,\, \sigma^2)$
- Random vector to estimate $X \in \mathbb{R}^n$ has iid Laplacian entries $\mathcal{L}(0, b)$

$$\widehat{x}_{\mathsf{MAP}} \ \in \ \arg\min_x \ \ \frac{1}{2}\big\|y - Ax\big\|_2^2 + \frac{\sigma^2}{b}\|x\|_1$$

Maximum a posteriori (MAP)

LASSO:

- Observations: $Y_i = a_i^\top X + V_i,\ i = 1, \ldots, m,$ with $V_i \overset{\text{iid}}{=} V \sim \mathcal{N}(0,\, \sigma^2)$
- Random vector to estimate $X \in \mathbb{R}^n$ has iid Laplacian entries $\mathcal{L}(0, b)$

$$\widehat{x}_{\text{MAP}} \ \in\ \arg\min_x \ \ \frac{1}{2}\left\|y - Ax\right\|_2^2 + \frac{\sigma^2}{b}\|x\|_1$$

Known as *basis pursuit denoising* or *LASSO*

Maximum a posteriori (MAP)

LASSO:

- Observations: $Y_i = a_i^\top X + V_i,\ i = 1, \ldots, m,$ with $V_i \overset{\text{iid}}{=} V \sim \mathcal{N}(0,\, \sigma^2)$

- Random vector to estimate $X \in \mathbb{R}^n$ has iid Laplacian entries $\mathcal{L}(0, b)$

$$\widehat{x}_{\text{MAP}} \ \in \ \arg\min_x \ \ \frac{1}{2}\left\|y - Ax\right\|_2^2 + \frac{\sigma^2}{b}\|x\|_1$$

Known as *basis pursuit denoising* or *LASSO* (Convex)

Nonparametric estimation

Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Example:

X: discrete RV taking values on 100 equidistant points in $[-1, 1]$

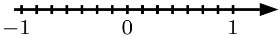
Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Example:

X: discrete RV taking values on 100 equidistant points in $[-1, 1]$



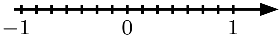
Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Example:

X: discrete RV taking values on 100 equidistant points in $[-1,\, 1]$



We require

$\mathbb{E}\, X \in [-0.1,\, 0.1]$

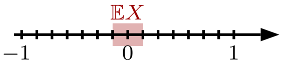
Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Example:

X: discrete RV taking values on 100 equidistant points in $[-1, 1]$



We require

$\mathbb{E}\, X \in [-0.1,\, 0.1]$

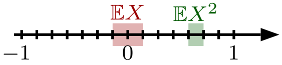
Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Example:

X: discrete RV taking values on 100 equidistant points in $[-1, 1]$



We require

$\mathbb{E}\, X \in [-0.1,\, 0.1]$ $\qquad$ $\mathbb{E}\, X^2 \in [0.5,\, 0.6]$

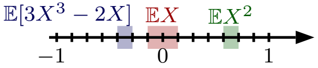
Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Example:

X: discrete RV taking values on 100 equidistant points in $[-1, 1]$



We require

$\mathbb{E}\,X \in [-0.1, 0.1]$ $\qquad$ $\mathbb{E}\,X^2 \in [0.5, 0.6]$

$\mathbb{E}\big[3X^3 - 2X\big] \in [-0.3, -0.2]$

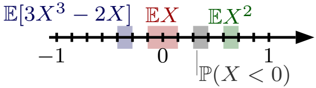
Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Example:

X: discrete RV taking values on 100 equidistant points in $[-1, 1]$



We require

$\mathbb{E}\, X \in [-0.1,\, 0.1]$ $\qquad$ $\mathbb{E}\, X^2 \in [0.5,\, 0.6]$

$\mathbb{E}\big[3X^3 - 2X\big] \in [-0.3,\, -0.2]$

$\mathbb{P}\big(X < 0\big) \in [0.3,\, 0.4]$

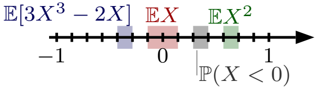
Nonparametric estimation

ML and MAP allow estimating parameters of given distributions

What about estimating non-canonical distributions?

Example:

X: discrete RV taking values on 100 equidistant points in $[-1,\, 1]$



We require

$\mathbb{E}\,X \in [-0.1,\, 0.1]$ $\qquad$ $\mathbb{E}\,X^2 \in [0.5,\, 0.6]$

$\mathbb{E}\big[3X^3 - 2X\big] \in [-0.3,\, -0.2]$

$\mathbb{P}\big(X < 0\big) \in [0.3,\, 0.4]$

Find a distribution satisfying these constraints & with maximum entropy

Nonparametric estimation

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*

$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \,:\, x \geq 0_n\,,\; 1_n^\top x = 1 \right\}$$

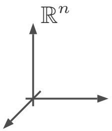
Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*

$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \,:\, x \geq 0_n \,,\; 1_n^\top x = 1 \right\}$$

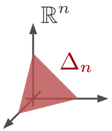


Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*



$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \,:\, x \geq 0_n \,,\; 1_n^\top x = 1 \right\}$$

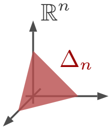
Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*

$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \,:\, x \geq 0_n\,,\ 1_n^\top x = 1 \right\}$$



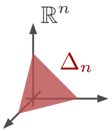
$$\mathbb{E}\, X \in [-0.1,\, 0.1]$$

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*



$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \,:\, x \geq 0_n \,,\; 1_n^\top x = 1 \right\}$$

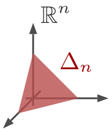
$$\mathbb{E}\, X \in [-0.1,\, 0.1] \iff -0.1 \leq \alpha^\top p \leq 0.1$$

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*



$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \, : \, x \geq 0_n \, , \; 1_n^\top x = 1 \right\}$$

$$\mathbb{E}\, X \in [-0.1, \, 0.1] \iff -0.1 \leq \alpha^\top p \leq 0.1$$

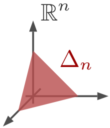
$$\mathbb{E}\, X^2 \in [0.5, \, 0.6]$$

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*



$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \,:\, x \geq 0_n\,,\; 1_n^\top x = 1 \right\}$$

$$\mathbb{E}\, X \in [-0.1,\, 0.1] \iff -0.1 \leq \alpha^\top p \leq 0.1$$

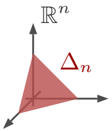
$$\mathbb{E}\, X^2 \in [0.5,\, 0.6] \iff 0.5 \leq \beta^\top p \leq 0.6 \qquad \beta_i = \alpha_i^2$$

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*



$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \,:\, x \geq 0_n \,,\; 1_n^\top x = 1 \right\}$$

$$\mathbb{E}\, X \in [-0.1,\, 0.1] \iff -0.1 \leq \alpha^\top p \leq 0.1$$

$$\mathbb{E}\, X^2 \in [0.5,\, 0.6] \iff 0.5 \leq \beta^\top p \leq 0.6 \qquad \beta_i = \alpha_i^2$$

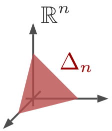
$$\mathbb{E}\big[3X^3 - 2X\big] \in [-0.3,\, -0.2] \iff -0.3 \leq \gamma^\top p \leq -0.2 \qquad \gamma_i = 3\alpha_i^2 - 2\alpha_i$$

Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*



$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \,:\, x \geq 0_n \,,\; 1_n^\top x = 1 \right\}$$

$$\mathbb{E}\, X \in [-0.1,\, 0.1] \iff -0.1 \leq \alpha^\top p \leq 0.1$$

$$\mathbb{E}\, X^2 \in [0.5,\, 0.6] \iff 0.5 \leq \beta^\top p \leq 0.6 \qquad \beta_i = \alpha_i^2$$

$$\mathbb{E}\left[3X^3 - 2X\right] \in [-0.3,\, -0.2] \iff -0.3 \leq \gamma^\top p \leq -0.2 \qquad \gamma_i = 3\alpha_i^2 - 2\alpha_i$$

$$\mathbb{P}\left(X < 0\right) \in [0.3,\, 0.4]$$

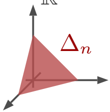
Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*

$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \, : \, x \geq 0_n \, , \; 1_n^\top x = 1 \right\}$$



$$\mathbb{E}\, X \in [-0.1,\, 0.1] \iff -0.1 \leq \alpha^\top p \leq 0.1$$

$$\mathbb{E}\, X^2 \in [0.5,\, 0.6] \iff 0.5 \leq \beta^\top p \leq 0.6 \qquad \beta_i = \alpha_i^2$$

$$\mathbb{E}\big[3X^3 - 2X\big] \in [-0.3,\, -0.2] \iff -0.3 \leq \gamma^\top p \leq -0.2 \qquad \gamma_i = 3\alpha_i^2 - 2\alpha_i$$

$$\mathbb{P}\big(X < 0\big) \in [0.3,\, 0.4] \iff 0.3 \leq \sigma^\top p \leq 0.4 \qquad \sigma_i = 1 \text{ if } i < 50$$

$$\sigma_i = 0 \text{ if } i \geq 50$$

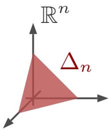
Nonparametric estimation

$\alpha = (-1, -0.9798, \ldots, 0.9798, 1) \in \mathbb{R}^{100}$ values that X takes

$p \in \mathbb{R}^{100}$ such that $p_i = \mathbb{P}(X = \alpha_i)$

p: pmf $\implies$ belongs to the *probability simplex*

$$p \in \Delta_n := \left\{ x \in \mathbb{R}^n \ : \ x \geq 0_n \, , \ 1_n^\top x = 1 \right\}$$



$$\mathbb{E}\, X \in [-0.1,\, 0.1] \iff -0.1 \leq \alpha^\top p \leq 0.1$$

$$\mathbb{E}\, X^2 \in [0.5,\, 0.6] \iff 0.5 \leq \beta^\top p \leq 0.6 \qquad \beta_i = \alpha_i^2$$

$$\mathbb{E}\big[3X^3 - 2X\big] \in [-0.3,\, -0.2] \iff -0.3 \leq \gamma^\top p \leq -0.2 \qquad \gamma_i = 3\alpha_i^2 - 2\alpha_i$$

$$\mathbb{P}\big(X < 0\big) \in [0.3,\, 0.4] \iff 0.3 \leq \sigma^\top p \leq 0.4 \qquad \sigma_i = 1 \text{ if } i < 50$$

$$\sigma_i = 0 \text{ if } i \geq 50$$

All constraints are *linear inequalities* in p!

Nonparametric estimation

Maximize entropy

Nonparametric estimation

Maximize entropy $\quad = \quad$ minimize *negative entropy:*

Nonparametric estimation

Maximize entropy $=$ minimize *negative entropy:*

$$-H(p) := \sum_{i=1}^{n} p_i \log p_i$$

Nonparametric estimation

Maximize entropy $=$ minimize *negative entropy:*

$$-H(p) := \sum_{i=1}^{n} p_i \log\, p_i$$

Convex because $\dfrac{\mathrm{d}^2}{\mathrm{d}x^2}\, x \log\, x = \frac{1}{x} > 0$, for all $x > 0$

Nonparametric estimation

Maximize entropy $=$ minimize *negative entropy:*

$$-H(p) := \sum_{i=1}^{n} p_i \log p_i$$

Convex because $\dfrac{\mathrm{d}^2}{\mathrm{d}x^2} x \log x = \frac{1}{x} > 0$, for all $x > 0$

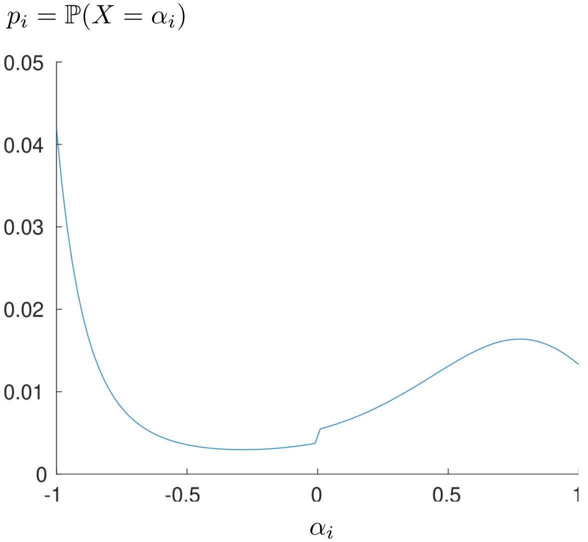
Optimization problem: *(convex)*

$$\begin{aligned}
\underset{p \in \mathbb{R}^{100}}{\text{minimize}} \quad & \textstyle\sum_{i=1}^{n} p_i \log p_i \\
\text{subject to} \quad & -0.1 \le \alpha^\top p \le 0.1 \\
& 0.5 \le \beta^\top p \le 0.6 \\
& -0.3 \le \gamma^\top p \le -0.2 \\
& 0.3 \le \sigma^\top p \le 0.4
\end{aligned}$$

Nonparametric estimation

```
n = 100;
alpha = linspace(-1,1,n)';
cvx_begin
    variable p(n,1);
    minimize( -sum( entr( p ) ) );
    subject to
        p >= 0;
        ones(1, n)*p == 1;
        -0.1 <= alpha'*p <= 0.1;
        0.5 <= (alpha.^2)'*p <= 0.6;
        -0.3 <= (3*alpha.^3 - 2*alpha)'*p <= -0.2;
        0.3 <= (alpha < 0)'*p <= 0.4;
cvx_end
```

Nonparametric estimation



Hypothesis testing & optimal detection

Hypothesis testing & optimal detection

X : discrete random variable w/ probability mass function (pmf) p_θ

Hypothesis testing & optimal detection

X : discrete random variable w/ probability mass function (pmf) p_θ

$$X \in \{1, \ldots, n\}$$

Hypothesis testing & optimal detection

X : discrete random variable w/ probability mass function (pmf) p_θ

$$X \in \{1, \ldots, n\}$$

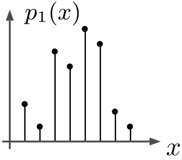
$$\theta \in \{1, \ldots, m\}$$

Hypothesis testing & optimal detection

X : discrete random variable w/ probability mass function (pmf) p_θ

$X \in \{1, \ldots, n\}$

$\theta \in \{1, \ldots, m\}$

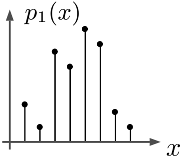


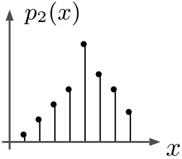
Hypothesis testing & optimal detection

X : discrete random variable w/ probability mass function (pmf) p_θ

$$X \in \{1, \ldots, n\}$$

$$\theta \in \{1, \ldots, m\}$$



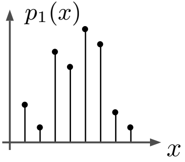


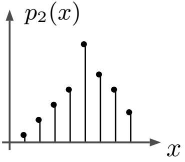
Hypothesis testing & optimal detection

X : discrete random variable w/ probability mass function (pmf) p_θ

$X \in \{1, \ldots, n\}$

$\theta \in \{1, \ldots, m\}$





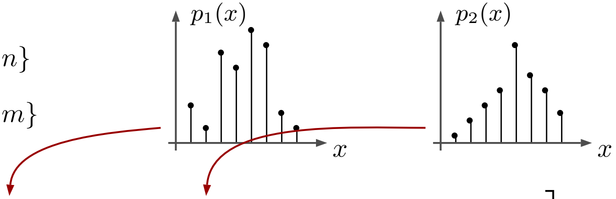
$$P = \begin{bmatrix} \mathbb{P}(X=1 \,|\, \theta=1) & \mathbb{P}(X=1 \,|\, \theta=2) & \cdots & \mathbb{P}(X=1 \,|\, \theta=m) \\ \mathbb{P}(X=2 \,|\, \theta=1) & \mathbb{P}(X=2 \,|\, \theta=2) & \cdots & \mathbb{P}(X=2 \,|\, \theta=m) \\ \vdots & \vdots & & \vdots \\ \mathbb{P}(X=n \,|\, \theta=1) & \mathbb{P}(X=n \,|\, \theta=2) & \cdots & \mathbb{P}(X=n \,|\, \theta=m) \end{bmatrix} \in \mathbb{R}^{n \times m}$$

Hypothesis testing & optimal detection

X : discrete random variable w/ probability mass function (pmf) p_θ

$$X \in \{1, \ldots, n\}$$

$$\theta \in \{1, \ldots, m\}$$



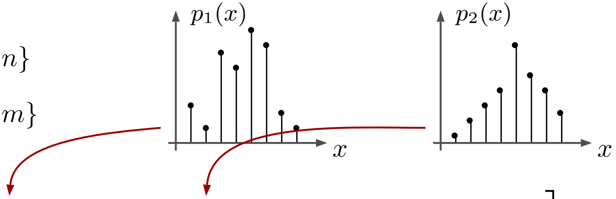
$$P = \begin{bmatrix} \mathbb{P}(X = 1 \,|\, \theta = 1) & \mathbb{P}(X = 1 \,|\, \theta = 2) & \cdots & \mathbb{P}(X = 1 \,|\, \theta = m) \\ \mathbb{P}(X = 2 \,|\, \theta = 1) & \mathbb{P}(X = 2 \,|\, \theta = 2) & \cdots & \mathbb{P}(X = 2 \,|\, \theta = m) \\ \vdots & \vdots & & \vdots \\ \mathbb{P}(X = n \,|\, \theta = 1) & \mathbb{P}(X = n \,|\, \theta = 2) & \cdots & \mathbb{P}(X = n \,|\, \theta = m) \end{bmatrix} \in \mathbb{R}^{n \times m}$$

Hypothesis testing & optimal detection

X : discrete random variable w/ probability mass function (pmf) p_θ

$X \in \{1, \ldots, n\}$

$\theta \in \{1, \ldots, m\}$



$$P = \begin{bmatrix} \mathbb{P}(X=1\,|\,\theta=1) & \mathbb{P}(X=1\,|\,\theta=2) & \cdots & \mathbb{P}(X=1\,|\,\theta=m) \\ \mathbb{P}(X=2\,|\,\theta=1) & \mathbb{P}(X=2\,|\,\theta=2) & \cdots & \mathbb{P}(X=2\,|\,\theta=m) \\ \vdots & \vdots & & \vdots \\ \mathbb{P}(X=n\,|\,\theta=1) & \mathbb{P}(X=n\,|\,\theta=2) & \cdots & \mathbb{P}(X=n\,|\,\theta=m) \end{bmatrix} \in \mathbb{R}^{n\times m}$$

Goal: Estimate θ based on an observation of X

Detector:

Detector: $\Psi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, m\}$

Detector: $\Psi : \underbrace{\{1, \ldots, n\}}_{X} \longrightarrow \underbrace{\{1, \ldots, m\}}_{\theta}$

Detector: $\Psi : \underbrace{\{1, \ldots, n\}}_{X} \longrightarrow \underbrace{\{1, \ldots, m\}}_{\theta}$

Example: maximum likelihood detector

Detector: $\Psi \, : \, \underbrace{\{1, \ldots, n\}}_{X} \longrightarrow \underbrace{\{1, \ldots, m\}}_{\theta}$

Example: maximum likelihood detector

$$\widehat{\theta}_{\mathsf{ML}} \;=\; \Psi_{\mathsf{ML}}(x) \;=\; \arg\max_{j} \; \mathbb{P}\big(X = x \,|\, \theta = j\big)$$

Detector: $\Psi : \underbrace{\{1, \ldots, n\}}_{X} \longrightarrow \underbrace{\{1, \ldots, m\}}_{\theta}$

Example: maximum likelihood detector

$$\widehat{\theta}_{\mathsf{ML}} = \Psi_{\mathsf{ML}}(x) = \arg\max_{j} \mathbb{P}\big(X = x \,|\, \theta = j\big)$$

Randomized detector: Random variable $\widehat{\theta}$ whose pmf depends on X

Detector: $\Psi : \underbrace{\{1, \ldots, n\}}_{X} \longrightarrow \underbrace{\{1, \ldots, m\}}_{\theta}$

Example: maximum likelihood detector

$$\widehat{\theta}_{\mathsf{ML}} = \Psi_{\mathsf{ML}}(x) = \arg\max_{j} \ \mathbb{P}\big(X = x \,|\, \theta = j\big)$$

Randomized detector: Random variable $\widehat{\theta}$ whose pmf depends on X

$$T = \begin{bmatrix} \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = n) \\ \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = n) \\ \vdots & \vdots & & \vdots \\ \mathbb{P}(\widehat{\theta} = m \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = m \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = m \,|\, X = n) \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Detector: $\Psi \,:\, \underbrace{\{1,\ldots,n\}}_{X} \longrightarrow \underbrace{\{1,\ldots,m\}}_{\theta}$

Example: maximum likelihood detector

$$\widehat{\theta}_{\mathsf{ML}} \;=\; \Psi_{\mathsf{ML}}(x) \;=\; \arg\max_{j} \; \mathbb{P}\big(X = x \,|\, \theta = j\big)$$

Randomized detector: Random variable $\widehat{\theta}$ whose pmf depends on X

$$T = \begin{bmatrix} \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = n) \\ \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = n) \\ \vdots & \vdots & & \vdots \\ \mathbb{P}(\widehat{\theta} = m \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = m \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = m \,|\, X = n) \end{bmatrix} \in \mathbb{R}^{m \times n}$$

Each column $t_i \in \mathbb{R}^m$ satisfies $1_m^\top t_i = 1$, $t_i \geq 0$

Detector: $\Psi : \underbrace{\{1,\ldots,n\}}_{X} \longrightarrow \underbrace{\{1,\ldots,m\}}_{\theta}$

Example: maximum likelihood detector

$$\widehat{\theta}_{\mathsf{ML}} = \Psi_{\mathsf{ML}}(x) = \arg\max_j \ \mathbb{P}\big(X = x \,|\, \theta = j\big)$$

Randomized detector: Random variable $\widehat{\theta}$ whose pmf depends on X

$$T = \begin{bmatrix} \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = n) \\ \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = n) \\ \vdots & \vdots & & \vdots \\ \mathbb{P}(\widehat{\theta} = m \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = m \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = m \,|\, X = n) \end{bmatrix} \in \mathbb{R}^{m\times n}$$

Each column $t_i \in \mathbb{R}^m$ satisfies $1_m^\top t_i = 1$, $t_i \geq 0$

If each t_i is a canonical vector $(0,\ldots,1,\ldots,0)$, then T is deterministic

Detection probability matrix:

Detection probability matrix: $D := TP \in \mathbb{R}^{m \times m}$

Detection probability matrix: $D := TP \in \mathbb{R}^{m \times m}$

$$D_{ij} = \sum_{k=1}^{n} \mathbb{P}\big(\widehat{\theta} = i \,|\, X = k\big) \cdot \mathbb{P}\big(X = k \,|\, \theta = j\big)$$

Detection probability matrix: $D := TP \in \mathbb{R}^{m\times m}$

$$D_{ij} = \sum_{k=1}^{n} \mathbb{P}\big(\widehat{\theta} = i \,|\, X = k\big) \cdot \mathbb{P}\big(X = k \,|\, \theta = j\big)$$

$$= \mathbb{P}\big(\, \widehat{\theta} = i \mid \theta = j \,\big)$$

Detection probability matrix: $D := TP \in \mathbb{R}^{m \times m}$

$$D_{ij} = \sum_{k=1}^{n} \mathbb{P}\big(\widehat{\theta} = i \,|\, X = k\big) \cdot \mathbb{P}\big(X = k \,|\, \theta = j\big)$$

$$= \mathbb{P}\big(\underbrace{\widehat{\theta} = i}_{\text{guess}} \mid \underbrace{\theta = j}_{\text{true}}\big)$$

Detection probability matrix: $D := TP \in \mathbb{R}^{m \times m}$

$$\begin{aligned} D_{ij} &= \sum_{k=1}^{n} \mathbb{P}\big(\widehat{\theta} = i \,|\, X = k\big) \cdot \mathbb{P}\big(X = k \,|\, \theta = j\big) \\ &= \mathbb{P}\big(\underbrace{\widehat{\theta} = i}_{\text{guess}} \mid \underbrace{\theta = j}_{\text{true}}\big) \end{aligned}$$

Goal: Design T such that the D_{ij}'s, for $i \neq j$, are as small as possible

Detection probability matrix: $D := TP \in \mathbb{R}^{m\times m}$

$$D_{ij} = \sum_{k=1}^{n} \mathbb{P}\big(\widehat{\theta} = i \,|\, X = k\big) \cdot \mathbb{P}\big(X = k \,|\, \theta = j\big)$$

$$= \mathbb{P}\big(\underbrace{\widehat{\theta} = i}_{\text{guess}} \mid \underbrace{\theta = j}_{\text{true}}\big)$$

Goal: Design T such that the D_{ij}'s, for $i \neq j$, are as small as possible

Multi-objective optimization

Minimax detector

Minimax detector

Note that

$$\mathbb{P}\big(\widehat{\theta} \neq j \,|\, \theta = j\big)$$

Minimax detector

Note that

$$\mathbb{P}\big(\widehat{\theta} \neq j \mid \theta = j\big) = 1 - \mathbb{P}\big(\widehat{\theta} = j \mid \theta = j\big)$$

Minimax detector

Note that

$$\mathbb{P}\big(\widehat{\theta} \neq j \,|\, \theta = j\big) = 1 - \mathbb{P}\big(\widehat{\theta} = j \,|\, \theta = j\big) = 1 - D_{jj}$$

Minimax detector

Note that

$$\mathbb{P}\big(\widehat{\theta} \neq j \,|\, \theta = j\big) = 1 - \mathbb{P}\big(\widehat{\theta} = j \,|\, \theta = j\big) = 1 - D_{jj}$$

Minimax detector:

Minimax detector

Note that

$$\mathbb{P}\big(\widehat{\theta} \neq j \mid \theta = j\big) = 1 - \mathbb{P}\big(\widehat{\theta} = j \mid \theta = j\big) = 1 - D_{jj}$$

Minimax detector:

$$\begin{array}{ll} \underset{T \in \mathbb{R}^{m \times n}}{\text{minimize}} & \displaystyle\max_{j=i,\ldots,m} \; 1 - D_{jj}(T) \\ \text{subject to} & T^\top 1_m = 1_n \,, \;\; T \geq 0 \end{array}$$

Minimax detector

Note that

$$\mathbb{P}\big(\widehat{\theta} \neq j \,|\, \theta = j\big) = 1 - \mathbb{P}\big(\widehat{\theta} = j \,|\, \theta = j\big) = 1 - D_{jj}$$

Minimax detector:

$$\begin{array}{ll} \underset{T \in \mathbb{R}^{m \times n}}{\text{minimize}} & \displaystyle\max_{j=i,\ldots,m} \; 1 - D_{jj}(T) \\ \text{subject to} & T^\top 1_m = 1_n \,, \;\; T \geq 0 \end{array}$$

$$\Longleftrightarrow \qquad \begin{array}{ll} \underset{T \in \mathbb{R}^{m \times n}}{\text{minimize}} & \displaystyle\max_{j=i,\ldots,m} \; 1 - \mathrm{tr}\big(TP\, e_j e_j^\top\big) \\ \text{subject to} & T^\top 1_m = 1_n \,, \;\; T \geq 0 \end{array}$$

Minimax detector

Note that

$$\mathbb{P}\big(\widehat{\theta} \neq j \,|\, \theta = j\big) = 1 - \mathbb{P}\big(\widehat{\theta} = j \,|\, \theta = j\big) = 1 - D_{jj}$$

Minimax detector:

$$\begin{array}{ll} \underset{T \in \mathbb{R}^{m \times n}}{\text{minimize}} & \underset{j=i,\ldots,m}{\max} \ 1 - D_{jj}(T) \\ \text{subject to} & T^\top 1_m = 1_n \,, \ \ T \geq 0 \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{T \in \mathbb{R}^{m \times n}}{\text{minimize}} & \underset{j=i,\ldots,m}{\max} \ 1 - \mathsf{tr}\big(TP\, e_j e_j^\top\big) \\ \text{subject to} & T^\top 1_m = 1_n \,, \ \ T \geq 0 \end{array}$$

Convex

Pareto-optimal detectors

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization:

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}(\widehat{\theta} = i \mid \theta = j)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

$$
\begin{array}{ll}
\underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m} \sum_{j=1}^{m} W_{ij} D_{ij}(T) \\
\text{subject to} & T^\top 1_m = 1_n \,, \;\; T \geq 0
\end{array}
$$

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

$$
\begin{array}{ll}
\underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\
\text{subject to} & T^\top 1_m = 1_n \,, \;\; T \geq 0
\end{array}
$$

By varying W we can find *Pareto optimal* solutions

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,, \;\; T \geq 0 \end{array}$$

By varying W we can find *Pareto optimal* solutions

Observations:

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,, \;\; T \geq 0 \end{array}$$

By varying W we can find *Pareto optimal* solutions

Observations:

- Constraints $\iff$ $1_m^\top t_i = 1, \; t_i \geq 0,$ for $i = 1, \ldots, n$

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m} \sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n \,, \;\; T \geq 0 \end{array}$$

By varying W we can find *Pareto optimal* solutions

Observations:

- Constraints $\iff$ $1_m^\top t_i = 1$, $t_i \geq 0$, for $i = 1, \ldots, n$
- Objective $\iff$ $\text{tr}\big(W^\top T P\big)$

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

$$
\begin{array}{ll}
\underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m} \sum_{j=1}^{m} W_{ij} D_{ij}(T) \\
\text{subject to} & T^\top 1_m = 1_n \,, \;\; T \geq 0
\end{array}
$$

By varying W we can find *Pareto optimal* solutions

Observations:

- Constraints $\iff$ $1_m^\top t_i = 1, \; t_i \geq 0,$ for $i = 1, \ldots, n$
- Objective $\iff$ $\text{tr}\big(W^\top T P\big) = \text{tr}\big((W P^\top)^\top T\big)$

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,, \;\; T \geq 0 \end{array}$$

By varying W we can find *Pareto optimal* solutions

Observations:

- Constraints $\iff$ $1_m^\top t_i = 1,\; t_i \geq 0$, for $i = 1, \ldots, n$
- Objective $\iff$ $\mathrm{tr}\big(W^\top T P\big) = \mathrm{tr}\big((W P^\top)^\top T\big) = \sum_{i=1}^{n} c_i^\top t_i$

Pareto-optimal detectors

Recall that we want to minimize all $D_{ij} = \mathbb{P}\big(\widehat{\theta} = i \,|\, \theta = j\big)$, for $i \neq j$

Scalarization: Fix $W \in \mathbb{R}^{m \times m}$ w/ $W_{ii} = 0$, and $W_{ij} > 0$ for $i \neq j$

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m} \sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n \,, \;\; T \geq 0 \end{array}$$

By varying W we can find *Pareto optimal* solutions

Observations:

- Constraints $\iff$ $1_m^\top t_i = 1,\; t_i \geq 0,$ for $i = 1, \ldots, n$
- Objective $\iff$ $\text{tr}\big(W^\top T P\big) = \text{tr}\big((W P^\top)^\top T\big) = \sum_{i=1}^n c_i^\top t_i$

  where c_i: ith column of $C := W P^\top$

Pareto-optimal detectors

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,, \;\; T \geq 0 \end{array}$$

Pareto-optimal detectors

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,, \;\; T \geq 0 \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \sum_{i=1}^{n} c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,, \;\; t_i \geq 0\,, \;\; i = 1, \ldots, n \end{array}$$

Pareto-optimal detectors

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,,\ \ T \geq 0 \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \sum_{i=1}^{n} c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\ \ t_i \geq 0\,,\ \ i = 1, \ldots, n \end{array}$$

Decouples into n independent optimization problems: for column i,

Pareto-optimal detectors

$$
\begin{array}{ll}
\underset{T}{\text{minimize}} & \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\
\text{subject to} & T^\top 1_m = 1_n\,,\;\; T \geq 0
\end{array}
$$

$$
\Longleftrightarrow \qquad
\begin{array}{ll}
\underset{t_1,\ldots,t_n}{\text{minimize}} & \sum_{i=1}^{n} c_i^\top t_i \\
\text{subject to} & 1_m^\top t_i = 1\,,\;\; t_i \geq 0\,,\;\; i = 1,\ldots,n
\end{array}
$$

Decouples into n independent optimization problems: for column i,

$$
\begin{array}{ll}
\underset{t_i}{\text{minimize}} & c_i^\top t_i \\
\text{subject to} & 1_m^\top t_i = 1\,,\;\; t_i \geq 0
\end{array}
$$

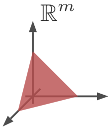
Pareto-optimal detectors

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,,\;\; T \geq 0 \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \sum_{i=1}^{n} c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\;\; t_i \geq 0\,,\;\; i = 1,\ldots,n \end{array}$$

Decouples into n independent optimization problems: for column i,

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\;\; t_i \geq 0 \end{array}$$



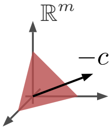
Pareto-optimal detectors

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij}D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,,\ \ T \geq 0 \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{t_1,\dots,t_n}{\text{minimize}} & \sum_{i=1}^{n} c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\ \ t_i \geq 0\,,\ \ i = 1,\dots,n \end{array}$$

Decouples into n independent optimization problems: for column i,

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\ \ t_i \geq 0 \end{array}$$



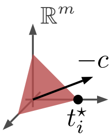
Pareto-optimal detectors

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,,\;\; T \geq 0 \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \sum_{i=1}^{n} c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\;\; t_i \geq 0\,,\;\; i = 1,\ldots,n \end{array}$$

Decouples into n independent optimization problems: for column i,

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\;\; t_i \geq 0 \end{array}$$



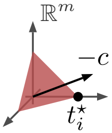
Pareto-optimal detectors

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij}D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,,\;\; T \geq 0 \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \sum_{i=1}^{n} c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\;\; t_i \geq 0\,,\;\; i = 1,\ldots,n \end{array}$$

Decouples into n independent optimization problems: for column i,

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\;\; t_i \geq 0 \end{array}$$



$t_i^\star = (0,\ldots,1,\ldots,0)$

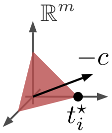
Pareto-optimal detectors

$$\begin{array}{ll} \underset{T}{\text{minimize}} & \sum_{i=1}^{m}\sum_{j=1}^{m} W_{ij} D_{ij}(T) \\ \text{subject to} & T^\top 1_m = 1_n\,,\ \ T \geq 0 \end{array}$$

$$\iff \qquad \begin{array}{ll} \underset{t_1,\ldots,t_n}{\text{minimize}} & \sum_{i=1}^{n} c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\ \ t_i \geq 0\,,\ \ i = 1,\ldots,n \end{array}$$

Decouples into n independent optimization problems: for column i,

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \\ \text{subject to} & 1_m^\top t_i = 1\,,\ \ t_i \geq 0 \end{array}$$



$t_i^\star = (0,\ldots,1,\ldots,0) \qquad \implies \qquad T$ is deterministic

Example: binary hypothesis testing

Example: binary hypothesis testing

X is generated by either p_1 $(\theta = 1)$ or p_2 $(\theta = 2)$

Example: binary hypothesis testing

X is generated by either p_1 $(\theta = 1)$ or p_2 $(\theta = 2)$

$$P = \begin{bmatrix} \mathbb{P}(X = 1 \,|\, \theta = 1) & \mathbb{P}(X = 1 \,|\, \theta = 2) \\ \mathbb{P}(X = 2 \,|\, \theta = 1) & \mathbb{P}(X = 2 \,|\, \theta = 2) \\ \vdots & \vdots \\ \mathbb{P}(X = n \,|\, \theta = 1) & \mathbb{P}(X = n \,|\, \theta = 2) \end{bmatrix} \in \mathbb{R}^{n \times 2}$$

Example: binary hypothesis testing

X is generated by either p_1 $(\theta = 1)$ or p_2 $(\theta = 2)$

$$P = \begin{bmatrix} \mathbb{P}(X = 1 \,|\, \theta = 1) & \mathbb{P}(X = 1 \,|\, \theta = 2) \\ \mathbb{P}(X = 2 \,|\, \theta = 1) & \mathbb{P}(X = 2 \,|\, \theta = 2) \\ \vdots & \vdots \\ \mathbb{P}(X = n \,|\, \theta = 1) & \mathbb{P}(X = n \,|\, \theta = 2) \end{bmatrix} \in \mathbb{R}^{n \times 2}$$

$$T = \begin{bmatrix} \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = n) \\ \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = n) \end{bmatrix} \in \mathbb{R}^{2 \times n}$$

Example: binary hypothesis testing

X is generated by either p_1 $(\theta = 1)$ or p_2 $(\theta = 2)$

$$P = \begin{bmatrix} \mathbb{P}(X = 1 \,|\, \theta = 1) & \mathbb{P}(X = 1 \,|\, \theta = 2) \\ \mathbb{P}(X = 2 \,|\, \theta = 1) & \mathbb{P}(X = 2 \,|\, \theta = 2) \\ \vdots & \vdots \\ \mathbb{P}(X = n \,|\, \theta = 1) & \mathbb{P}(X = n \,|\, \theta = 2) \end{bmatrix} \in \mathbb{R}^{n \times 2}$$

$$T = \begin{bmatrix} \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = n) \\ \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = n) \end{bmatrix} \in \mathbb{R}^{2 \times n}$$

$$D = \begin{bmatrix} \mathbb{P}\big(\widehat{\theta} = 1 \,|\, \theta = 1\big) & \mathbb{P}\big(\widehat{\theta} = 1 \,|\, \theta = 2\big) \\ \mathbb{P}\big(\widehat{\theta} = 2 \,|\, \theta = 1\big) & \mathbb{P}\big(\widehat{\theta} = 2 \,|\, \theta = 2\big) \end{bmatrix}$$

Example: binary hypothesis testing

X is generated by either p_1 $(\theta = 1)$ or p_2 $(\theta = 2)$

$$P = \begin{bmatrix} \mathbb{P}(X = 1 \,|\, \theta = 1) & \mathbb{P}(X = 1 \,|\, \theta = 2) \\ \mathbb{P}(X = 2 \,|\, \theta = 1) & \mathbb{P}(X = 2 \,|\, \theta = 2) \\ \vdots & \vdots \\ \mathbb{P}(X = n \,|\, \theta = 1) & \mathbb{P}(X = n \,|\, \theta = 2) \end{bmatrix} \in \mathbb{R}^{n \times 2}$$

$$T = \begin{bmatrix} \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 1 \,|\, X = n) \\ \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 1) & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = 2) & \cdots & \mathbb{P}(\widehat{\theta} = 2 \,|\, X = n) \end{bmatrix} \in \mathbb{R}^{2 \times n}$$

$$D = \begin{bmatrix} \mathbb{P}\big(\widehat{\theta} = 1 \,|\, \theta = 1\big) & \mathbb{P}\big(\widehat{\theta} = 1 \,|\, \theta = 2\big) \\ \mathbb{P}\big(\widehat{\theta} = 2 \,|\, \theta = 1\big) & \mathbb{P}\big(\widehat{\theta} = 2 \,|\, \theta = 2\big) \end{bmatrix} = \begin{bmatrix} 1 - P_{\mathsf{fp}} & P_{\mathsf{fn}} \\ P_{\mathsf{fp}} & 1 - P_{\mathsf{fn}} \end{bmatrix}$$

Pareto-optimal detectors:

Pareto-optimal detectors:

Select $W \in \mathbb{R}^{2\times 2}$, set $C = WP^\top$, and column i of T_{PO} solves

Pareto-optimal detectors:

Select $W \in \mathbb{R}^{2\times 2}$, set $C = WP^\top$, and column i of T_{PO} solves

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \qquad\qquad (c_i\text{: column } i \text{ of } C) \\ \text{subject to} & 1_2^\top t_i = 1\,,\ \ t_i \geq 0 \end{array}$$

Pareto-optimal detectors:

Select $W \in \mathbb{R}^{2\times 2}$, set $C = WP^\top$, and column i of T_{PO} solves

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \qquad\qquad (c_i\text{: column } i \text{ of } C) \\ \text{subject to} & 1_2^\top t_i = 1\,,\ \ t_i \geq 0 \end{array}$$

$$C = \begin{bmatrix} 0 & W_{12} \\ W_{21} & 0 \end{bmatrix} \begin{bmatrix} \mathbb{P}(X = 1 \,|\, \theta = 1) & \cdots & \mathbb{P}(X = n \,|\, \theta = 1) \\ \mathbb{P}(X = 1 \,|\, \theta = 2) & \cdots & \mathbb{P}(X = n \,|\, \theta = 2) \end{bmatrix}$$

Pareto-optimal detectors:

Select $W \in \mathbb{R}^{2\times 2}$, set $C = WP^\top$, and column i of T_{PO} solves

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \qquad\qquad (c_i\text{: column } i \text{ of } C) \\ \text{subject to} & 1_2^\top t_i = 1\,, \;\; t_i \geq 0 \end{array}$$

$$C = \begin{bmatrix} 0 & W_{12} \\ W_{21} & 0 \end{bmatrix} \begin{bmatrix} \mathbb{P}(X=1\,|\,\theta=1) & \cdots & \mathbb{P}(X=n\,|\,\theta=1) \\ \mathbb{P}(X=1\,|\,\theta=2) & \cdots & \mathbb{P}(X=n\,|\,\theta=2) \end{bmatrix}$$

$$= \begin{bmatrix} W_{12}\,\mathbb{P}(X=1\,|\,\theta=2) & \cdots & W_{12}\,\mathbb{P}(X=n\,|\,\theta=2) \\ W_{21}\,\mathbb{P}(X=1\,|\,\theta=1) & \cdots & W_{21}\,\mathbb{P}(X=n\,|\,\theta=1) \end{bmatrix}$$

Pareto-optimal detectors:

Select $W \in \mathbb{R}^{2\times 2}$, set $C = WP^\top$, and column i of T_{PO} solves

$$\begin{array}{ll} \underset{t_i}{\text{minimize}} & c_i^\top t_i \qquad\qquad (c_i\text{: column } i \text{ of } C) \\ \text{subject to} & 1_2^\top t_i = 1\,, \;\; t_i \ge 0 \end{array}$$

$$C = \begin{bmatrix} 0 & W_{12} \\ W_{21} & 0 \end{bmatrix} \begin{bmatrix} \mathbb{P}(X=1\,|\,\theta=1) & \cdots & \mathbb{P}(X=n\,|\,\theta=1) \\ \mathbb{P}(X=1\,|\,\theta=2) & \cdots & \mathbb{P}(X=n\,|\,\theta=2) \end{bmatrix}$$

$$= \begin{bmatrix} W_{12}\,\mathbb{P}(X=1\,|\,\theta=2) & \cdots & W_{12}\,\mathbb{P}(X=n\,|\,\theta=2) \\ W_{21}\,\mathbb{P}(X=1\,|\,\theta=1) & \cdots & W_{21}\,\mathbb{P}(X=n\,|\,\theta=1) \end{bmatrix}$$

Therefore,

$$c_i^\top t_i = t_{1i}\,W_{12}\,\mathbb{P}(X=i\,|\,\theta=2) + t_{2i}\,W_{21}\,\mathbb{P}(X=i\,|\,\theta=1)$$

- If $W_{12}\,\mathbb{P}(X = i \,|\, \theta = 2) < W_{21}\,\mathbb{P}(X = i \,|\, \theta = 1)$,

- If $W_{12}\,\mathbb{P}(X = i \,|\, \theta = 2) < W_{21}\,\mathbb{P}(X = i \,|\, \theta = 1)$,

$$(t^\star_{1i}\,,\, t^\star_{2i}) = (1\,,\, 0)$$

- If $W_{12}\,\mathbb{P}(X=i\,|\,\theta=2) < W_{21}\,\mathbb{P}(X=i\,|\,\theta=1)$,

$$(t^\star_{1i}\,,\,t^\star_{2i}) = (1\,,\,0) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 1$$

- If $W_{12}\,\mathbb{P}(X=i\,|\,\theta=2) < W_{21}\,\mathbb{P}(X=i\,|\,\theta=1)$,

$$(t^\star_{1i}\,,\,t^\star_{2i}) = (1\,,\,0) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 1$$

- If $W_{12}\,\mathbb{P}(X=i\,|\,\theta=2) \geq W_{21}\,\mathbb{P}(X=i\,|\,\theta=1)$

- If $W_{12}\,\mathbb{P}(X = i \,|\, \theta = 2) < W_{21}\,\mathbb{P}(X = i \,|\, \theta = 1)$,

$$(t^\star_{1i}\,,\, t^\star_{2i}) = (1\,,\, 0) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 1$$

- If $W_{12}\,\mathbb{P}(X = i \,|\, \theta = 2) \geq W_{21}\,\mathbb{P}(X = i \,|\, \theta = 1)$,

$$(t^\star_{1i}\,,\, t^\star_{2i}) = (0\,,\, 1)$$

- If $W_{12}\,\mathbb{P}(X=i\,|\,\theta=2) < W_{21}\,\mathbb{P}(X=i\,|\,\theta=1)$,

$$(t^\star_{1i}\,,\,t^\star_{2i}) = (1\,,\,0) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 1$$

- If $W_{12}\,\mathbb{P}(X=i\,|\,\theta=2) \geq W_{21}\,\mathbb{P}(X=i\,|\,\theta=1)$,

$$(t^\star_{1i}\,,\,t^\star_{2i}) = (0\,,\,1) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 2$$

- If $W_{12}\,\mathbb{P}(X=i\,|\,\theta=2) < W_{21}\,\mathbb{P}(X=i\,|\,\theta=1)$,

$$(t^\star_{1i}\,,\,t^\star_{2i}) = (1\,,\,0) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 1$$

- If $W_{12}\,\mathbb{P}(X=i\,|\,\theta=2) \geq W_{21}\,\mathbb{P}(X=i\,|\,\theta=1)$,

$$(t^\star_{1i}\,,\,t^\star_{2i}) = (0\,,\,1) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 2$$

This is the *likelihood-ratio test:*

- If $W_{12}\,\mathbb{P}(X = i \,|\, \theta = 2) < W_{21}\,\mathbb{P}(X = i \,|\, \theta = 1)$,

$$(t^\star_{1i}\,,\, t^\star_{2i}) = (1\,,\,0) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 1$$

- If $W_{12}\,\mathbb{P}(X = i \,|\, \theta = 2) \geq W_{21}\,\mathbb{P}(X = i \,|\, \theta = 1)$,

$$(t^\star_{1i}\,,\, t^\star_{2i}) = (0\,,\,1) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 2$$

This is the *likelihood-ratio test:* Decide $\theta = 2$ if

- If $W_{12}\,\mathbb{P}(X = i \,|\, \theta = 2) < W_{21}\,\mathbb{P}(X = i \,|\, \theta = 1)$,

$$(t^\star_{1i}\,,\,t^\star_{2i}) = (1\,,\,0) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 1$$

- If $W_{12}\,\mathbb{P}(X = i \,|\, \theta = 2) \geq W_{21}\,\mathbb{P}(X = i \,|\, \theta = 1)$,

$$(t^\star_{1i}\,,\,t^\star_{2i}) = (0\,,\,1) \qquad \Longrightarrow \qquad \widehat{\theta}_{\mathsf{PO}} = 2$$

This is the *likelihood-ratio test:* Decide $\theta = 2$ if

$$\frac{\mathbb{P}(X = i \,|\, \theta = 2)}{\mathbb{P}(X = i \,|\, \theta = 1)} \geq \frac{W_{21}}{W_{12}} =: \alpha$$

Neyman-Pearson lemma:

For each $\alpha > 0$, the likelihood-ratio test yields a (deterministic) Pareto-optimal detector.

Deterministic vs randomized detectors

$$P = \begin{bmatrix} 0.70 & 0.10 \\ 0.20 & 0.10 \\ 0.05 & 0.70 \\ 0.05 & 0.10 \end{bmatrix}$$

Deterministic vs randomized detectors

$$P = \begin{bmatrix} 0.70 & 0.10 \\ 0.20 & 0.10 \\ 0.05 & 0.70 \\ 0.05 & 0.10 \end{bmatrix}$$

Varying α yields 3 different Pareto-optimal detectors (excl. extremes):

Deterministic vs randomized detectors

$$P = \begin{bmatrix} 0.70 & 0.10 \\ 0.20 & 0.10 \\ 0.05 & 0.70 \\ 0.05 & 0.10 \end{bmatrix}$$

Varying α yields 3 different Pareto-optimal detectors (excl. extremes):

$$T_{\mathsf{PO}}^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \qquad T_{\mathsf{PO}}^{(2)} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad T_{\mathsf{PO}}^{(3)} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Deterministic vs randomized detectors

$$P = \begin{bmatrix} 0.70 & 0.10 \\ 0.20 & 0.10 \\ 0.05 & 0.70 \\ 0.05 & 0.10 \end{bmatrix}$$

Varying α yields 3 different Pareto-optimal detectors (excl. extremes):

$$T_{\mathsf{PO}}^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \qquad T_{\mathsf{PO}}^{(2)} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad T_{\mathsf{PO}}^{(3)} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Minimax detector (random):

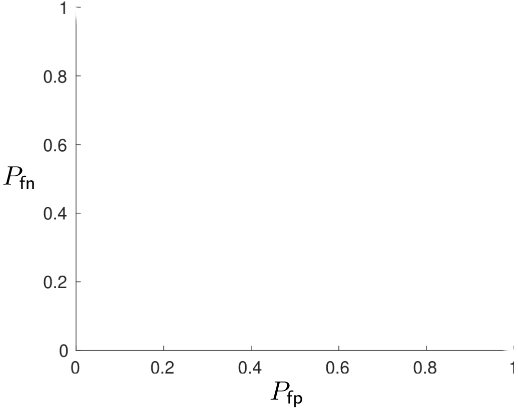
$$T_{\mathsf{MM}} = \begin{bmatrix} 1 & 2/3 & 0 & 0 \\ 0 & 1/3 & 1 & 1 \end{bmatrix}$$

Deterministic vs randomized detectors

Receiver operating characteristic (ROC)

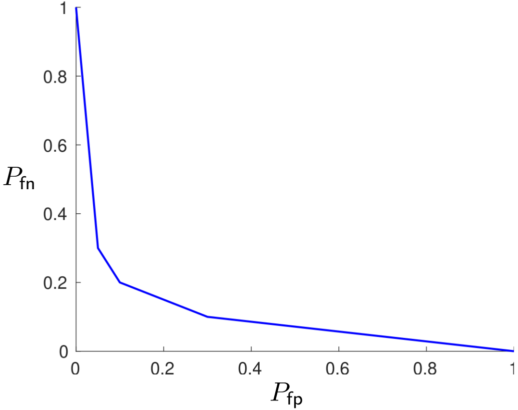
Deterministic vs randomized detectors

Receiver operating characteristic (ROC)



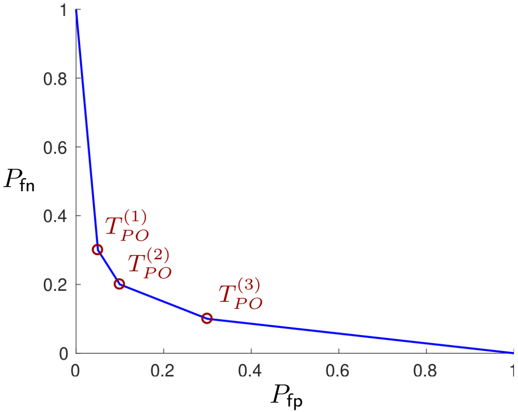
Deterministic vs randomized detectors

Receiver operating characteristic (ROC)



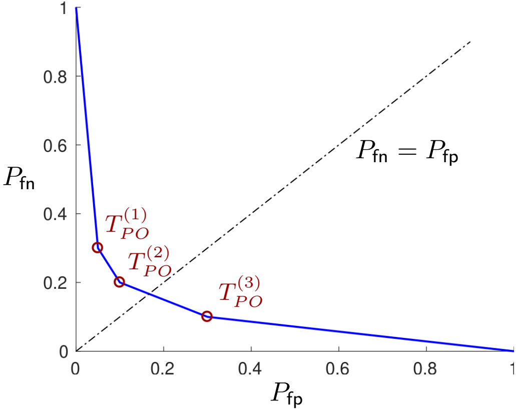
Deterministic vs randomized detectors

Receiver operating characteristic (ROC)



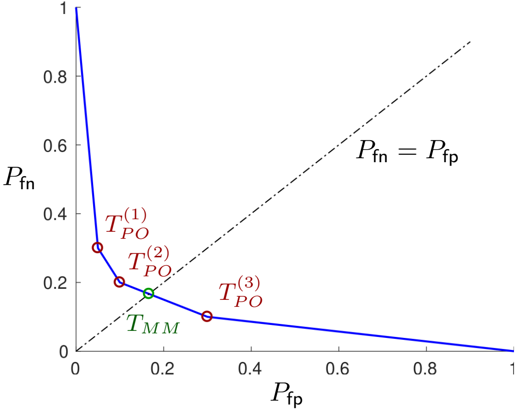
Deterministic vs randomized detectors

Receiver operating characteristic (ROC)



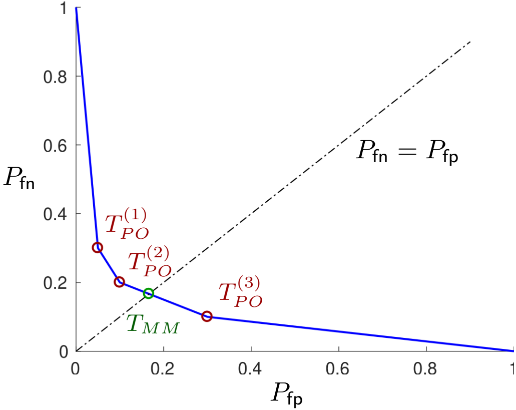
Deterministic vs randomized detectors

Receiver operating characteristic (ROC)



Deterministic vs randomized detectors

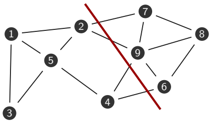
Receiver operating characteristic (ROC)



Minimax estimator has $(P_{\mathsf{fp}}, P_{\mathsf{fn}}) = (\frac{1}{6}, \frac{1}{6})$ and outperforms any deterministic estimator

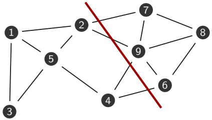
Conclusions

Conclusions



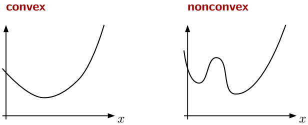
- Optimization problems arise in many areas

Conclusions

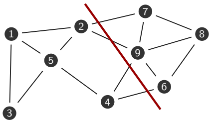


- Optimization problems arise in many areas

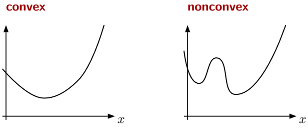
- Essential to distinguish easy *(convex)* from hard *(nonconvex)* probs



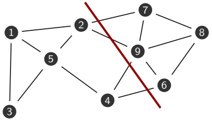
Conclusions



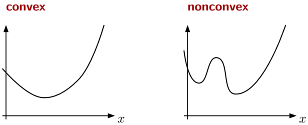
- Optimization problems arise in many areas
- Essential to distinguish easy *(convex)* from hard *(nonconvex)* probs
- Use CVX/CVXPY/Convex for small-scale problems



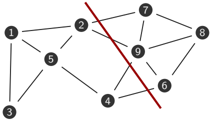
Conclusions



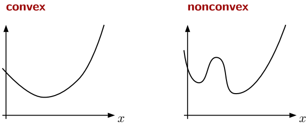
- Optimization problems arise in many areas
- Essential to distinguish easy *(convex)* from hard *(nonconvex)* probs
- Use CVX/CVXPY/Convex for small-scale problems
- *Didn't cover* optimality conditions & theory



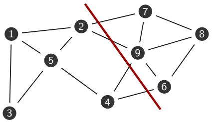
Conclusions



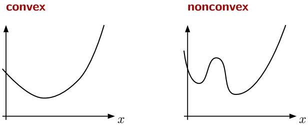
- Optimization problems arise in many areas
- Essential to distinguish easy *(convex)* from hard *(nonconvex)* probs
- Use CVX/CVXPY/Convex for small-scale problems
- *Didn't cover* optimality conditions & theory
- *Didn't cover* optimization algorithms



Conclusions



- Optimization problems arise in many areas
- Essential to distinguish easy *(convex)* from hard *(nonconvex)* probs
- Use CVX/CVXPY/Convex for small-scale problems
- *Didn't cover* optimality conditions & theory
- *Didn't cover* optimization algorithms

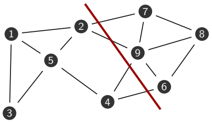


- Statistical estimation:

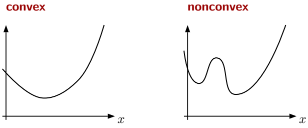
  Find parameters of distributions (ML/MAP)

  Entire distributions (nonparametric)

Conclusions



- Optimization problems arise in many areas
- Essential to distinguish easy *(convex)* from hard *(nonconvex)* probs
- Use CVX/CVXPY/Convex for small-scale problems
- *Didn't cover* optimality conditions & theory
- *Didn't cover* optimization algorithms



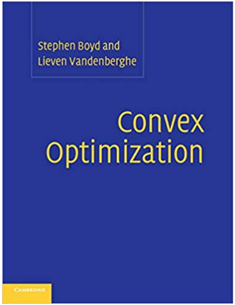
- Statistical estimation:

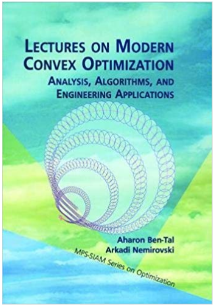
  Find parameters of distributions (ML/MAP)

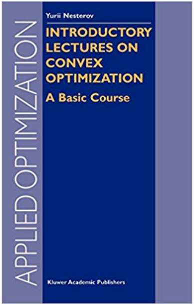
  Entire distributions (nonparametric)

- Multiple hypothesis testing via optimization

References and Resources







Lectures:

- web.stanford.edu/∼boyd/cvxbook/
- users.isr.ist.utl.pt/∼jxavier/NonlinearOptimization18799-2018
- www.seas.ucla.edu/∼vandenbe/ee236c