Disciplined Machine Learning

Hao Zhu & Joschka Boedecker

1. Introduction

- mathematical modeling
- model descriptions
- disciplined machine learning
- examples
- course goals and topics
- mathematical optimization
- convex optimization
- brief history of convex optimization
- nonlinear optimization

Mathematical modeling

build (mathematical) **models**

- that can
  - learn from data
  - make predictions or decisions based on the learned knowledge

- usually represented as functions that map some *inputs* (or *features*) to outputs (or *responses*, *observations*)

- characterized by a set of *parameters* that determine their behavior

Model descriptions

natural language descriptions define a model using math symbols and texts, *e.g.*,

> Given observations $a_1, \ldots, a_m \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$, a linear regression model assumes $y_i \approx a_i^T x$ and chooses $x \in \mathbf{R}^n$ to minimize the sum of squared errors.

algorithmic descriptions define a model in terms of a procedure to compute the model's parameters, *e.g.*,

---

Linear regression.

given $A = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}^T \in \mathbf{R}^{m \times n}$, $y \in \mathbf{R}^m$

return $x := (A^T A)^{-1} A^T y$

---

inverse problem descriptions define a model in terms of the associated fitting problem, $e.g.$,

$$\text{minimize} \quad \sum_{i=1}^{m} \left(a_i^T x - y_i\right)^2$$

with variable $x \in \mathbf{R}^n$ and data $a_1, \ldots, a_m \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$ (describes a linear regression model)

Disciplined machine learning

$$\begin{array}{ll} \text{minimize} & f_0(x) + \phi(x) \\ \text{subject to} & x \in C \end{array}$$

- $x \in \mathbf{R}^n$ is the model *parameter*
- $f_0 \colon \mathbf{R}^n \to \mathbf{R}$ is the *cost* or *loss function*
- $\phi \colon \mathbf{R}^n \to \mathbf{R}$ is the *regularization function* or *regularizer*
- $C \subseteq \mathbf{R}^n$ is the *constraint set*

- f_0 represents the primary objective of the model, such as fitting or prediction error
- ϕ and C represent prior information of required model properties

Least squares

$$\text{minimize} \quad \|Ax - b\|_2^2$$

- variable $x \in \mathbf{R}^n$
- data $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$
- a convex optimization problem
- analytical solution $x^\star = (A^T A)^{-1} A^T b$

- easy to recognize and a mature technology
- reliable and efficient algorithms and software

regularized least squares

$$\text{minimize} \quad \|Ax - b\|_2^2 + \lambda \|x\|_2^2$$

- $\|x\|_2^2$ called ridge or Tikhonov regularization
- analytical solution $x^\star = (A^T A + \lambda I)^{-1} A^T b$
- induce 'small' solutions $x^\star$

$$\text{minimize} \quad \|Ax - b\|_2^2 + \lambda \|x\|_1$$

- $\|x\|_1$ called lasso regularization
- no analytical solution, but is a convex optimization problem
- can be solved efficiently by numerical methods
- induce sparse solutions $x^\star$ (many components are zero)

(linear inequality) constrained least squares

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & Cx \preceq d \end{array}$$

- a convex optimization problem; no analytical solution
- $A = I$ and $b = 0$, $\preceq$ takes equality $\implies$ least norm problem

$$\begin{array}{ll} \text{minimize} & \|x\|_2^2 \\ \text{subject to} & Cx = d \end{array}$$

  finds the minimum (ℓ_2-) norm solution of the linear system $Cx = d$

- $C = -I$ and $d = 0$ $\implies$ nonnegative least squares problem

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & x \succeq 0 \end{array}$$

  finds the best nonnegative solution of the linear system $Ax = b$

Principal component analysis (PCA)

$$\begin{array}{ll} \text{minimize} & \|Z - A\|_F^2 \\ \text{subject to} & \mathbf{rank}\, Z \leq k \end{array}$$

- variable $Z \in \mathbf{R}^{m \times n}$
- data (matrix) $A \in \mathbf{R}^{m \times n}$, rank k ($\ll n$)
- $\|\cdot\|_F$ is the Frobenius norm, *i.e.*,

$$\|U\|_F = \left(\mathbf{tr}(U^T U)\right)^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} U_{ij}^2\right)^{1/2}$$

- finds the best rank-k approximation of $A \in \mathbf{R}^{m \times n}$

factorized formulation of PCA

$$\text{minimize} \quad \|XY - A\|_F^2$$

- variables $X \in \mathbf{R}^{m\times k}$, $Y \in \mathbf{R}^{k\times n}$
- *not* a convex problem
- but has analytical solution

$$X^\star = U_k\Sigma_k^{1/2}, \qquad Y^\star = \Sigma_k^{1/2}V_k^T$$

from SVD $A = U\Sigma V^T$

- $U_k \in \mathbf{R}^{m\times k}$, $V_k \in \mathbf{R}^{n\times k}$ are the first k left and right singular vectors
- $\Sigma_k = \mathbf{diag}(\sigma_1, \ldots, \sigma_k) \in \mathbf{R}^{k\times k}$ is the diagonal matrix of the first k singular values

regularized PCA

- quadratically regularized PCA

$$\text{minimize} \quad \|XY - A\|_F^2 + \lambda(\|X\|_F^2 + \|Y\|_F^2)$$

  - not a convex problem, also has analytical solution via SVD
  - finds a low-rank approximation of A with small entries in X and Y

- sparse PCA

$$\text{minimize} \quad \|XY - A\|_F^2 + \lambda(\|X\|_{\text{sav}} + \|Y\|_{\text{sav}})$$

  - sum-absolute-value norm $\|U\|_{\text{sav}} = \sum_{i=1}^m \sum_{j=1}^n |U_{ij}|$
  - not a convex problem; good approximation method exists
  - finds a low-rank approximation of A with sparse entries in X and Y

constrained PCA: nonnegative matrix factorization

$$\begin{array}{ll} \text{minimize} & \|XY - A\|_F^2 \\ \text{subject to} & X_{ij} \geq 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, k \\ & Y_{ij} \geq 0, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n \end{array}$$

- finds nonnegative low-rank approximation of A
- nonconvex problem; good approximation method exists

Course goals and topics

goals

- recognize/formulate disciplined inverse problems
- develop code scripts to solve inverse problems
- characterize solution quality

topics

- mathematical optimization
  - convex sets, functions, and optimization problems
  - sequential methods for nonlinear optimization
- disciplined modules
  - objective functions and regularizers
  - constraints, relaxation, and duality
- applications, $e.g.$, robust and latent factor models

Mathematical optimization

(mathematical) optimization problem

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

- $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$: *optimization variables*
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: *objective function*
- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, m$: *inequality constraint functions*
- $h_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, p$: *equality constraint functions*

optimal point $x^\star$ has smallest value of f_0 among all vectors that satisfy the constraints

Solving optimization problems

general optimization problem

- very difficult to solve
- usually involve some compromise
  - very long computation time
  - only finds an approximate solution

exceptions: certain problem classes can be solved efficiently and reliably

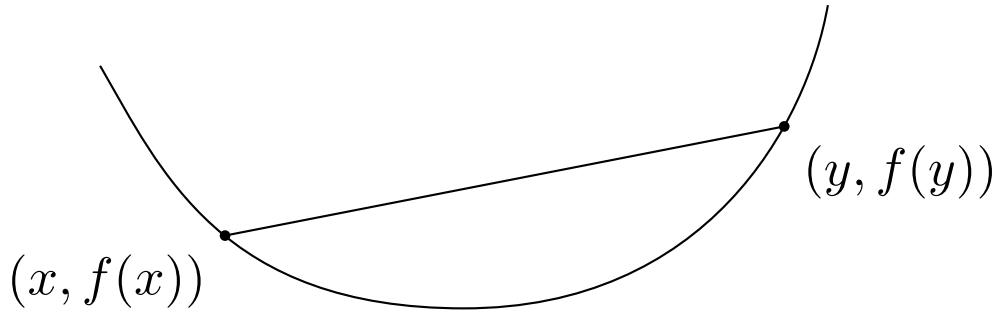
Convex optimization

convex optimization problem satisfies

- $f_i : \mathbf{R}^n \to \mathbf{R}$, $i = 0, \ldots, m$, are convex, *i.e.*,

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

  for all $x, y \in \mathbf{R}^n$ and $\theta \in [0, 1]$



- $h_i : \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, p$, are affine

examples: least squares, least norm problems, . . .

solving convex optimization problems

- (usually) no analytical solution
- reliable and efficient numerical algorithms
- a technology; we could now say

  If a problem can be shown/formulated convex, then it is solved.

using convex optimization

- often difficult to recognize
- many tricks for transforming problems into convex form
- surprisingly many problems can be solved via convex optimization (however less known)

Brief history of convex optimization

theory (convex analysis): approximately 1900–1970

algorithms (convex optimization)

- 1947: simplex algorithm for linear programming (Dantzig)
- 1960s: early interior-point methods (Fiacco & McCormick, Dikin, . . .)
- 1970s: ellipsoid method and other subgradient methods
- 1980s: polynomial-time interior-point methods for linear programming (Karmarkar 1984)
- late 1980s–now: polynomial-time interior-point methods for nonlinear convex optimization (Nesterov & Nemirovski 1994)

applications

- before 1990: mostly in operations research; few in engineering

- since 1990:
  - many new applications in engineering (control, signal processing, communications, circuit design, . . .)
  - new problem classes
    * semidefinite and second-order cone programming; robust optimization
    * exponential and power cone programming (many applications in machine learning)

disciplined convex programming: 2000–now (Boyd & Vandenberghe, Grant, Diamond, . . .)

- human-readable symbolic modeling languages for convex optimization
- since 2010: extensions to many nonconvex problem classes (geometric, quasiconvex, convex-concave, . . .)
- close to the idea

  Say what you want, not how to get it.

Nonlinear optimization

general nonconvex optimization involve compromises

local optimization (traditional nonlinear optimization)

- find a point that minimizes f_0 among feasible points *near* it
- fast, can handle large problems
- require initial guess that influences the solution quality
- no information about distance to (global) optimum

global optimization

- find the (global) solution
- worst-case complexity grows exponentially with problem size

these methods are often based on solving convex subproblems

2. Convex optimization

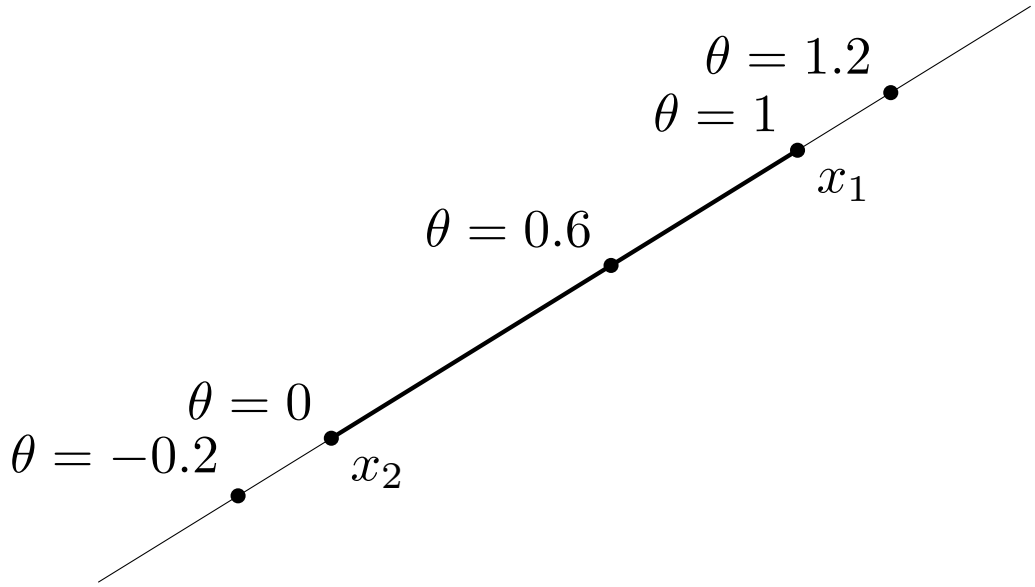
- lines, line segments, and affine sets
- convex sets and cones
- examples and algebra of convex sets
- convex functions and basic properties
- examples of convex functions
- functional operations that preserve convexity
- optimization problems
- equivalent problems and relaxations
- convex optimization problems and optimality conditions
- typical convex optimization problems

Lines and line segments

line through $x_1, x_2 \in \mathbf{R}^n$: all the points

$$y = \theta x_1 + (1 - \theta) x_2 \qquad (\theta \in \mathbf{R})$$

if $\theta \in [0, 1]$, then y is a **line segment** between x_1 and x_2



Affine sets

affine set contains the line through any two points in the set, $i.e.$,

$$x_1, x_2 \in C, \quad \theta \in \mathbf{R} \quad \Longrightarrow \quad \theta x_1 + (1-\theta)x_2 \in C$$

affine combination of points $x_1, \ldots, x_k \in \mathbf{R}^n$:

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

where $\theta \in \mathbf{R}^k$ and

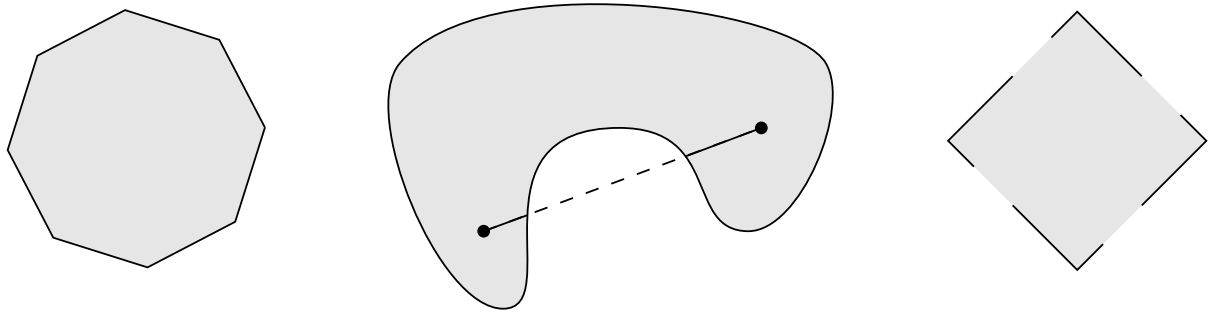
$$\mathbf{1}^T\theta = \theta_1 + \cdots + \theta_k = 1$$

- affine set contains all affine combinations of its points

Convex sets

convex set contains line segment between any two points in the set, $i.e.$,

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



convex combination of points $x_1, \ldots, x_k \in \mathbf{R}^n$:

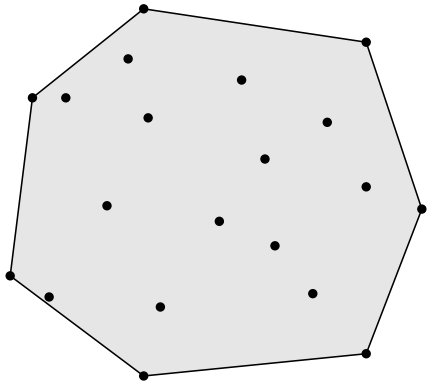
$$\theta_1 x_1 + \cdots + \theta_k x_k$$

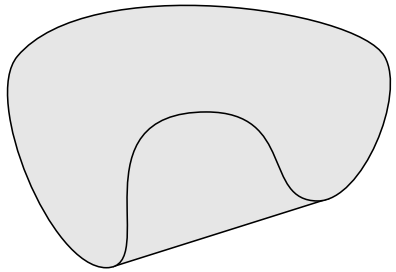
where $\theta \in \mathbf{R}^k$ and

$$\mathbf{1}^T\theta = 1, \quad \theta \succeq 0 \ (i.e., \ \theta_i \geq 0, \ i = 1, \ldots, k)$$

convex hull of set C: set of all convex combinations of points in C, $i.e.$,

$$\mathbf{conv}\, C = \left\{ \sum_{i=1}^{k} \theta_i x_i \;\middle|\; \begin{array}{ll} k \in \mathbf{Z}_{++} & \\ x_i \in C, & i = 1, \ldots, k \\ \mathbf{1}^T \theta = 1, & \theta \succeq 0 \end{array} \right\}$$





Convex combinations of infinite points

let $C \subseteq \mathbf{R}^n$ be a convex set:

- $x_1, x_2, \ldots \in C$, $\theta_1, \theta_2, \ldots \geq 0$, $\sum_{i=1}^{\infty} \theta_i = 1 \implies$

$$\sum_{i=1}^{\infty} \theta_i x_i \in C$$

- $p\colon \mathbf{R}^n \to \mathbf{R}$, $\int_C p(x)\, dx = 1$, $p(x) \geq 0$ for all $x \in C \implies$

$$\int_C p(x) x \, dx \in C$$

- random vector $x \in \mathbf{R}^n$ with $\mathbf{prob}(x \in C) = 1 \implies$

$$\mathbf{E}\, x \in C$$

Convex cones

cone contains all rays through the origin and any point in the set, $i.e.$,

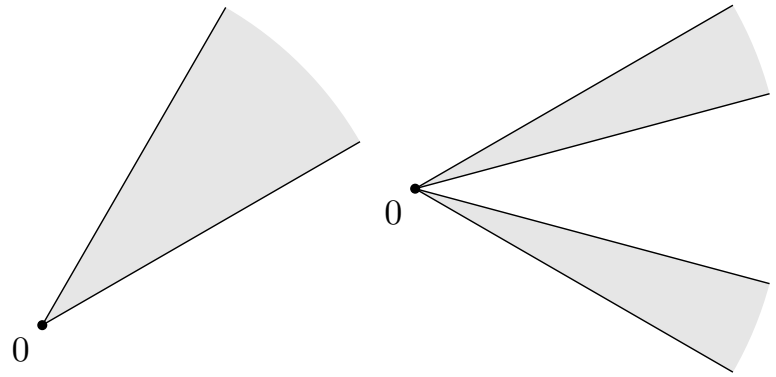
$$x \in C, \quad \theta \geq 0 \quad \Longrightarrow \quad \theta x \in C$$

conic (nonnegative) combination of points $x_1, \ldots, x_k \in \mathbf{R}^n$:

$$\theta_1 x_1 + \cdots + \theta_k x_k$$

where $\theta \in \mathbf{R}^k$ and $\theta \succeq 0$

convex cone contains all conic combinations of its points



Basic examples

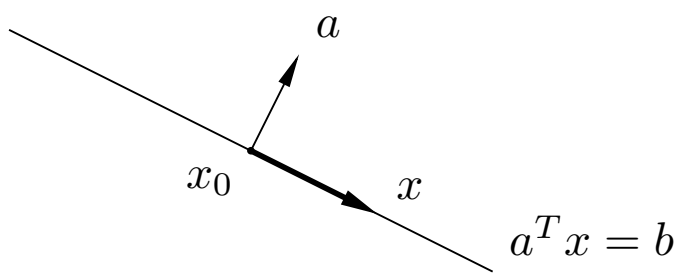
- subspace of $\mathbf{R}^n$ is an affine set and a convex cone, and hence convex
- empty set $\emptyset$, singleton $\{x_0\}$, and $\mathbf{R}^n$ are affine sets
- line through two different points is affine
- line segment between two different points is convex but not affine
- ray $\{x_0 + \theta v \mid \theta \geq 0\}$ $(v \neq 0)$ is convex
  (is a convex cone if contains the origin)
- translation $x_0 + C$ of a convex cone C is convex but not a cone
- solution set of linear equations

$$C = \{x \in \mathbf{R}^n \mid Ax = b\}$$

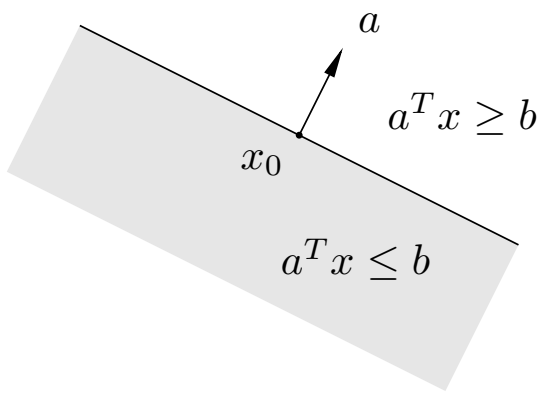
  for $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$ is an affine set

Hyperplanes and halfspaces

hyperplane: set of the form $\{x \in \mathbf{R}^n \mid a^T x = b\}$ $(a \neq 0)$



halfspace: set of the form $\{x \in \mathbf{R}^n \mid a^T x \leq b\}$ $(a \neq 0)$



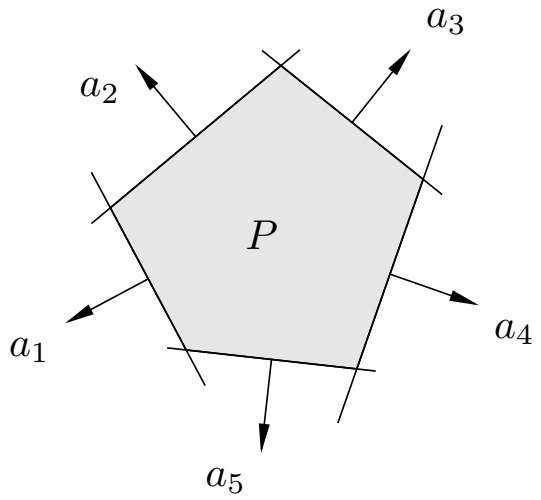
- a is the *normal vector*
- hyperplanes are affine; halfspaces are convex

Polyhedra and simplexes

polyhedron: intersection of finite hyperplanes and halfspaces, $i.e.$,

$$\begin{aligned} P &= \left\{ x \in \mathbf{R}^n \;\middle|\; \begin{array}{ll} a_i^T x \leq b_i, & i = 1, \ldots, m \\ c_i^T x = d_i, & i = 1, \ldots, p \end{array} \right\} \\ &= \{ x \in \mathbf{R}^n \mid Ax \preceq b, \quad Cx = d \} \end{aligned}$$

- $A = \left[\begin{array}{ccc} a_1 & \cdots & a_m \end{array} \right]^T \in \mathbf{R}^{m \times n}$
- $C = \left[\begin{array}{ccc} c_1 & \cdots & c_p \end{array} \right]^T \in \mathbf{R}^{p \times n}$



example: nonnegative orthant $\mathbf{R}^n_+ = \{x \in \mathbf{R}^n \mid x \succeq 0\}$ is a polyhedron (where positive orthant $\mathbf{R}^n_{++} = \{x \in \mathbf{R}^n \mid x \succ 0\}$ is its interior)

simplex: convex hull of $k+1$ affinely independent points, *i.e.*,

$$C = \mathbf{conv}\{x_0, x_1, \ldots, x_k\}$$

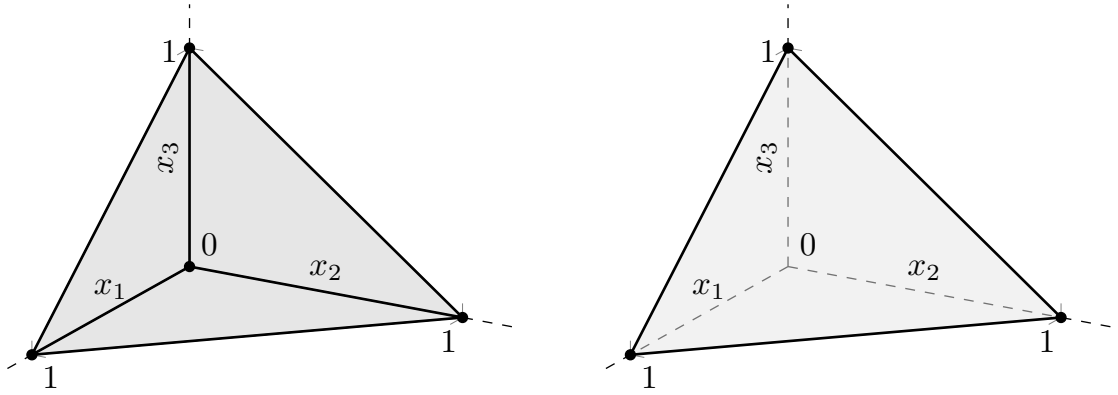
examples:

- *unit simplex* in $\mathbf{R}^n$:

$$\mathbf{conv}\{0, e_1, e_2, \ldots, e_n\} = \{x \in \mathbf{R}^n \mid x \succeq 0,\ \mathbf{1}^T x \leq 1\}$$

- *probability simplex* in $\mathbf{R}^n$:

$$\mathbf{conv}\{e_1, e_2, \ldots, e_n\} = \{x \in \mathbf{R}^n \mid x \succeq 0,\ \mathbf{1}^T x = 1\}$$



Norm balls

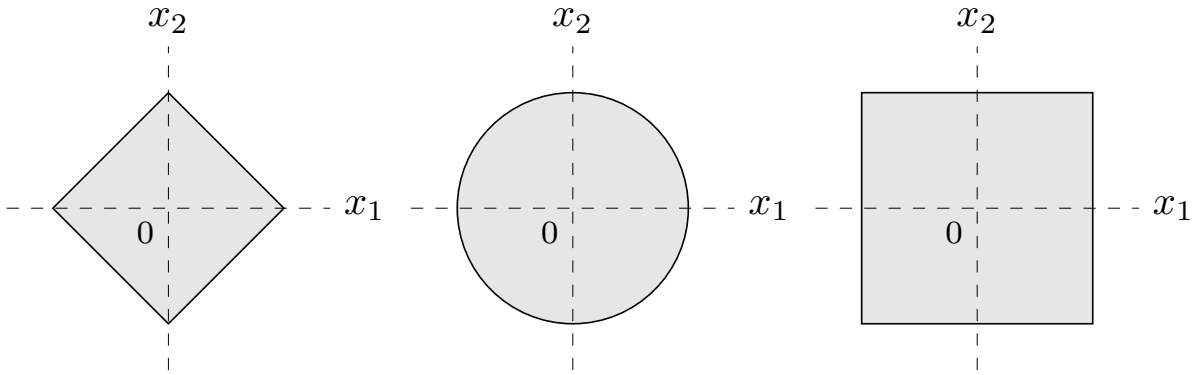
norm: a function $\|\cdot\| : \mathbf{R}^n \to \mathbf{R}$ satisfying

- $\|x\| \geq 0$ for all $x \in \mathbf{R}^n$, and $\|x\| = 0$ if and only if $x = 0$
- $\|tx\| = |t|\|x\|$ for all $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$
- $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbf{R}^n$

we use $\|\cdot\|$ for general norm, and $\|\cdot\|_{\mathrm{symb}}$ for specific norm

norm ball of radius r and center x_c: $\{x \in \mathbf{R}^n \mid \|x - x_c\| \leq r\}$

examples: unit balls of ℓ_1-, ℓ_2- (Euclidean), and ℓ_∞-norms in $\mathbf{R}^2$:



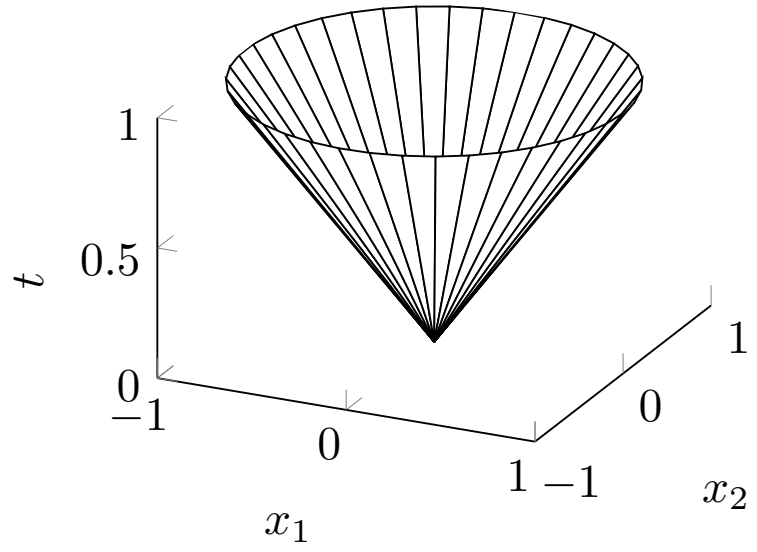
Norm cones

norm cone associated with norm $\|\cdot\|$:

$$\{(x,t) \in \mathbf{R}^n \times \mathbf{R} \mid \|x\| \leq t\}$$

- norm cones are convex cones
- example: **second-order cone** in $\mathbf{R}^3$:

$$\{(x,t) \in \mathbf{R}^2 \times \mathbf{R} \mid \|x\|_2 \leq t\}$$



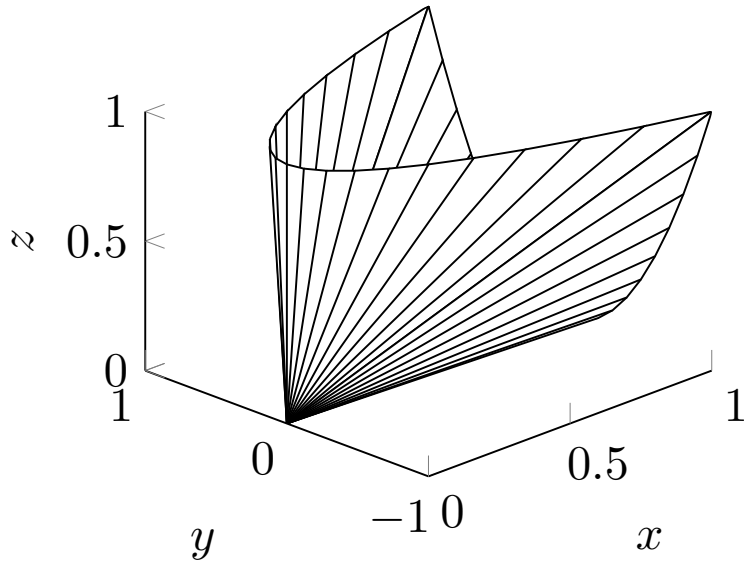
Positive semidefinite cones

- $\mathbf{S}^n$: set of symmetric $n \times n$ matrices
- $\mathbf{S}^n_+ = \{X \in \mathbf{S}^n \mid X \succeq 0\}$: positive semidefinite $n \times n$ matrices, *i.e.*,

$$X \in \mathbf{S}^n_+ \quad \Longleftrightarrow \quad z^T X z \geq 0 \text{ for all } z \in \mathbf{R}^n$$

  - $\mathbf{S}^n_+$ is a convex cone
- $\mathbf{S}^n_{++} = \{X \in \mathbf{S}^n \mid X \succ 0\}$: positive definite $n \times n$ matrices

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



Algebra of convex sets

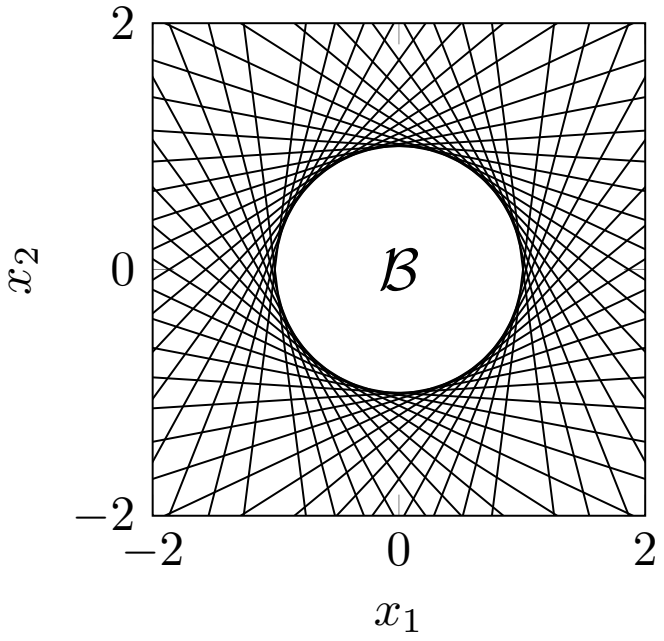
intersection of convex sets is convex, $i.e.$,

- $C_1, C_2, \ldots, C_k$ convex $\implies \bigcap_{i=1}^k C_i$ convex
- $\{C_i\}_{\mathcal{I}}$ set of convex sets indexed by $\mathcal{I}$ $\implies \bigcap_{i\in\mathcal{I}} C_i$ convex

example: unit Euclidean ball in $\mathbf{R}^2$

$$\mathcal{B} = \bigcap_{t\in[0,2\pi)} \{x \mid x_1 \cos t + x_2 \sin t \leq 1\}$$

is intersection of infinite halfspaces



Cartesian product of convex sets is convex, *i.e.*,

$$C \times D = \{(x, y) \mid x \in C,\ y \in D\}$$

is convex if C and D are convex

affine transformation: suppose $f : \mathbf{R}^n \to \mathbf{R}^m$ is affine, *i.e.*,

$$f(x) = Ax + b \text{ with } A \in \mathbf{R}^{m \times n},\ b \in \mathbf{R}^m$$

- the image of a convex set under f is convex, *i.e.*,

$$C \subseteq \mathbf{R}^n \text{ convex} \implies f(C) = \{f(x) \mid x \in C\} \text{ convex}$$

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

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

examples:

- scaling and translation of a convex set is convex, $i.e.$,

$$C \subseteq \mathbf{R}^n \text{ convex}, \quad \lambda \in \mathbf{R}, \quad v \in \mathbf{R}^n$$
$$\Longrightarrow \quad \lambda C + v = \{\lambda x + v \mid x \in C\} \text{ convex}$$

- (Minkowski) addition of convex sets is convex, $i.e.$,

$$C, D \subseteq \mathbf{R}^n \text{ convex}$$
$$\Longrightarrow \quad C + D = \{x + y \mid x \in C,\ y \in D\} \text{ convex}$$

- solution set of linear matrix inequality

$$\{x \in \mathbf{R}^n \mid x_1 A_1 + \cdots + x_n A_n \preceq B\}$$

  with $A_1, \ldots, A_n, B \in \mathbf{S}^m$ is convex

- an ellipsoid

$$\begin{aligned}\mathcal{E} &= \{x \in \mathbf{R}^n \mid (x - x_0)^T P^{-1}(x - x_0) \leq 1\} \\ &= \{x \in \mathbf{R}^n \mid \|P^{-1/2}(x - x_0)\|_2 \leq 1\}\end{aligned}$$

  with center $x_0 \in \mathbf{R}^n$ and $P \in \mathbf{S}^n_{++}$ is the inverse image of the unit ball

$$\{x \in \mathbf{R}^n \mid \|x\|_2 \leq 1\}$$

  under the affine transformation

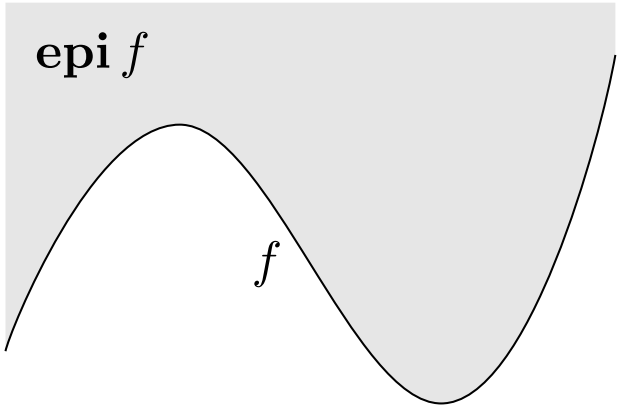
$$f(x) = P^{-1/2}(x - x_0)$$

  and hence is convex

Convex functions

epigraph of function $f : \mathbf{R}^n \to \mathbf{R}$:

$$\mathbf{epi}\, f = \left\{ (x,t) \in \mathbf{R}^{n+1} \;\middle|\; \begin{array}{l} f(x) \leq t \\ x \in \mathbf{dom}\, f \end{array} \right\}$$



- $f : \mathbf{R}^n \to \mathbf{R}$ is **convex** if $\mathbf{epi}\, f$ is convex
- $f : \mathbf{R}^n \to \mathbf{R}$ is **concave** if $-f$ is convex, *i.e.*, its *hypograph*

$$\mathbf{hypo}\, f = \{(x,t) \in \mathbf{R}^{n+1} \mid f(x) \geq t,\ x \in \mathbf{dom}\, f\}$$

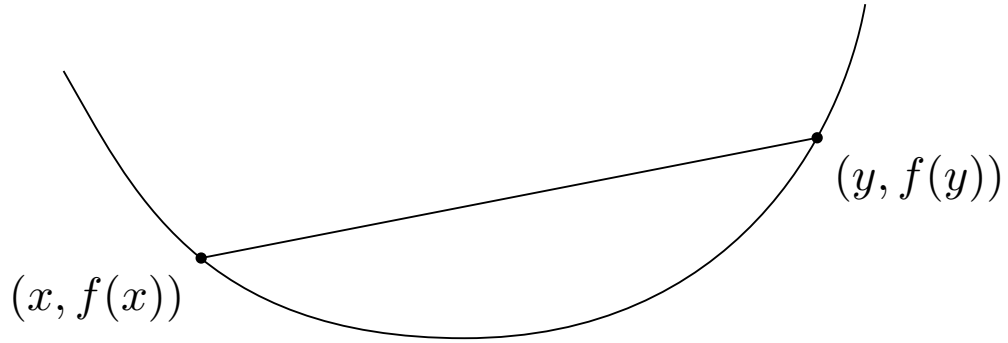
is convex

Jensen's inequality

$f : \mathbf{R}^n \to \mathbf{R}$ is convex if $\mathbf{dom}\, f$ is a convex set and

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

for all $x, y \in \mathbf{dom}\, f$ and $\theta \in [0,1]$



- f is strictly convex if $\mathbf{dom}\, f$ is convex and

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

  for all $x, y \in \mathbf{dom}\, f$, $x \neq y$, and $\theta \in (0,1)$

extensions: let $f : \mathbf{R}^n \to \mathbf{R}$ be convex

- $x_1, \ldots, x_k \in \mathbf{dom}\, f$, $\theta \in \mathbf{R}^k$ with $\mathbf{1}^T\theta = 1$, $\theta \succeq 0 \implies$

$$f(\theta_1 x_1 + \cdots + \theta_k x_k) \leq \theta_1 f(x_1) + \cdots + \theta_k f(x_k)$$

- $p : S \to \mathbf{R}$, $S \subseteq \mathbf{dom}\, f$, $\int_S p(x)\, dx = 1$, $p(x) \geq 0$ for all $x \in S \implies$

$$f\left(\int_S p(x) x\, dx\right) \leq \int_S p(x) f(x)\, dx$$

- random vector $z \in \mathbf{R}^n$ with $\mathbf{prob}(z \in \mathbf{dom}\, f) = 1 \implies$

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

basic inequality is a special case with discrete distribution

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

Sublevel sets

α-**sublevel set** of function $f : \mathbf{R}^n \to \mathbf{R}$:

$$C_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

- f is convex $\implies$ C_α is convex for all $\alpha \in \mathbf{R}$

- converse is false: C_α convex for all $\alpha \in \mathbf{R}$ does not imply f convex
  - $f(x) = x^3$ has convex sublevel sets $(-\infty, \alpha^{1/3})$ but is not convex
  - $f(x) = \log x$ has convex sublevel sets $(0, e^\alpha]$ but is not convex

Convex envelope

suppose $F \subseteq \mathbf{R}^{n+1}$ is convex, the function $f\colon \mathbf{R}^n \to \mathbf{R}$ defined by

$$f(x) = \inf\{t \mid (x,t) \in F\}$$

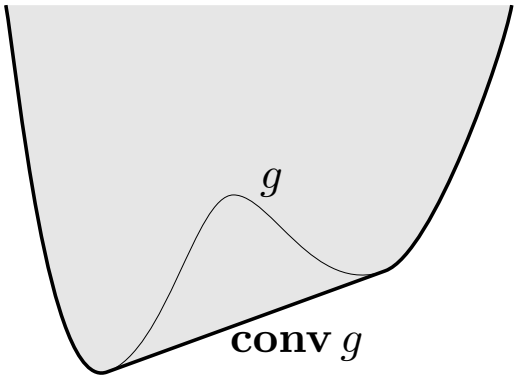
is a convex function with $\mathbf{epi}\, f = F$

convex envelope of $g\colon \mathbf{R}^n \to \mathbf{R}$: the function $f\colon \mathbf{R}^n \to \mathbf{R}$ obtained from the operation above with

$$F = \mathbf{conv}\, \mathbf{epi}\, g$$

- denoted $\mathbf{conv}\, g$
- largest convex global underestimator of g, *i.e.*,

$$\mathbf{conv}\, g(x) = \sup \left\{ h(x) \;\middle|\; \begin{array}{l} h \text{ convex} \\ h(z) \leq g(z) \\ \text{for all } z \in \mathbf{dom}\, g \end{array} \right\}$$

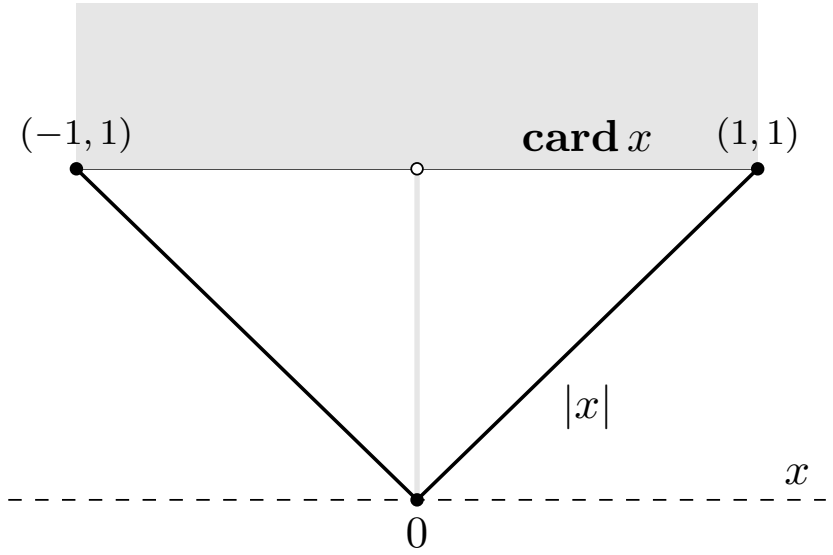


example: convex envelope of the *cardinality function* for $x \in \mathbf{R}$:

$$\mathbf{card}\, x = \begin{cases} 0, & x = 0 \\ 1, & x \neq 0 \end{cases}$$

on the interval $[-1, 1]$ is the absolute value function, *i.e.*,

$$\mathbf{conv}\, \mathbf{card}\, x = |x|$$

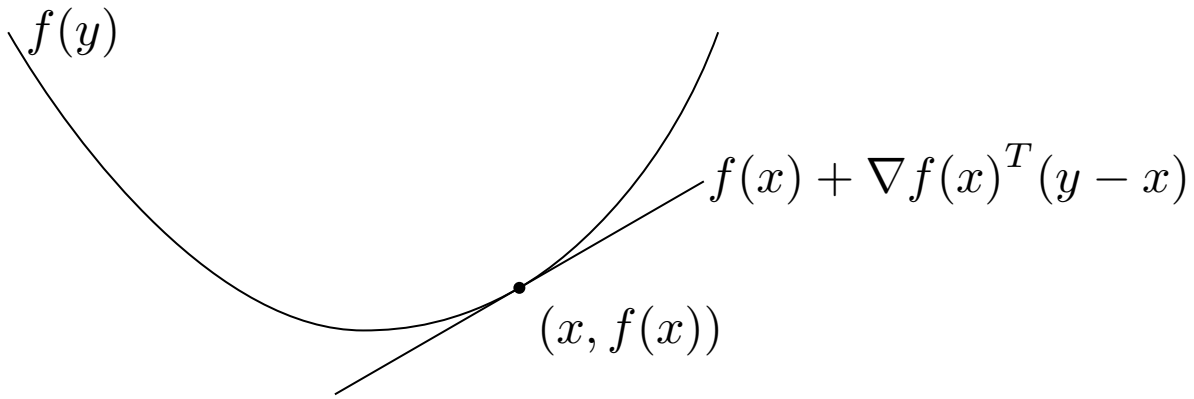


First-order conditions

differentiable $f : \mathbf{R}^n \to \mathbf{R}$ with convex domain is convex iff

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

where $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)$ is the gradient of f at $x \in \mathbf{R}^n$



- 1st-order approximation of convex f is global underestimator
- strictly convex f if $>$ holds for $x \neq y$ (converse also true)

Second-order conditions

twice differentiable $f : \mathbf{R}^n \to \mathbf{R}$ with convex domain is convex iff

$$\nabla^2 f(x) \succeq 0 \quad \text{for all } x \in \mathbf{dom}\, f$$

where $\nabla^2 f(x) \in \mathbf{S}^n$ is the Hessian of f at $x \in \mathbf{R}^n$, *i.e.*,

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

- strictly convex f if $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$
- converse not true: $f(x) = x^4$ is strictly convex but $\nabla^2 f(0) = 0$

convex domain cannot be dropped from 1st- and 2nd-order conditions
e.g., $f(x) = 1/x^2$ nonconvex but $\nabla^2 f(x) \succ 0$ for all $x \neq 0$

Examples on $\mathbf{R}$

convex:

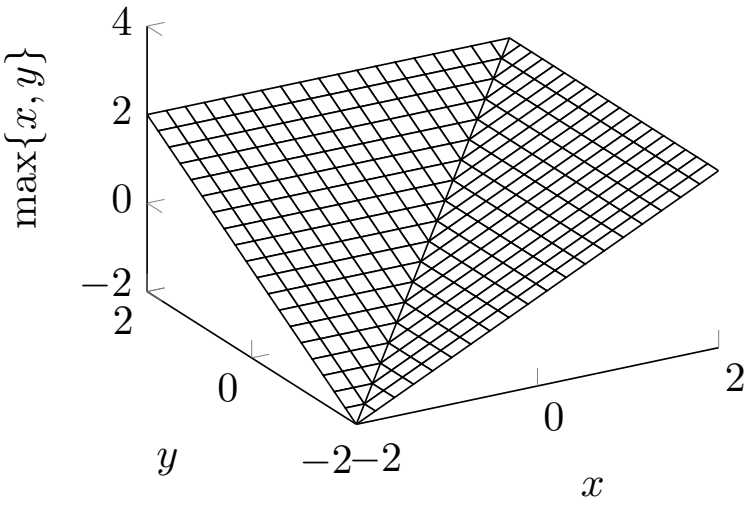
- affine: $f(x) = ax + b$ on $\mathbf{R}$ for any $a, b \in \mathbf{R}$
- quadratic: $f(x) = ax^2 + bx + c$ on $\mathbf{R}$ for any $a, b, c \in \mathbf{R}$, $a > 0$
- absolute value: $f(x) = |x|$ on $\mathbf{R}$
- powers: $f(x) = x^p$ on $\mathbf{R}_{++}$ for $p \leq 0$ or $p \geq 1$
- exponential: $f(x) = e^x$ on $\mathbf{R}$
- negative entropy: $f(x) = x \log x$ on $\mathbf{R}_{++}$

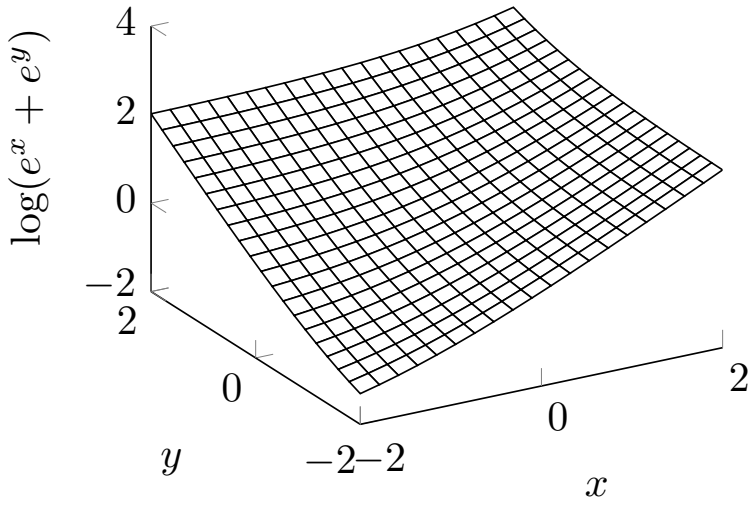
concave:

- affine: $f(x) = ax + b$ on $\mathbf{R}$ for any $a, b \in \mathbf{R}$
- powers: $f(x) = x^p$ on $\mathbf{R}_{++}$ for $0 < p \leq 1$
- logarithm: $f(x) = \log x$ on $\mathbf{R}_{++}$

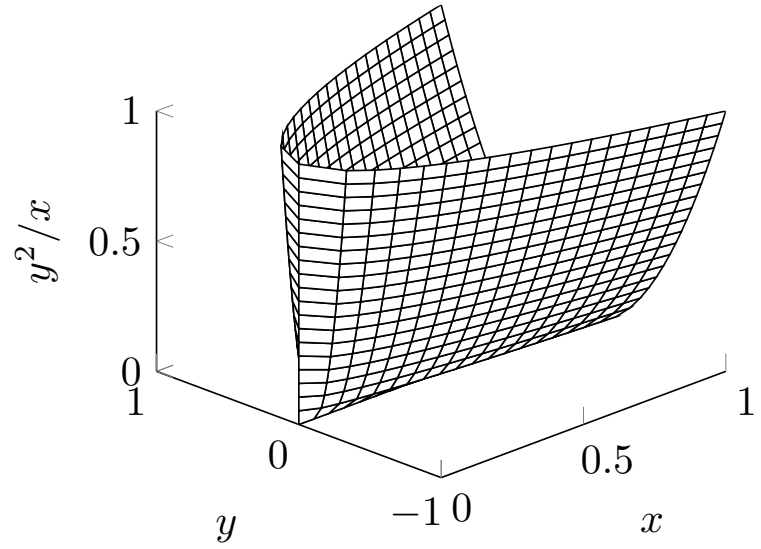
Examples on $\mathbf{R}^n$

- affine: $f(x) = a^T x + b$ convex and concave on $\mathbf{R}^n$
- quadratic: $f(x) = x^T P x + q^T x + r$ convex on $\mathbf{R}^n$ for $P \succeq 0$
- norms: $f(x) = \|x\|$ convex on $\mathbf{R}^n$ for any norm $\|\cdot\|$
  e.g., $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ for $p \geq 1$; $\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|$
- max: $f(x) = \max\{x_1, \ldots, x_n\}$ convex on $\mathbf{R}^n$
- log-sum-exp: $f(x) = \log \sum_{i=1}^n e^{x_i}$ convex on $\mathbf{R}^n$





- quadratic-over-linear: $f(x,y) = y^2/x$ convex on $\mathbf{R}_{++} \times \mathbf{R}$



- relative entropy: $f(x,y) = \sum_{i=1}^n x_i \log(x_i/y_i)$ convex on $\mathbf{R}^n_{++} \times \mathbf{R}^n_{++}$
- Kullback-Leibler (KL) divergence:

$$f(x,y) = \sum_{i=1}^n (x_i \log(x_i/y_i) - x_i + y_i)$$

convex on $\mathbf{R}^n_{++} \times \mathbf{R}^n_{++}$

Examples on $\mathbf{R}^{m\times n}$

- affine:

$$f(X) = \mathbf{tr}(A^T X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$$

  convex and concave on $\mathbf{R}^{m\times n}$ for any $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}$

- spectral norm:

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = \left(\lambda_{\max}(X^T X)\right)^{1/2}$$

  convex on $\mathbf{R}^{m\times n}$ ($\sigma_{\max}$, $\lambda_{\max}$: largest singular value, eigenvalue)

- log-determinant:

$$f(X) = \log\det X$$

  concave on $\mathbf{S}^n_{++}$ (analogous to $f(x) = \log x$ on $\mathbf{R}_{++}$)

Probability distributions

logarithm of many probability distributions are concave ($i.e.$, log-concave):

- multivariate normal:

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

  where $\mu \in \mathbf{R}^n$ is the mean and $\Sigma \in \mathbf{S}^n_{++}$ is the covariance matrix

- uniform:

$$f(x) = \begin{cases} 1/\alpha, & x \in C \\ 0, & \text{otherwise} \end{cases}$$

  where $C \subseteq \mathbf{R}^n$ is a convex set and $\alpha > 0$ is its volume; log given by

$$\log f(x) = \begin{cases} -\log \alpha, & x \in C \\ -\infty, & \text{otherwise} \end{cases}$$

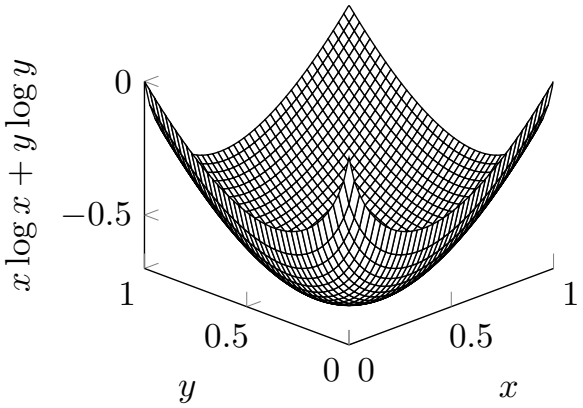
Basic functional operations

nonnegative weighted sum of convex functions is convex, $i.e.$,

$$\begin{array}{l} f_1, \ldots, f_k \text{ convex} \\ w_1, \ldots, w_k \geq 0 \end{array} \quad \Longrightarrow \quad w_1 f_1 + \cdots + w_k f_k \text{ convex}$$

- example: negative entropy

$$f(x) = \sum_{i=1}^{n} x_i \log x_i$$



  convex on $\mathbf{R}^n_{++}$

- extension: $f(x, y)$ convex in x, $w(y) \geq 0$ for all fixed $y \in S$ $\Longrightarrow$

$$g(x) = \int_S w(y) f(x, y)\, dy$$

  convex in x (basic property as discrete case)

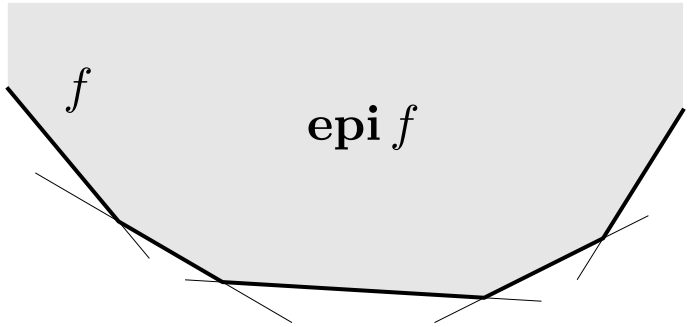
pointwise maximum of convex functions is convex, $i.e.$,

$$f_1, \ldots, f_k \text{ convex} \quad \Longrightarrow \quad f(x) = \max\{f_1(x), \ldots, f_k(x)\} \text{ convex}$$

- $\mathbf{dom}\, f = \bigcap_{i=1}^k \mathbf{dom}\, f_i$
- proof via epigraph: $\mathbf{epi}\, f = \bigcap_{i=1}^k \mathbf{epi}\, f_i$ is intersection of convex sets
- example: piecewise linear function

  $$f(x) = \max_{i=1,\ldots,k} (a_i^T x + b_i) \quad (a_i \neq 0)$$

  convex on $\mathbf{R}^n$



- pointwise minimum of concave functions is concave

pointwise supremum: $\{f_i\}_{i\in\mathcal{I}}$ convex $\Longrightarrow$

$$f(x) = \sup_{i\in\mathcal{I}} f_i(x)$$

with

$$\mathbf{dom}\, f = \left\{ x \in \bigcap_{i\in\mathcal{I}} \mathbf{dom}\, f_i \;\middle|\; \sup_{i\in\mathcal{I}} f_i(x) < \infty \right\}$$

is convex

- example: distance to the farthest point in a set $C \subseteq \mathbf{R}^n$:

$$f(x) = \sup_{y\in C} \|x - y\|$$

  is convex on $\mathbf{R}^n$

- corollary: any convex function $f\colon \mathbf{R}^n \to \mathbf{R}$ can be expressed as

$$f(x) = \sup\{a^T x + b \mid a^T y + b \le f(y) \text{ for all } y \in \mathbf{R}^n\}$$

  $i.e.$, the supremum of all its affine underestimators

Composition with affine

given $f: \mathbf{R}^m \to \mathbf{R}$, then $g: \mathbf{R}^n \to \mathbf{R}$ with

$$g(x) = f(Ax + b), \quad \mathbf{dom}\, g = \{x \mid Ax + b \in \mathbf{dom}\, f\}$$

has the same convexity as f for any $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$

examples:

- log-barrier function for linear inequalities $a_i^T x \leq b_i$, $i = 1, \ldots, m$:

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

  convex on $\mathbf{dom}\, f = \{x \mid a_i^T x < b_i,\ i = 1, \ldots, m\}$
- linear approximation cost $f(x) = \|Ax - b\|$ convex on $\mathbf{R}^n$ for any norm
- least squares cost $f(x) = \|Ax - b\|_2^2$ convex on $\mathbf{R}^n$

General composition

given $h\colon \mathbf{R}^k \to \mathbf{R}$ and $g_i\colon \mathbf{R}^n \to \mathbf{R}$, $i = 1, \ldots, k$, define $f\colon \mathbf{R}^n \to \mathbf{R}$ as

$$f(x) = h(g_1(x), \ldots, g_k(x))$$

where

$$\mathbf{dom}\, f = \left\{ x \in \mathbf{R}^n \;\middle|\; \begin{array}{l} x \in \mathbf{dom}\, g_i,\ i = 1, \ldots, k \\ (g_1(x), \ldots, g_k(x)) \in \mathbf{dom}\, h \end{array} \right\}$$

- f convex if $\mathbf{dom}\, f$ convex, h convex, and for $i = 1, \ldots, k$:
  - g_i affine, or
  - h nondecreasing in its ith argument, and g_i convex, or
  - h nonincreasing in its ith argument, and g_i concave

- similar conditions holds for concavity of f

proof for $k = n = 1$: let $f = h \circ g$ with $\mathbf{dom}\, f$ convex,

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

- assume h convex and nondecreasing, g convex
- $\Longrightarrow$ $h''(g(x)) \geq 0$, $h'(g(x)) \geq 0$, $g''(x) \geq 0$ for all $x \in \mathbf{dom}\, f$
- $\Longrightarrow$ $f''(x) \geq 0$ for all $x \in \mathbf{dom}\, f$, *i.e.*, f convex

examples:

- $f(x) = |x|^p$ convex on $\mathbf{R}$ for $p \geq 1$
- $f(x) = 1/g(x)$ convex if g concave and positive
- $f(x) = \exp g(x)$ convex if g convex
- $f(x) = \sum_{i=1}^k \log g_i(x)$ concave if g_i concave and positive
- $f(x) = \left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$ $(p \geq 1)$ convex if g_i convex and nonnegative
- $f(x) = \log \sum_{i=1}^k \exp g_i(x)$ convex if g_i convex

Optimization problem in standard form

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

- $x \in \mathbf{R}^n$: optimization or decision *variable*
- $f_0 : \mathbf{R}^n \to \mathbf{R}$: *objective* function
- $f_1, \ldots, f_m : \mathbf{R}^n \to \mathbf{R}$: *inequality constraint* functions
- $h_1, \ldots, h_p : \mathbf{R}^n \to \mathbf{R}$: *equality constraint* functions

goal: find x that minimizes $f_0(x)$ while satisfying the constraints

Optimal value and points

optimal value of a problem:

$$p^\star = \inf\{f_0(x) \mid x \in \mathcal{D}\}$$

where $\mathcal{D}$ is the **feasible set**:

$$\mathcal{D} = \left\{ x \in \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i \;\middle|\; \begin{array}{l} f_i(x) \leq 0,\ i = 1, \ldots, m \\ h_i(x) = 0,\ i = 1, \ldots, p \end{array} \right\}$$

- $p^\star = \infty$: problem is **infeasible**, *i.e.*, $\mathcal{D} = \emptyset$
- $p^\star = -\infty$: problem is **unbounded below**
- a point x is **feasible** if $x \in \mathcal{D}$, otherwise **infeasible**

a feasible point $x^\star$ is **optimal** if $f_0(x^\star) = p^\star$

x is **locally optimal** if there is $\epsilon > 0$ such that

$$f_0(x) = \inf\{f_0(z) \mid z \in \mathcal{D},\ \|z - x\|_2 \leq \epsilon\}$$

examples ($n = 1$, $m = p = 0$):

- $f_0(x) = x \log x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -1/e$, $x^\star = 1/e$
- $f_0(x) = 1/x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = 0$, no optimal point
- $f_0(x) = -\log x$, $\mathbf{dom}\, f_0 = \mathbf{R}_{++}$: $p^\star = -\infty$
- $f_0(x) = x^3 - 3x$: $p^\star = -\infty$, local optimum at $x = 1$

Feasibility problem

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

can be considered a special case of the standard form with $f_0(x) = 0$:

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

- $p^\star = 0$ if constraints are feasible
- any feasible x is optimal
- $p^\star = \infty$ if constraints are infeasible

Equivalent problems

two problems are (informally) **equivalent** if the solution of one is readily obtained from the solution of the other, and vice versa

examples:

- change of variables; scaling and translation of objective
- introduce slack variables:

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

is equivalent to

$$\begin{array}{ll} \text{minimize (over } x, s) & f_0(x) \\ \text{subject to} & f_i(x) + s_i = 0, \quad i = 1, \ldots, m \\ & s \succeq 0 \end{array}$$

where s_i is the *slack variable* associated with $f_i(x) \leq 0$

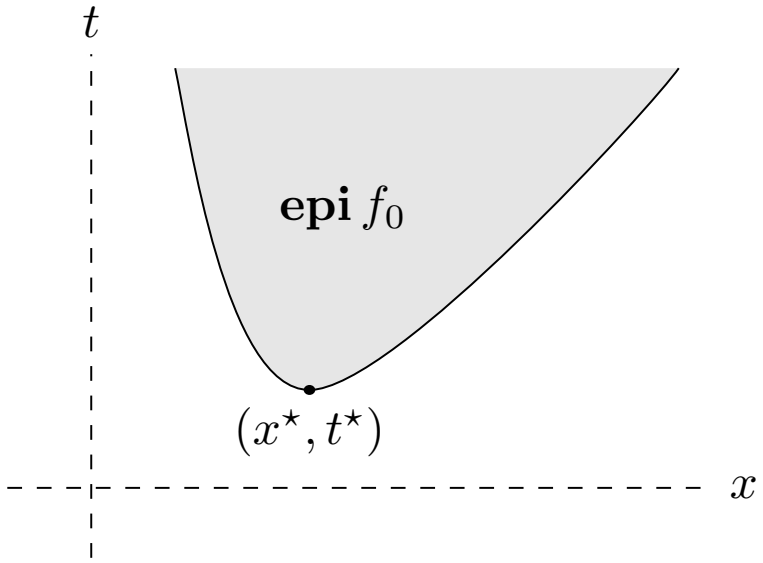
- introduce epigraph variable:

$$\text{minimize (over } x) \quad f_0(x)$$

is equivalent to

$$\begin{array}{ll} \text{minimize (over } x, t) & t \\ \text{subject to} & f_0(x) - t \leq 0 \end{array}$$

where t is the epigraph variable associated with $f_0(x)$



Relaxations

one problem is a **relaxation** of another if the feasible set of the first contains that of the second

basic properties:

- relaxation lower bounds optimal value of the original problem
- relaxation optimal point is original feasible
  $\Longrightarrow$ it is optimal for the original problem

example: the combinatorial problem

$$\begin{array}{ll} \text{minimize} & f_0(x_1, \ldots, x_m) \\ \text{subject to} & x_i \in \{0,1\}^n, \quad \mathbf{card}\, x_i = 1, \quad i = 1, \ldots, m \end{array}$$

is often relaxed to

$$\begin{array}{ll} \text{minimize} & f_0(x_1, \ldots, x_m) \\ \text{subject to} & x_i \succeq 0, \quad \mathbf{1}^T x_i = 1, \quad i = 1, \ldots, m \end{array}$$

Convex optimization problem

standard form of a convex optimization problem:

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

- $f_0, f_1, \ldots, f_m$ convex functions
- equality constraints are affine
- often written as

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

- important property: feasible set of a convex problem is convex

example:

$$\begin{array}{ll} \text{minimize} & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} & f_1(x) = x_1/(1+x_2^2) \leq 0 \\ & h_1(x) = (x_1 + x_2)^2 = 0 \end{array}$$

- f_0 convex; feasible set $\{(x_1, x_2) \mid x_1 = x_2 \leq 0\}$ convex
- (by definition) *not* a convex problem: f_1 not convex, h_1 not affine
- equivalent (but not identical) to convex problem:

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & x_1 \leq 0 \\ & x_1 + x_2 = 0 \end{array}$$

Optimal points

any locally optimal point of a convex problem is global optimal

proof: let $x \in \mathcal{D}$ be locally optimal, *i.e.*, for some $\epsilon > 0$,

$$f_0(x) = \inf\{f_0(z) \mid z \in \mathcal{D},\ \|z - x\|_2 \leq \epsilon\}$$

- suppose $y \in \mathcal{D}$ is feasible with $f_0(y) < f_0(x)$, then $\|y - x\|_2 > \epsilon$
- $\mathcal{D}$ convex $\implies$ $z = \theta x + (1-\theta)y \in \mathcal{D}$ for all $\theta \in (0,1)$
- choose $\theta = \epsilon/(2\|y - x\|_2) \in (0,1)$, then

$$\|z - x\|_2 = \epsilon/2 < \epsilon \quad \implies \quad f_0(z) \geq f_0(x)$$

- from convexity of f_0:

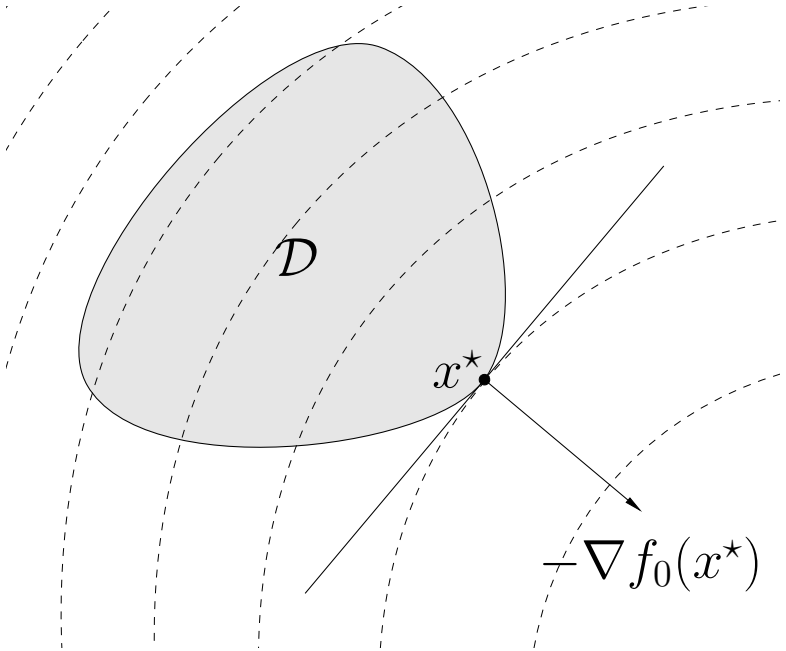
$$f_0(z) = f_0(\theta x + (1-\theta)y) \leq \theta f_0(x) + (1-\theta) f_0(y) < f_0(x)$$

  contradiction $\implies$ no such y exists, *i.e.*, x is globally optimal

Optimality conditions

for differentiable f_0, $x \in \mathcal{D}$ is optimal $\iff$

$$\nabla f_0(x)^T (y - x) \geq 0 \text{ for all } y \in \mathcal{D}$$

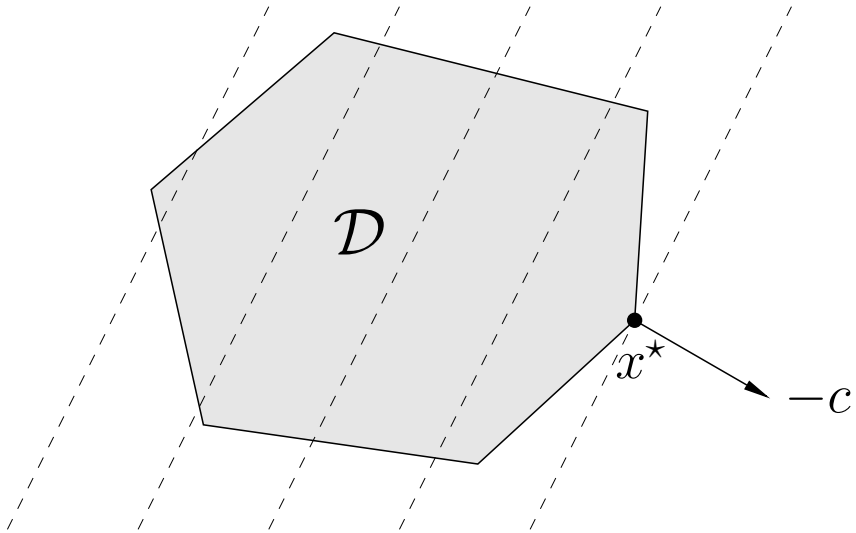


- nonzero $\nabla f_0(x)$ defines a *supporting hyperplane* to $\mathcal{D}$ at x
- unconstrained problem: x optimal $\iff$ $x \in \mathbf{dom}\, f_0$, $\nabla f_0(x) = 0$

Linear program (LP)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

- convex problem with affine objective and constraint functions
- feasible set is a polyhedron



example: least ℓ_1-norm fitting

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

- equivalent to the LP

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T t \\ \text{subject to} & -t \preceq x \preceq t \\ & Ax = b, \end{array}$$

  with variables x and t

- another equivalent LP formulation:

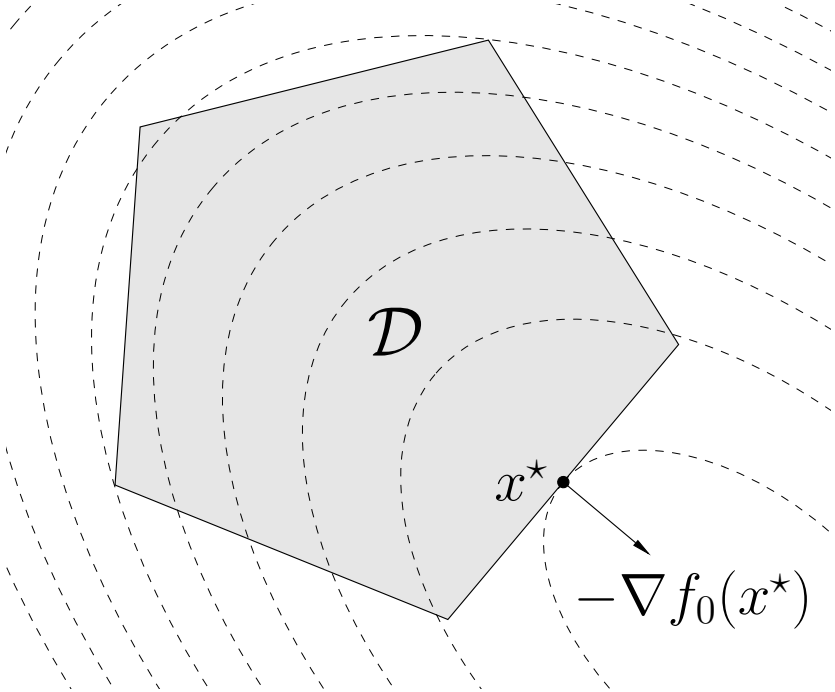
$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T x_+ + \mathbf{1}^T x_- \\ \text{subject to} & x_+ \succeq 0, \quad x_- \succeq 0 \\ & A(x_+ - x_-) = b \end{array}$$

  with variables $x_+, x_- \in \mathbf{R}^n$

Quadratic program (QP)

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Gx \preceq h \\ & Ax = b \end{array}$$

- $P \in \mathbf{S}_+^n$ so objective is a convex quadratic function



quadratically constrained quadratic program (QCQP):

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

- $P_0, \ldots, P_m \in \mathbf{S}_+^n$ so objective and constraints are convex quadratics
- if $P_1, \ldots, P_m \in \mathbf{S}_{++}^n$, then feasible set is intersection of m ellipsoids and an affine set
- QCQP reduces to
  - QP if $P_1, \ldots, P_m = 0$
  - LP if $P_0, P_1, \ldots, P_m = 0$

examples:

- least squares problem

$$\text{minimize} \quad \|Ax - b\|_2^2 = x^T A^T A x - 2b^T A x + b^T b$$

  is a QP with $P = 2A^T A$, $q = -2A^T b$, $r = b^T b$

- norm bounded least squares problem

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & \|x\|_2 \leq 1 \end{array}$$

  is equivalent to the QCQP

$$\begin{array}{ll} \text{minimize} & x^T A^T A x - 2b^T A x + b^T b \\ \text{subject to} & x^T x - 1 \leq 0 \end{array}$$

Semidefinite program (SDP)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b \end{array}$$

- $F_1, \ldots, F_n, G \in \mathbf{S}^m$
- inequality constraint is called *linear matrix inequality* (LMI)
- includes problems with multiple LMI constraints, *e.g.*,

$$x_1 F_1^{(i)} + \cdots + x_n F_n^{(i)} + G^{(i)} \preceq 0, \quad i = 1, \ldots, k$$

  is equivalent to a single LMI constraint

$$\sum_{i=1}^{n} x_i \begin{bmatrix} F_i^{(1)} & & \\ & \ddots & \\ & & F_i^{(k)} \end{bmatrix} + \begin{bmatrix} G^{(1)} & & \\ & \ddots & \\ & & G^{(k)} \end{bmatrix} \preceq 0$$

examples:

- the inequality form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

  is equivalent to the SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{diag}(Ax - b) \preceq 0 \end{array}$$

  (note different parsing of the inequality $\preceq$)

- optimal value of the SDP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & A - tI \succeq 0 \end{array}$$

  finds the smallest eigenvalue $\lambda_{\min}(A)$ for $A \in \mathbf{S}^n$

3. Disciplined convex analysis and programming

- methods of convexity verification
- restriction of a function to a line
- second-order conditions
- constructive convex analysis
- disciplined convex programming

Methods of convexity verification

practical methods for establishing convexity of a set $C \subseteq \mathbf{R}^n$

1. apply definition

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

2. show that C is obtained from simple convex sets, *e.g.*,

   - hyperplanes and halfspaces
   - polyhedra and simplexes
   - balls and ellipsoids
   - . . .

   by operations that preserve convexity, *e.g.*,

   - intersection
   - Cartesian product
   - affine transformation
   - . . .

for establishing convexity of a function $f : \mathbf{R}^n \to \mathbf{R}$

1. apply definition (Jensen's inequality)

$$x_1, x_2 \in \mathbf{R}^n, \quad \theta \in [0,1] \implies f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta) f(x_2)$$

2. prove second-order condition (for twice differentiable f)

$$\nabla^2 f(x) \succeq 0 \text{ for all } x \in \mathbf{R}^n$$

3. show that f is obtained from simple convex functions by operations that preserve convexity, *e.g.*,
   - nonnegative weighted sum
   - pointwise maximum and supremum
   - composition with affine function
   - general composition
   - . . .

Restriction of a function to a line

show Jensen's inequality for $f\colon \mathbf{R}^n \to \mathbf{R}$ is equivalent to proving

$$g(t) = f(x+tv) \text{ convex (in } t\text{) for all } x \in \mathbf{dom}\, f,\ v \in \mathbf{R}^n$$

on its domain $\{t \in \mathbf{R} \mid x + tv \in \mathbf{dom}\, f\}$

i.e., f convex $\iff$ f convex along all lines in its domain

example: for $f\colon \mathbf{S}^n \to \mathbf{R}$ with $f(X) = \log\det X$, $\mathbf{dom}\, f = \mathbf{S}^n_+$,

$$\begin{aligned} g(t) &= \log\det(Z+tV) \\ &= \log\det Z + \log\det(I + tZ^{-1/2}VZ^{-1/2}) \\ &= \log\det Z + \textstyle\sum_{i=1}^n \log(1+t\lambda_i) \end{aligned}$$

where λ_i are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$

g concave in t for all $Z \succ 0$, V $\implies$ f concave

Second-order conditions

consider log-sum-exp function $f : \mathbf{R}^n \to \mathbf{R}$ given by

$$f(x) = \log \sum_{i=1}^{n} \exp x_i, \quad \mathbf{dom}\, f = \mathbf{R}^n$$

we have

$$\nabla^2 f(x) = \frac{1}{\left(\mathbf{1}^T z\right)^2} \left((\mathbf{1}^T z)\, \mathbf{diag}(z) - z z^T \right)$$

where $z = (e^{x_1}, \ldots, e^{x_n})$

to show $\nabla^2 f(x) \succeq 0$, we need to show for all $v \in \mathbf{R}^n$,

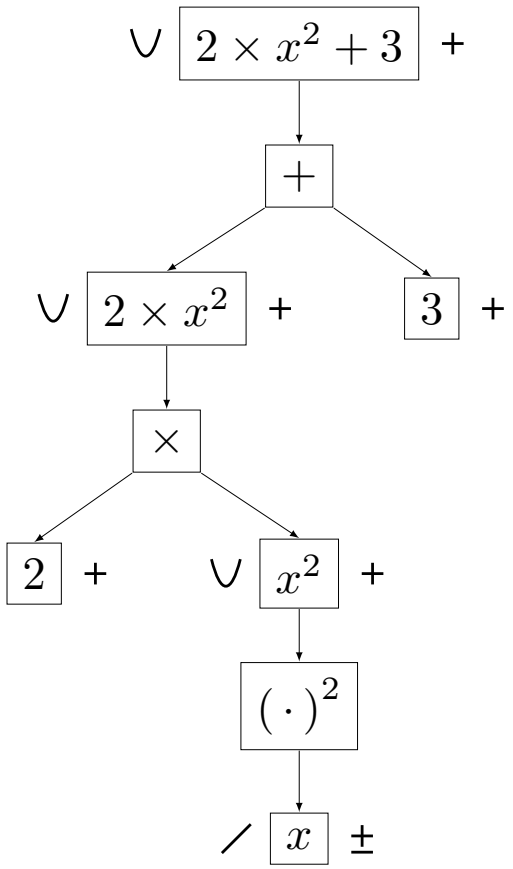
$$v^T \nabla^2 f(x) v = \frac{1}{\left(\mathbf{1}^T z\right)^2} \left(\left(\sum_{i=1}^{n} z_i \right) \left(\sum_{i=1}^{n} z_i v_i^2 \right) - \left(\sum_{i=1}^{n} z_i v_i \right)^2 \right) \geq 0$$

which holds by Cauchy-Schwarz inequality

Constructive convex analysis

basic ideas:

- parse target function f into a *composition tree*
- nodes consist of
  - atomic functions with known convexity, or
  - convexity preserving operations
- traverse the tree from leaves to root, check whether convexity is preserved



example: $f(x) = 2x^2 + 3$

Disciplined convex programming (DCP)

a DCP problem has the form

$$\begin{array}{ll} \text{minimize/maximize} & f(x) \\ \text{subject to} & p_i(x) \sim q_i(x), \quad i = 1, \ldots, m \end{array}$$

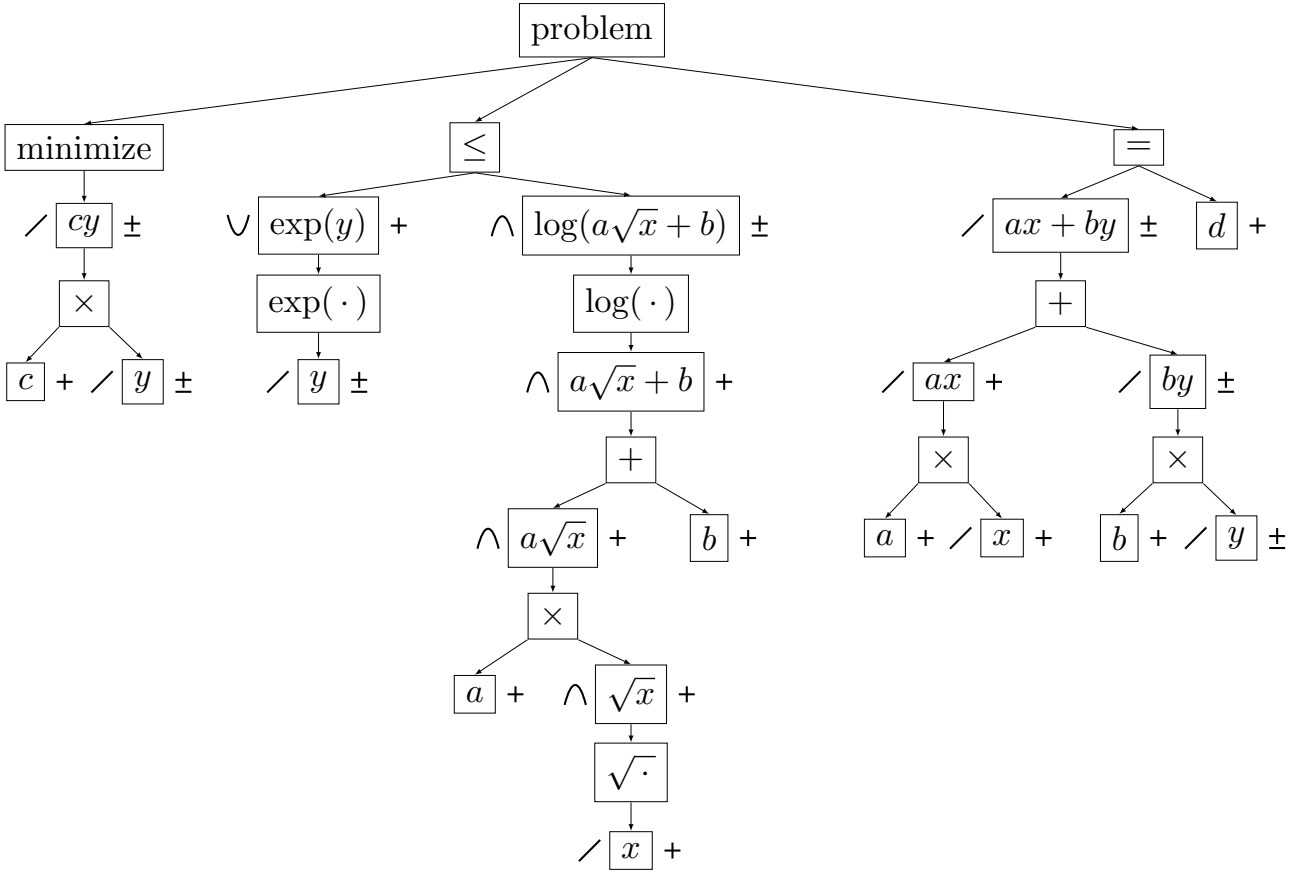
- variable $x \in \mathbf{R}^n$, objective $f\colon \mathbf{R}^n \to \mathbf{R}$, constraints $p_i, q_i\colon \mathbf{R}^n \to \mathbf{R}$
- relational operator $\sim$ is one of $=$, $\leq$, and $\geq$

convexity follows from the following rules:

- f convex if minimization, concave if maximization
- p_i convex, q_i concave if $\sim$ is $\leq$
- p_i concave, q_i convex if $\sim$ is $\geq$
- p_i and q_i affine if $\sim$ is $=$

example: DCP problem with variables $x \in \mathbf{R}_+$, $y \in \mathbf{R}$

$$
\begin{array}{ll}
\text{minimize} & cy \\
\text{subject to} & \exp(y) \leq \log(a\sqrt{x} + b) \\
& ax + by = d,
\end{array}
$$



4. Sequential convex programming

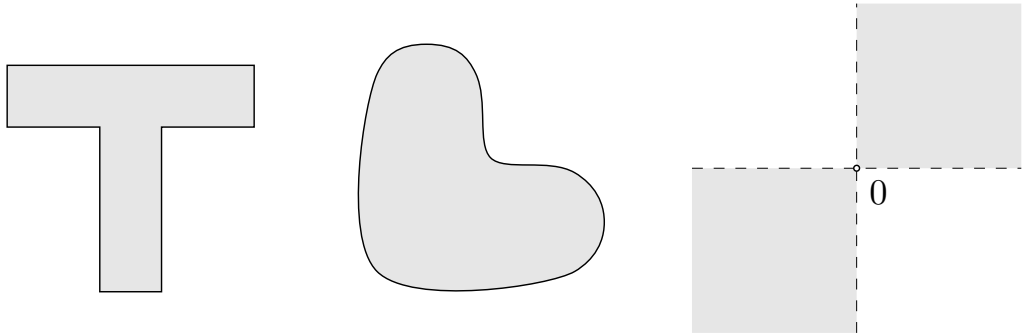
- biconvex sets and functions
- functional operations preserving biconvexity
- biconvex optimization problems
- alternate convex search
- difference-of-convex functions
- first-order majorization
- difference-of-convex programming
- convex-concave procedure
- sequential convex approximation
- Taylor expansions and particle methods

Biconvex sets

biconvex set $B \subseteq \mathbf{R}^n \times \mathbf{R}^k$ is convex when restricted to either variable:

- $B_{\tilde{y}} = \{x \in \mathbf{R}^n \mid (x, \tilde{y}) \in B\}$ convex for all $\tilde{y} \in \mathbf{R}^k$
- $B_{\tilde{x}} = \{y \in \mathbf{R}^k \mid (\tilde{x}, y) \in B\}$ convex for all $\tilde{x} \in \mathbf{R}^n$

examples:



- solution set of a system of *bilinear equations* is biconvex:

$$B = \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^k \mid (Ax + b)^T (Cy + d) = 0\}$$

- intersection of (potentially infinite) biconvex sets is biconvex

Biconvex functions

$f : \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R}$ is **biconvex function** if

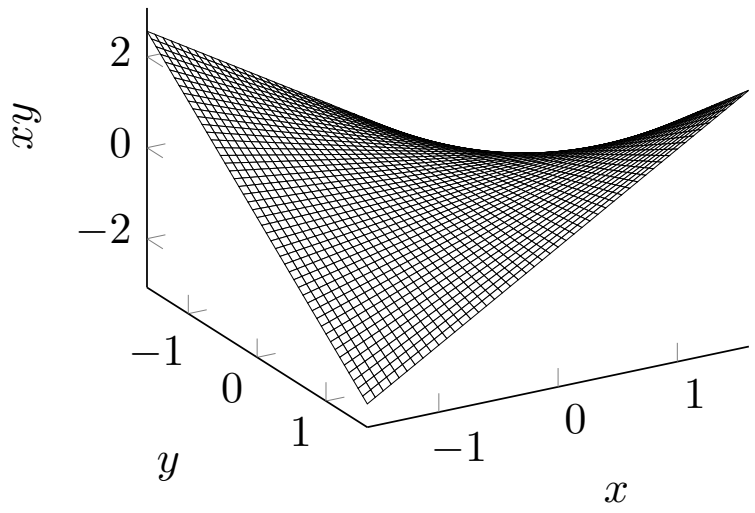
- $\mathbf{dom}\, f$ is a biconvex set
- $f_{\tilde{y}} : \mathbf{R}^n \to \mathbf{R}$ given by $x \mapsto f(x, \tilde{y})$ is convex in x
- $f_{\tilde{x}} : \mathbf{R}^k \to \mathbf{R}$ given by $y \mapsto f(\tilde{x}, y)$ is convex in y

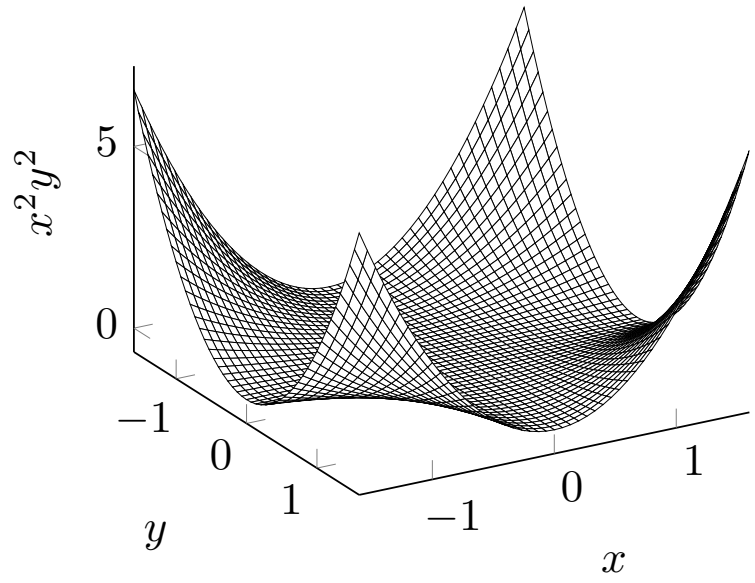
i.e., f is convex in each variable when the other is fixed

biconcave, **bilinear**, and **biaffine** functions are defined similarly

examples:

- $f(x, y) = x^T y$ is bilinear (hence biaffine)
- $f(x, y) = (Ax + b)^T (Cy + d)$ is biaffine (hence biconvex and biconcave)
- $f(X, Y) = XY - A$ is biaffine (on $\mathbf{R}^{n \times m} \times \mathbf{R}^{m \times k}$)
- $f(x, y) = x^2 y^2$ is biconvex





Sublevel sets

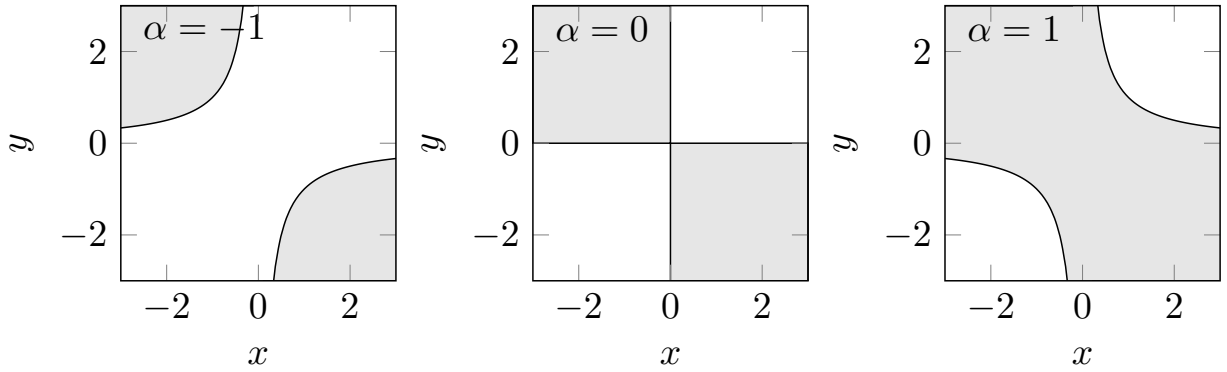
α-sublevel set of biconvex function $f : \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R}$:

$$C_\alpha = \{(x, y) \in \mathbf{dom}\, f \mid f(x, y) \leq \alpha\}$$

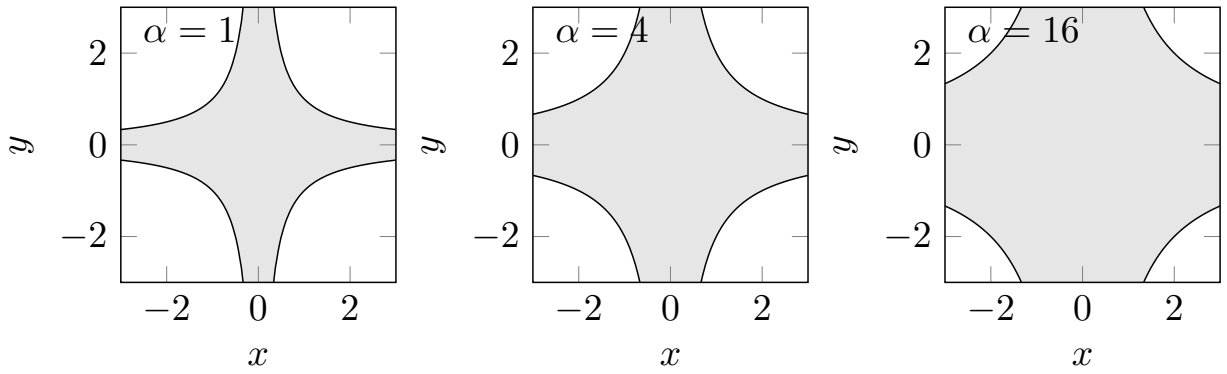
is biconvex

examples:

- $f(x, y) = xy$:



- $f(x, y) = x^2y^2$:



Functional operations preserving biconvexity

nonnegative weighted sum of biconvex functions is biconvex

pointwise maximum and supremum of biconvex functions is biconvex

biaffine composition:

$$h\colon \mathbf{R} \to \mathbf{R} \text{ convex} \implies h((Ax+b)^T(Cy+d)) \text{ biconvex}$$

general composition:

$$\begin{array}{l} h\colon \mathbf{R} \to \mathbf{R} \text{ convex nondecreasing} \\ g\colon \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R} \text{ biconvex} \end{array} \implies h(g(x,y)) \text{ biconvex}$$

- can be generalized to multivariate h

example: matrix factorization cost

$$f(X,Y) = \|XY - A\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{k}\left((XY)_{ij} - A_{ij}\right)^2$$

- $A \in \mathbf{R}^{n\times k}$ is a given matrix

- composition of convex function $h(Z) = \|Z\|_F^2$ with biaffine function $g(X,Y) = XY - A$

Biconvex optimization problems

standard form biconvex optimization problem:

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

- $f_i\colon \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R}$ biconvex for $i = 0, 1, \ldots, m$
- $h_i\colon \mathbf{R}^n \times \mathbf{R}^k \to \mathbf{R}$ biaffine for $i = 1, \ldots, p$
- not convex in the most general case

feasible set:

$$\mathcal{D} = \left\{ (x,y) \in \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i \;\middle|\; \begin{array}{l} f_i(x,y) \leq 0,\ i=1,\ldots,m \\ h_i(x,y) = 0,\ i=1,\ldots,p \end{array} \right\}$$

is the intersection of biconvex sets and hence biconvex

Optimal value and points

feasible point $(x^\star, y^\star) \in \mathcal{D}$ is **partially optimal** if

- $f_0(x^\star, y^\star) \leq f_0(x, y^\star)$ for all $x \in \mathcal{D}_{y^\star} = \{x \mid (x, y^\star) \in \mathcal{D}\}$
- $f_0(x^\star, y^\star) \leq f_0(x^\star, y)$ for all $y \in \mathcal{D}_{x^\star} = \{y \mid (x^\star, y) \in \mathcal{D}\}$

$f(x^\star, y^\star)$ called **partially optimal value**

- every stationary point of differentiable f_0 over $\mathcal{D}$ is partially optimal, and vice versa
- not necessarily a local optimum (and of course global optimum)
- turns out to be quite useful in practice
- no stronger optimality conditions exist for general biconvex problems

Alternate convex search (ACS)

iterate between solving convex subproblems in x and y:

Alternate convex search.

given feasible starting point $(x^{(0)}, y^{(0)})$.
$k := 0$.
repeat

1. $x^{(k+1)} := \underset{x \in \mathbf{R}^n}{\operatorname{argmin}} \left\{ f_0(x, y^{(k)}) \;\middle|\; \begin{array}{l} f_i(x, y^{(k)}) \leq 0, \; i = 1, \ldots, m \\ h_i(x, y^{(k)}) = 0, \; i = 1, \ldots, p \end{array} \right\}.$

2. $y^{(k+1)} := \underset{y \in \mathbf{R}^k}{\operatorname{argmin}} \left\{ f_0(x^{(k+1)}, y) \;\middle|\; \begin{array}{l} f_i(x^{(k+1)}, y) \leq 0, \; i = 1, \ldots, m \\ h_i(x^{(k+1)}, y) = 0, \; i = 1, \ldots, p \end{array} \right\}.$

3. $k := k + 1$.

until stopping criteria are satisfied.

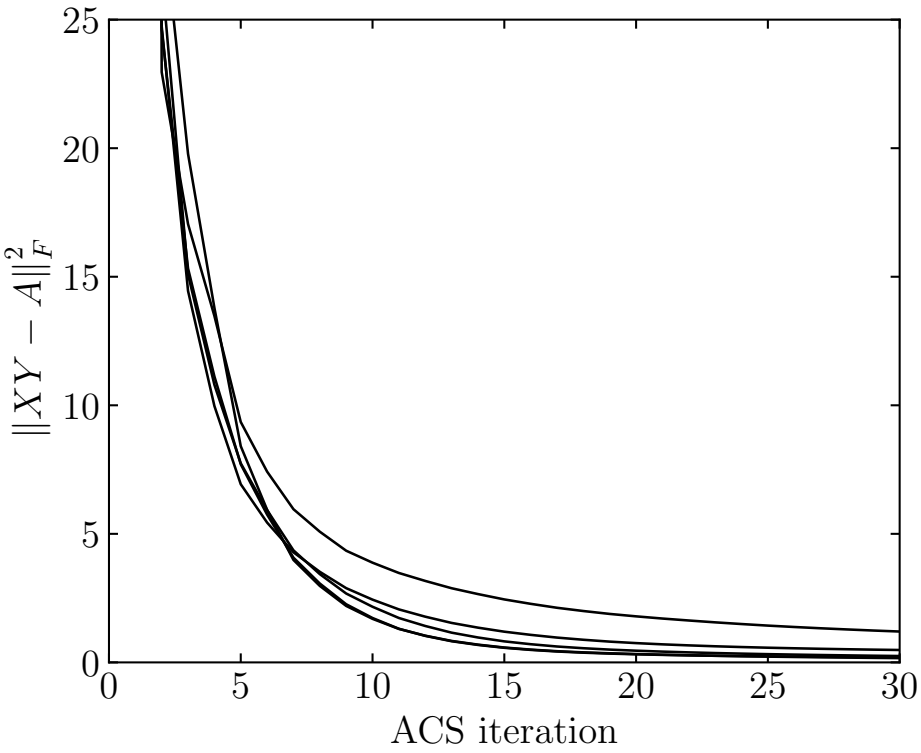
- objective value is nonincreasing
- converges to stationary point (hence partially optimal)

example: nonnegative matrix factorization

$$\begin{array}{ll} \text{minimize} & \|XY - A\|_F^2 \\ \text{subject to} & X_{ij} \geq 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, k \\ & Y_{ij} \geq 0, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n \end{array}$$

- variables $X \in \mathbf{R}^{m \times k}$ and $Y \in \mathbf{R}^{k \times n}$, $A \in \mathbf{R}^{m \times n}$ is given matrix

for $m = n = 50$, $k = 5$,
multiple random initializations:



Difference-of-convex functions

difference-of-convex (DC) function $h\colon \mathbf{R}^n \to \mathbf{R}$ has the form

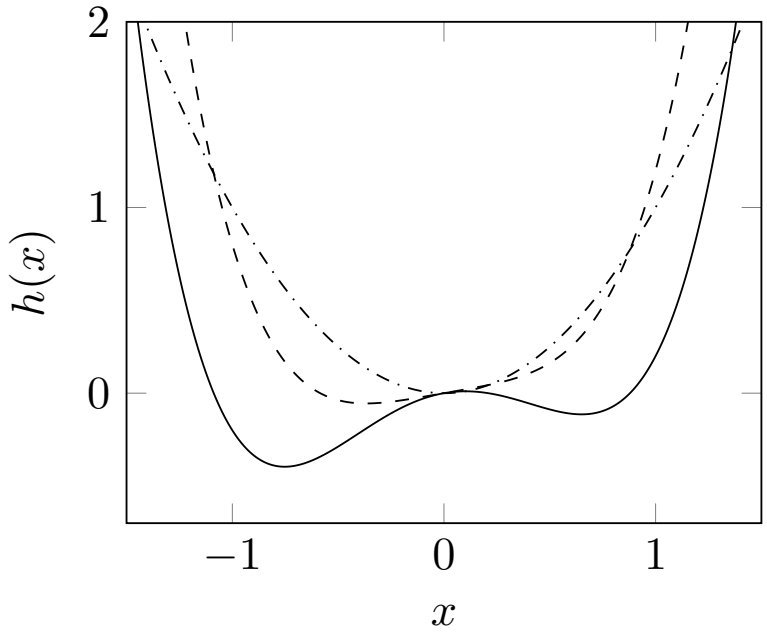
$$h(x) = f(x) - g(x), \quad f,\, g \text{ convex}$$

- any convex f is DC with $f(x) = f(x) - 0$
- any concave h is DC with $h(x) = 0 - (-h(x))$

examples:

- $h(x) = x^4 + (1/5)x - x^2$ is DC with

  $$f(x) = x^4 + (1/5)x, \quad g(x) = x^2$$



- nonconvex quadratic function

$$h(x) = x^T P x + q^T x + r, \quad P \in \mathbf{S}^n,\ q \in \mathbf{R}^n,\ r \in \mathbf{R}$$

  is expressed in DC form as

$$h(x) = x^T P_+ x + q^T x + r - x^T P_- x$$

  where $P_+ \succeq 0$, $P_- \succeq 0$, and $P = P_+ - P_-$
  (obtained from eigenvalue decomposition)

First-order majorization

for DC function $h(x) = f(x) - g(x)$ and differentiable g:

- 1st-order Taylor expansion of g at $\tilde{x}$ is affine function $\hat{g}_{\tilde{x}} : \mathbf{R}^n \to \mathbf{R}$:

$$\hat{g}_{\tilde{x}}(x) = g(\tilde{x}) + \nabla g(\tilde{x})^T (x - \tilde{x}) \leq g(x)$$

  which is a global underestimator of g

- define $\hat{h}_{\tilde{x}} : \mathbf{R}^n \to \mathbf{R}$ as

$$\hat{h}_{\tilde{x}}(x) = f(x) - \hat{g}_{\tilde{x}}(x) = f(x) - g(\tilde{x}) - \nabla g(\tilde{x})^T (x - \tilde{x})$$

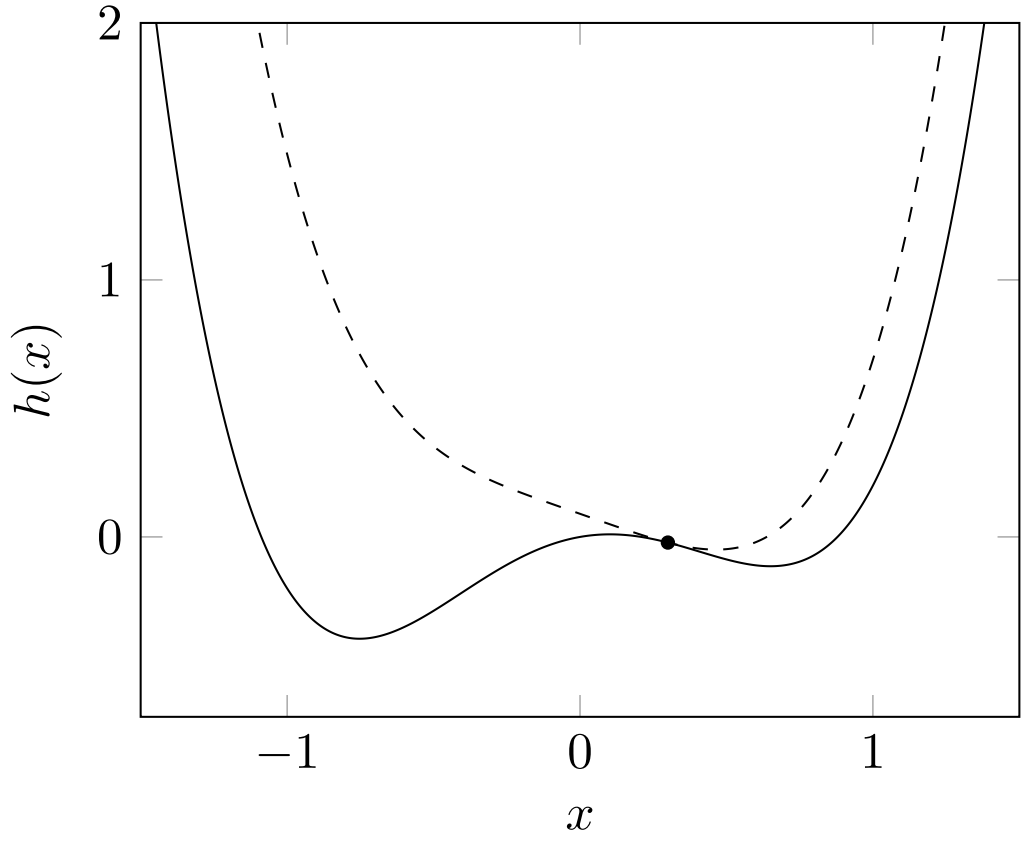
  – $\hat{h}_{\tilde{x}}(x) \geq f(x) - g(x) = h(x) \implies \hat{h}_{\tilde{x}}$ is a global *majorizer* of h

  – $\hat{h}_{\tilde{x}}(\tilde{x}) = h(\tilde{x}) \implies$ majorizer is tight at $\tilde{x}$

example: 1st-order majorization of DC function $h(x) = x^4 + (1/5)x - x^2$

$$\begin{aligned}\hat{h}_{\tilde{x}}(x) &= x^4 + (1/5)x - (\tilde{x}^2 + 2\tilde{x}(x - \tilde{x})) \\ &= x^4 + (1/5)x - 2\tilde{x}x + \tilde{x}^2\end{aligned}$$

at $\tilde{x} = 0.3$:



Difference-of-convex programming

difference-of-convex optimization problem has the form

$$\begin{array}{ll} \text{minimize} & f_0(x) - g_0(x) \\ \text{subject to} & f_i(x) - g_i(x) \leq 0, \quad i = 1, \ldots, m \end{array}$$

- f_i, g_i convex for $i = 0, 1, \ldots, m$
- nonconvex in general, reduces to convex program if g_i are all affine
- includes equality constraints, *e.g.*,

$$p_i(x) = q_i(x), \quad i = 1, \ldots, k$$

with convex p_i, q_i as

$$p_i(x) - q_i(x) \leq 0, \quad q_i(x) - p_i(x) \leq 0, \quad i = 1, \ldots, k$$

example: boolean least squares problem

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & x_i \in \{0,1\}, \quad i=1,\ldots,n \end{array}$$

is expresses as a DC program as

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & x_i^2 - x_i \leq 0, \quad i=1,\ldots,n \\ & x_i - x_i^2 \leq 0, \quad i=1,\ldots,n, \end{array}$$

since for $i=1,\ldots,n$:

$$\begin{array}{lcll} x_i \in \{0,1\} & \iff & x_i(1-x_i)=0 & \\ & \iff & x_i(1-x_i) \leq 0 & \text{and} \quad -x_i(1-x_i) \leq 0 \end{array}$$

Convex-concave procedure (CCP)

iteratively solve convex approximations of DC program:

Convex-concave procedure.

given a feasible starting point $x^{(0)}$.
$k := 0$.
repeat
1. *Convexify.*

$$\hat{g}_{i,x^{(k)}}(x) := g_i(x^{(k)}) + \nabla g_i(x^{(k)})^T(x - x^{(k)}), \quad i = 0, 1, \ldots, m$$

2. *Solve the convex approximation.*

$$\begin{array}{lll} x^{(k+1)} := \mathrm{argmin}_x & f_0(x) - \hat{g}_{0,x^{(k)}}(x) & \\ \text{subject to} & f_i(x) - \hat{g}_{i,x^{(k)}}(x) \leq 0, & i = 1, \ldots, m \end{array}$$

3. $k := k + 1$.

until stopping criteria are satisfied.

---

interpretation: iteratively find a point

- minimizing a global upper bound of the DC objective, since

$$f_0(x) - g_0(x) \leq f_0(x) - \hat{g}_{0,x^{(k)}}(x), \quad k = 0, 1, \ldots$$

- over a convex restriction of the original feasible set, since

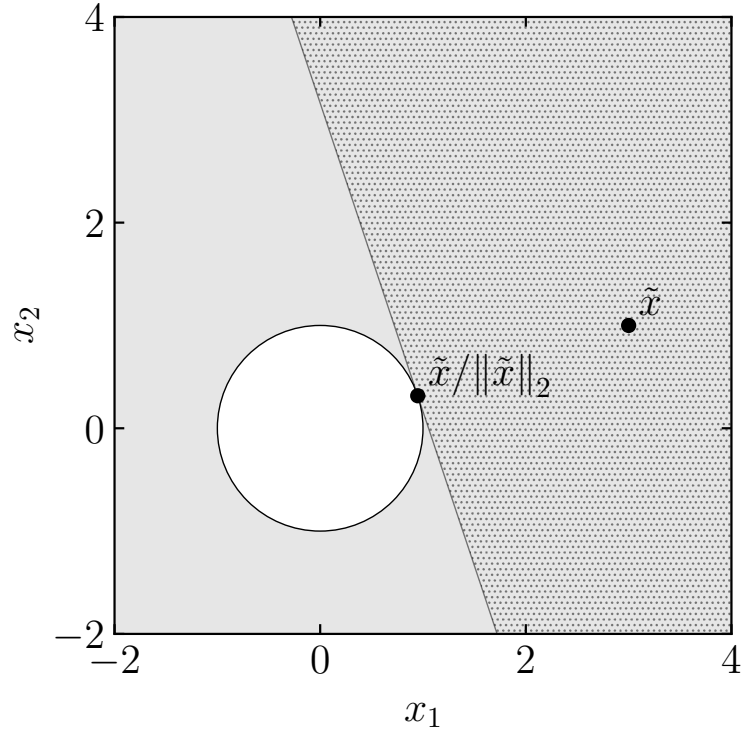
$$f_i(x) - g_i(x) \leq f_i(x) - \hat{g}_{i,x^{(k)}}(x) \leq 0, \quad k = 0, 1, \ldots$$

example:

- shaded region: $\|x\|_2 \geq 1$
- dotted region: approximation

$$\frac{\tilde{x}^T}{\|\tilde{x}\|_2} x \geq 1$$

  in each CCP iteration at $\tilde{x}$



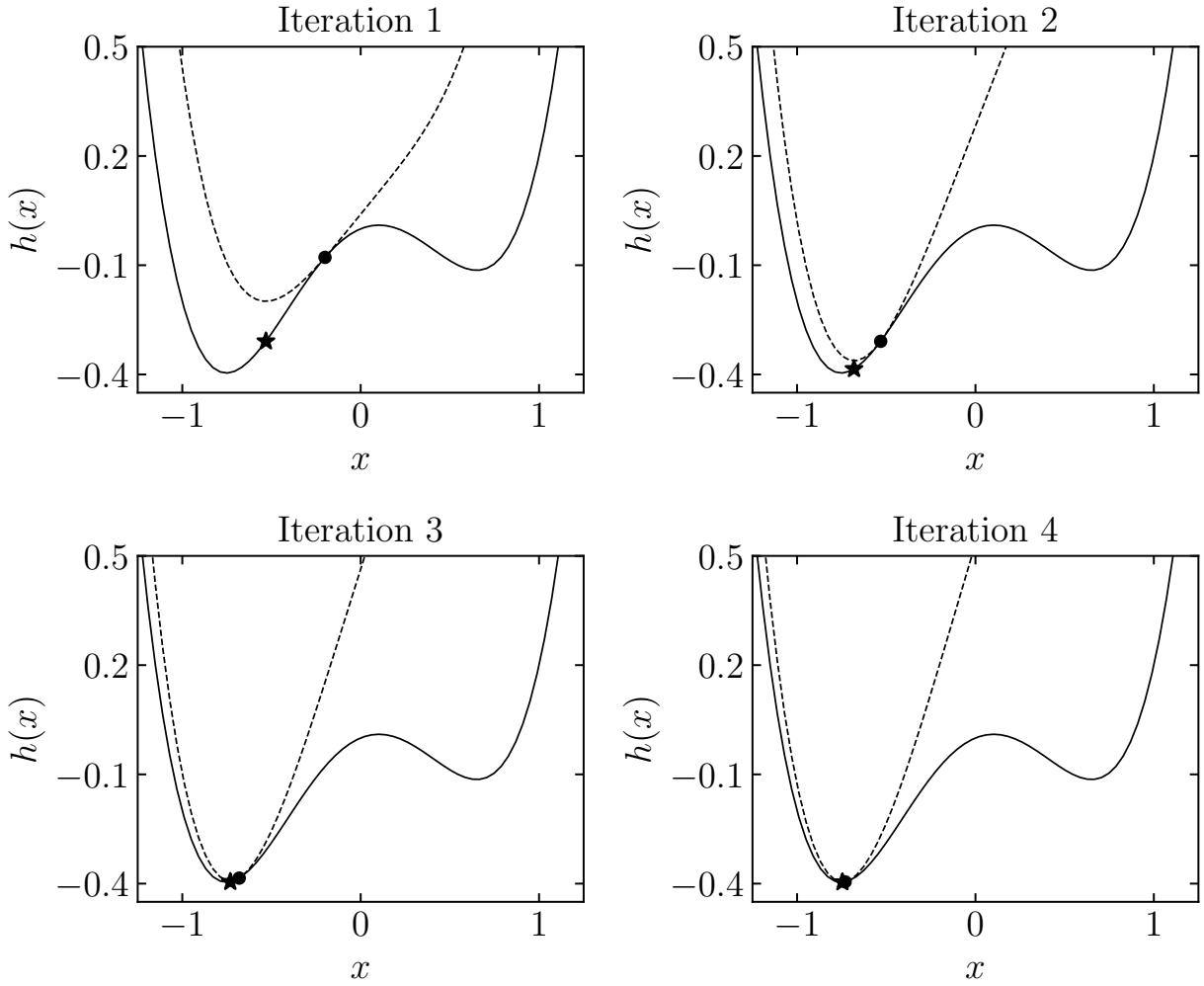
Minimizing a DC function

unconstrained DC program

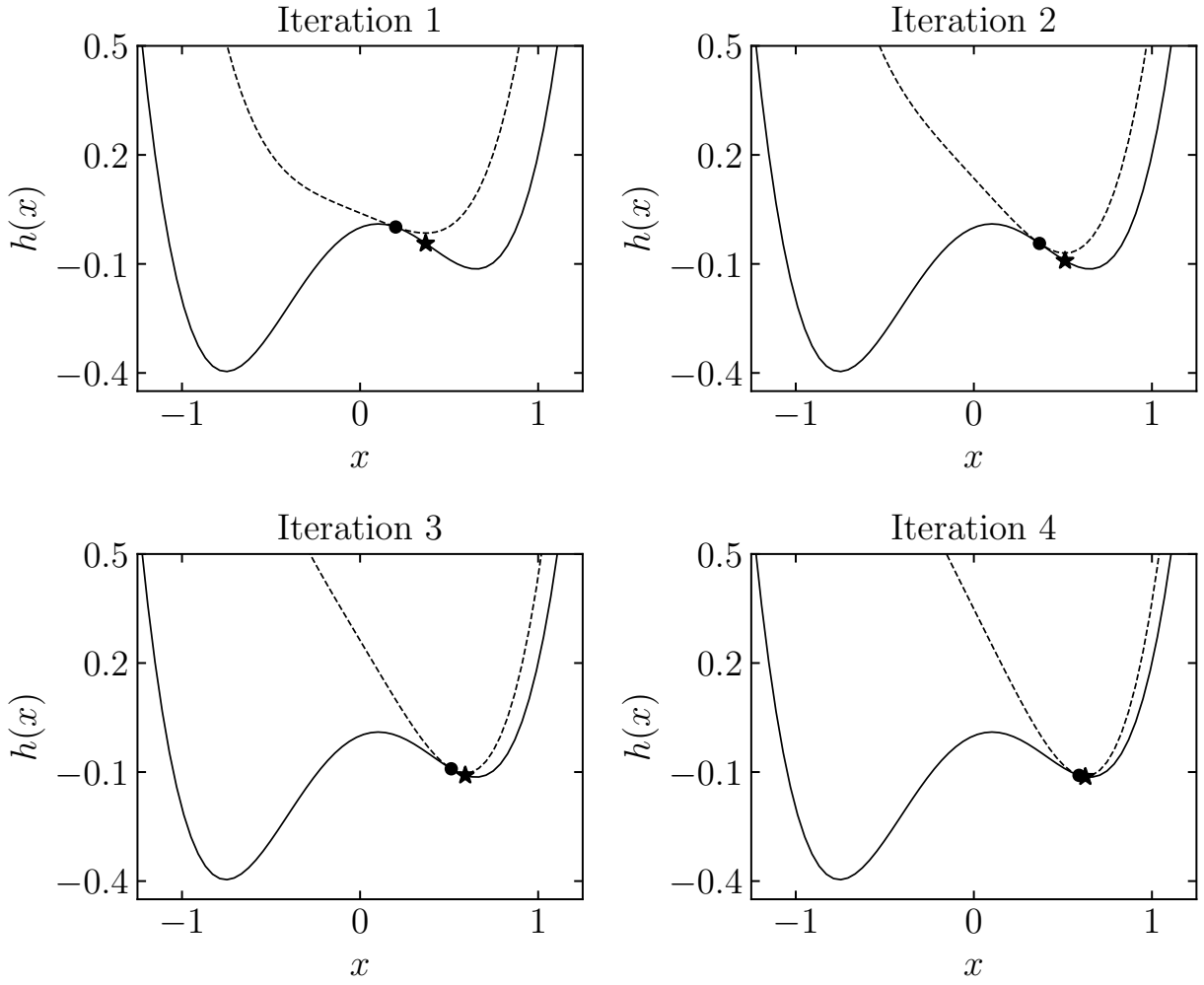
minimize $x^4 + (1/5)x - x^2$

CCP iteration at $\tilde{x}$ solves

minimize $x^4 + (1/5)x - 2\tilde{x}x + \tilde{x}^2$



CCP is initialization-dependent:



Closest point outside unit ball

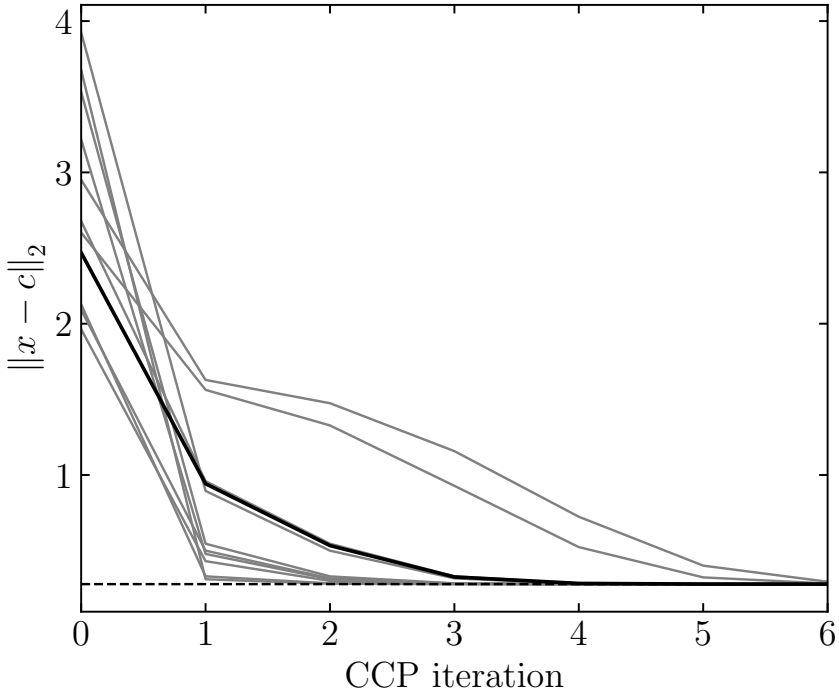
find $x \in \mathbf{R}^n$ closest to given point $c \in \mathbf{R}^n$ outside the unit ball:

$$\begin{array}{ll} \text{minimize} & \|x - c\|_2 \\ \text{subject to} & \|x\|_2 \geq 1 \end{array}$$

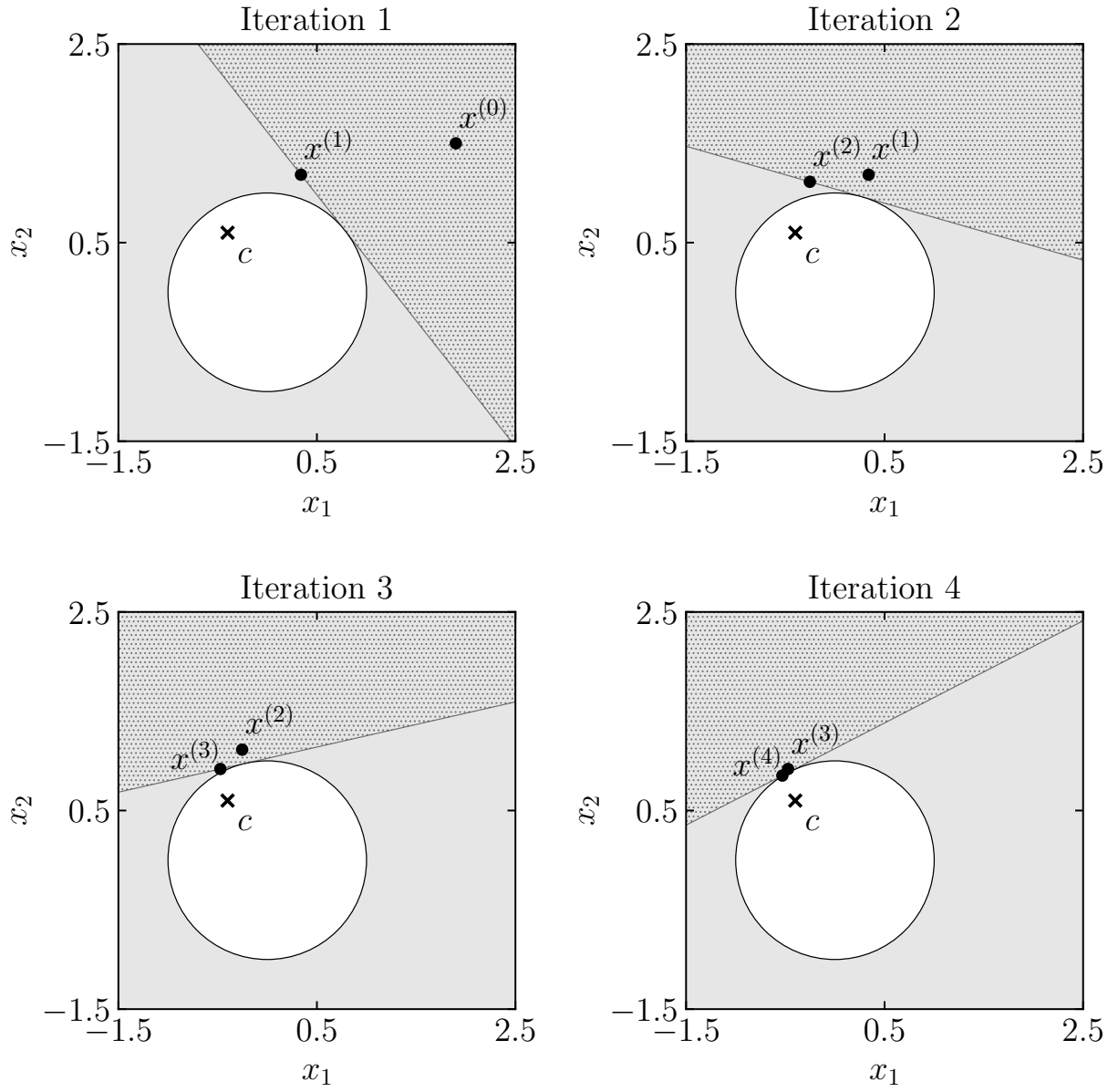
CCP iteration at $\tilde{x}$ solves

$$\begin{array}{ll} \text{minimize} & \|x - c\|_2 \\ \text{subject to} & (1/\|\tilde{x}\|_2)\tilde{x}^T x \geq 1 \end{array}$$

right shows example for $n = 2$,
10 random initializations:



results of initial 4 iterations:



Sequential convex approximation

given general nonlinear program with variable $x \in \mathbf{R}^n$

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

- $f_i \colon \mathbf{R}^n \to \mathbf{R}$ not necessarily convex for $i = 0, 1, \ldots, m$
- $h_i \colon \mathbf{R}^n \to \mathbf{R}$ not necessarily affine for $i = 1, \ldots, p$

basic idea: at current point $\tilde{x} \in \mathbf{R}^n$ iteratively form

- convex approximation $\hat{f}_{i,\tilde{x}}$ of f_i for $i = 0, 1, \ldots, m$
- affine approximation $\hat{h}_{i,\tilde{x}}$ of h_i for $i = 1, \ldots, p$
- over some *trust region* $\mathcal{T}_{\tilde{x}} \subseteq \mathbf{R}^n$ around $\tilde{x}$

so we end up solving a sequence of convex approximations

$$\begin{array}{ll} \text{minimize} & \hat{f}_{0,\tilde{x}}(x) \\ \text{subject to} & \hat{f}_{i,\tilde{x}}(x) \leq 0, \quad i = 1, \ldots, m \\ & \hat{h}_{i,\tilde{x}}(x) = 0, \quad i = 1, \ldots, p \\ & x \in \mathcal{T}_{\tilde{x}} \end{array}$$

trust region $\mathcal{T}_{\tilde{x}}$:

- Euclidean ball with center $\tilde{x}$ and radius $\rho > 0$:

$$\mathcal{T}_{\tilde{x}} = \{x \in \mathbf{R}^n \mid \|x - \tilde{x}\|_2 \leq \rho\}$$

- box centered at $\tilde{x}$ bounded by $\rho \in \mathbf{R}^n_{++}$:

$$\mathcal{T}_{\tilde{x}} = \{x \in \mathbf{R}^n \mid |x_i - \tilde{x}_i| \leq \rho_i,\ i = 1, \ldots, n\}$$

  - $\rho_i = \infty$ if x_i appears only in convex f_i or affine h_i

interpretation: the next point should not be too far from the current point $\tilde{x}$, so that the approximations are valid

Taylor expansions

for twice differentiable $f : \mathbf{R}^n \to \mathbf{R}$

affine approximation at $\tilde{x}$:

$$\hat{f}_{\tilde{x}}(x) = f(\tilde{x}) + \nabla f(\tilde{x})^T (x - \tilde{x})$$

convex approximation at $\tilde{x}$:

$$\hat{f}_{\tilde{x}}(x) = f(\tilde{x}) + \nabla f(\tilde{x})^T (x - \tilde{x}) + (1/2)(x - \tilde{x})^T H_+ (x - \tilde{x})$$

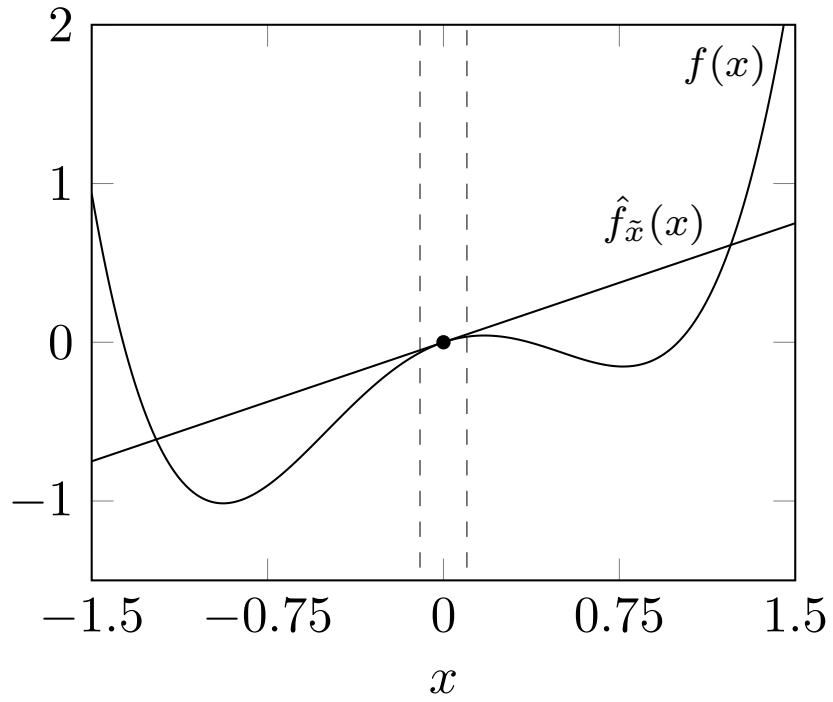
- $H_+ \succeq 0$ is positive semidefinite part of Hessian $\nabla^2 f(\tilde{x})$, *i.e.*,

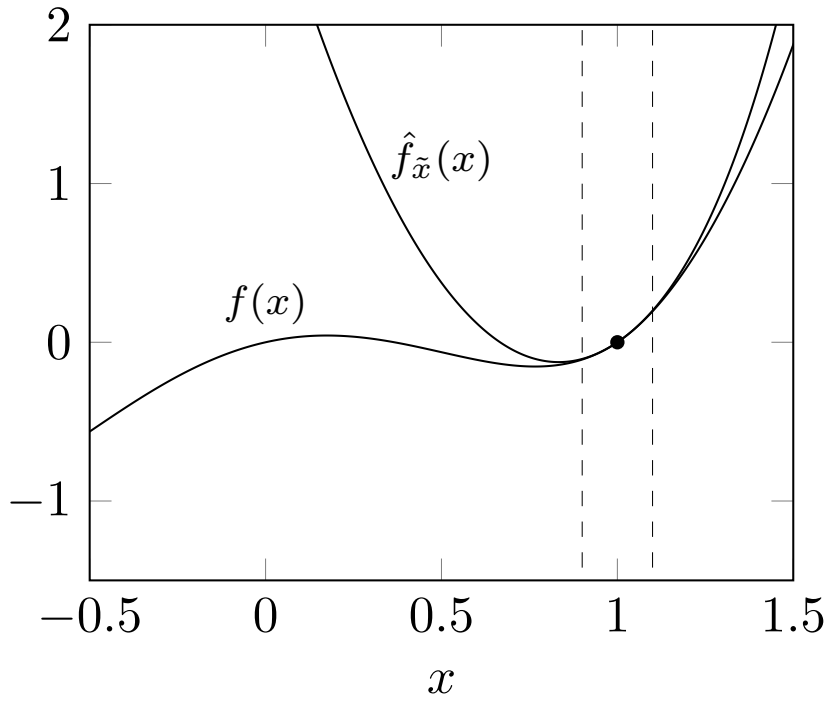
$$H_+ = Q \Lambda_+ Q^T$$

  where $\nabla^2 f(\tilde{x}) = Q \Lambda Q^T$,

$$\Lambda_+ = \mathbf{diag}(\max\{\lambda_1, 0\}, \ldots, \max\{\lambda_n, 0\})$$

examples:





Taylor expansions are **local** approximations, *i.e.*,

- if x is close to $\tilde{x}$, then approximation is very accurate
- approximation invariant to selection of $\mathcal{T}_{\tilde{x}}$
- but large $\mathcal{T}_{\tilde{x}}$ may lead to very poor approximation (in the whole $\mathcal{T}_{\tilde{x}}$)

Particle methods

given target function $f : \mathbf{R}^n \to \mathbf{R}$ and trust region $\mathcal{T}_{\tilde{x}}$:

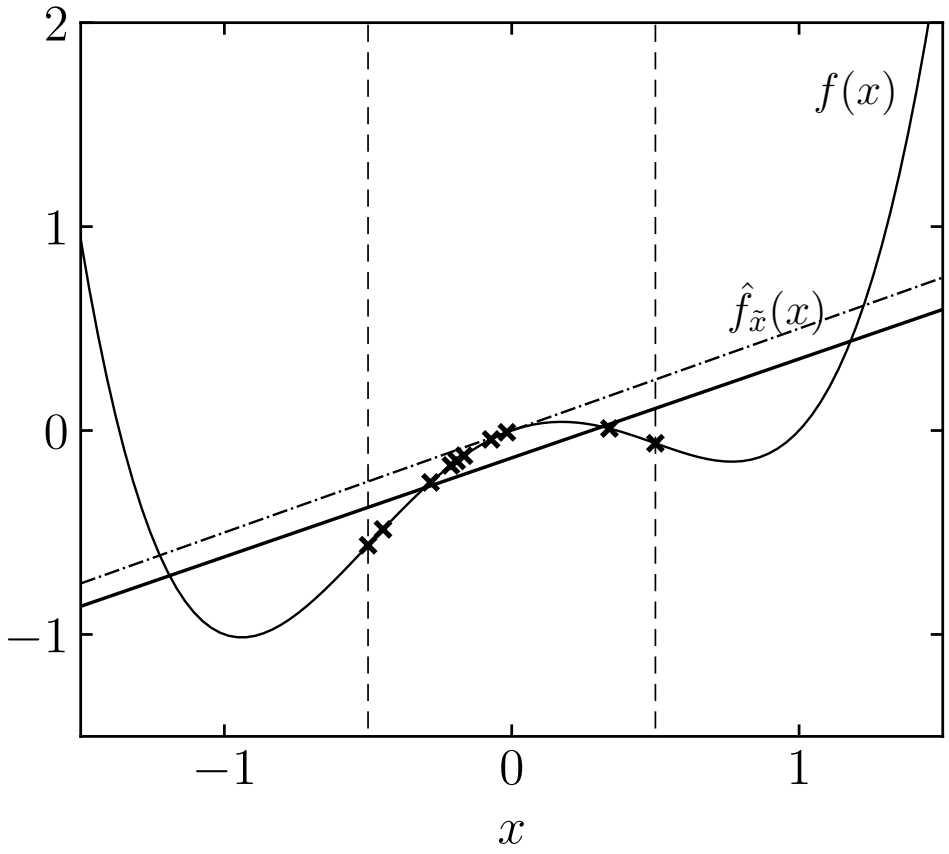
- sample K points $z_1, \ldots, z_K \in \mathcal{T}_{\tilde{x}}$, evaluate $y_i = f(z_i)$ for $i = 1, \ldots, K$
- so we have *particles* (z_i, y_i), $i = 1, \ldots, K$
- fit affine or convex function $\hat{f}_{\tilde{x}}$ to the particles (via, $e.g.$, least squares)

- allow approximation of nondifferentiable functions
- provide **'regional'** approximations, roughly speaking,
  - approximation is valid in the whole $\mathcal{T}_{\tilde{x}}$
  - accuracy not so sensitive to the size of $\mathcal{T}_{\tilde{x}}$ (compared to local approximations)
- approximation depends on $\mathcal{T}_{\tilde{x}}$ and the particles (z_i, y_i)
- needs extraordinary sampling and function evaluation

examples:

- affine approximation $\hat{f}_{\tilde{x}}(x) = a^T(x - \tilde{x}) + b$ fitted to particles (z_i, y_i):

$$\text{minimize} \quad \sum_{i=1}^{K} \left(a^T(z_i - \tilde{x}) + b - y_i\right)^2$$

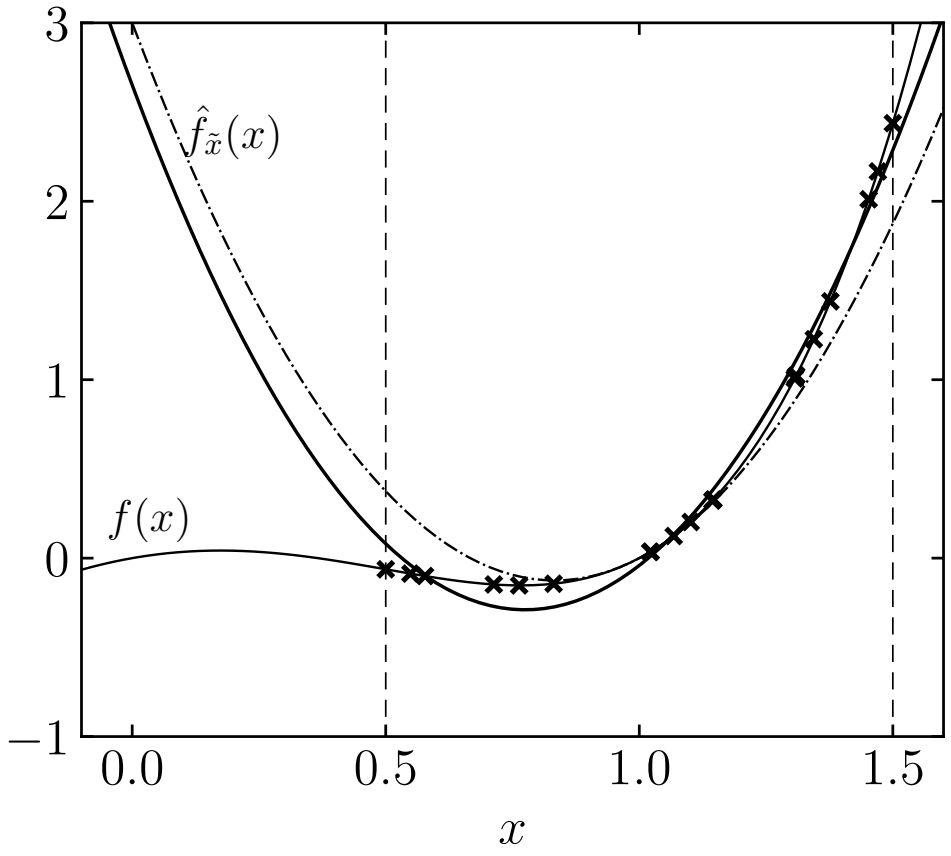


dashdotted line shows Taylor expansion at $\tilde{x}$

- fit convex quadratic $\hat{f}_{\tilde{x}}(x) = (1/2)(x-\tilde{x})^T P(x-\tilde{x}) + q^T(x-\tilde{x}) + r$:

  minimize $\sum_{i=1}^K \left((1/2)(z_i-\tilde{x})^T P(z_i-\tilde{x}) + q^T(z_i-\tilde{x}) + r - y_i\right)^2$
  subject to $P \succeq 0$

  (least squares problem with variables $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, $r \in \mathbf{R}$)



Fit convex envelope to particles

given particles $(z_i, y_i = f(z_i))$, $i = 1, \ldots, K$, fit $h\colon \mathbf{R}^n \to \mathbf{R}$,

$$h(x) = \max\{h_i + g_i^T(x - z_i) \mid i = 1, \ldots, K\}$$

with parameters $h_i \in \mathbf{R}$, $g_i \in \mathbf{R}^n$ for $i = 1, \ldots, K$

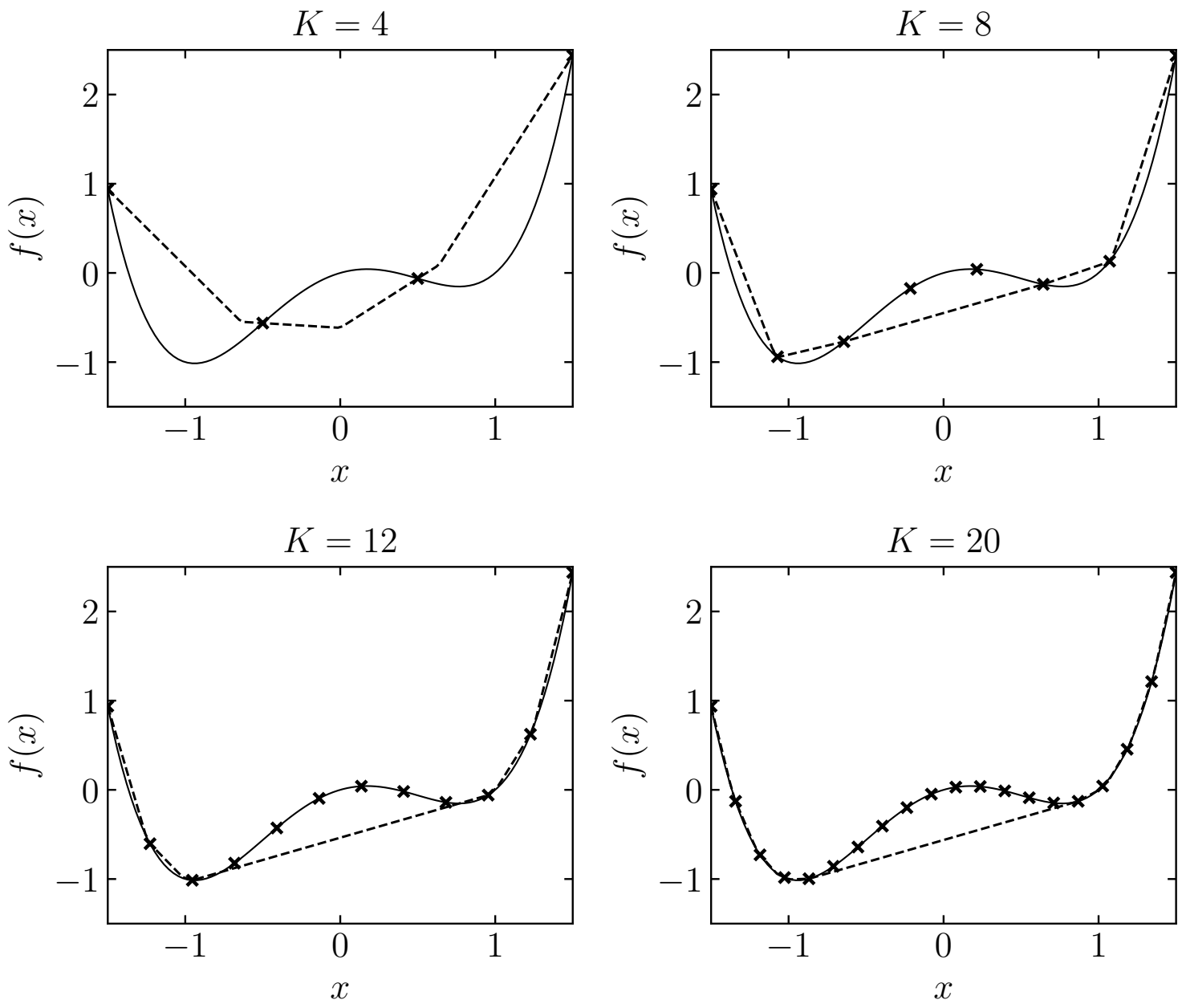
- h is always convex, since it is the pointwise maximum of affine functions
- require h lower bounds f at the particles, for $i = 1, \ldots, K$:

$$\begin{aligned} & y_i \geq h(z_i) = \max\{h_j + g_j^T(z_i - z_j) \mid j = 1, \ldots, K\} \\ \iff \quad & y_i \geq h_j + g_j^T(z_i - z_j), \quad j = 1, \ldots, K \end{aligned}$$

- maximize $\sum_{i=1}^K h_i$ make h as tight as possible (since $h(z_i) \geq h_i$)

put together, we have the LP

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^K h_i \\ \text{subject to} & y_i \geq h_j + g_j^T(z_i - z_j), \quad i,\ j = 1, \ldots, K \end{array}$$


$K = 4$
$K = 8$
$K = 12$
$K = 20$
$f(x)$
x
2
1
0
−1
−1
0
1

Example nonconvex QP

for the nonconvex QP with $P \in \mathbf{S}^n$ indefinite

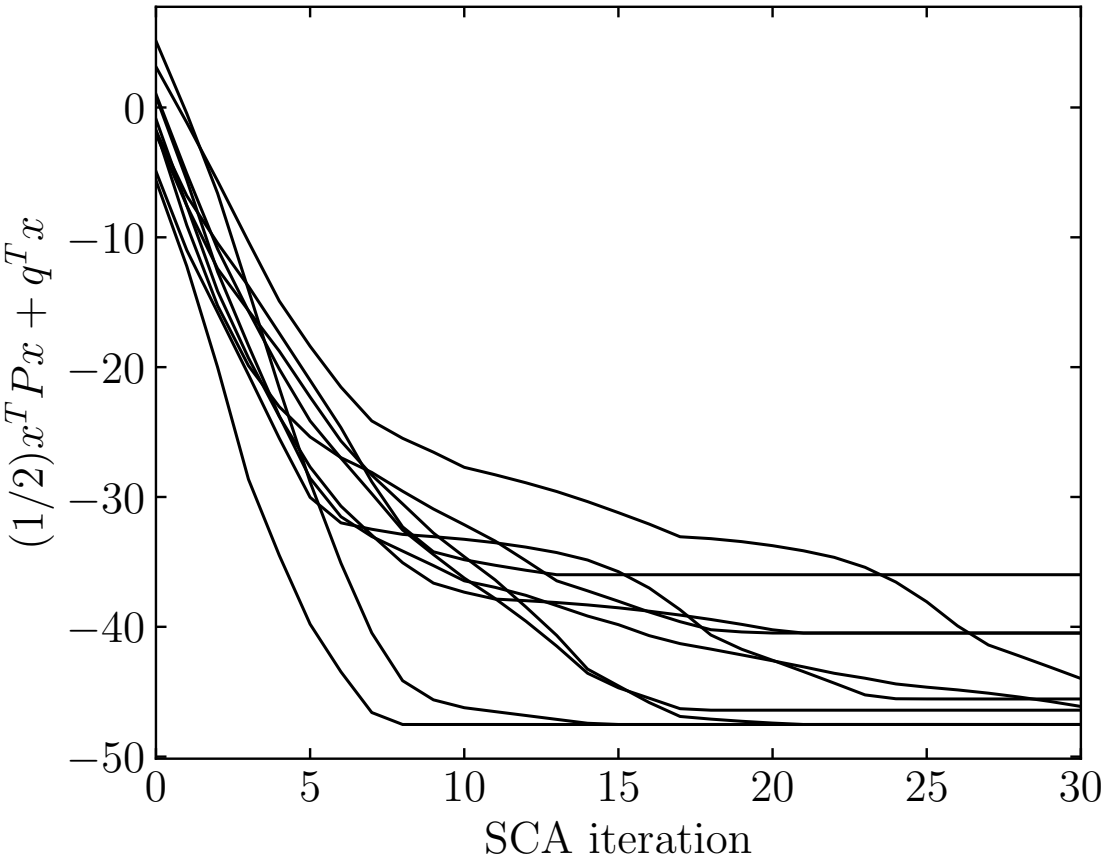
$$\begin{array}{ll} \text{minimize} & f(x) = (1/2)x^T P x + q^T x \\ \text{subject to} & \|x\|_\infty \leq 1 \end{array}$$

iteratively solve convex approximations

$$\begin{array}{ll} \text{minimize} & f(\tilde{x}) + (P\tilde{x} + q)^T (x - \tilde{x}) + (1/2)(x - \tilde{x})^T P_+ (x - \tilde{x}) \\ \text{subject to} & \|x\|_\infty \leq 1 \\ & \|x - \tilde{x}\|_\infty \leq \rho \end{array}$$

- convex objective from 2nd-order Taylor approximation at $\tilde{x}$
- $P_+ \succeq 0$ is positive semidefinite part of P
- $\rho > 0$ is fixed trust region radius

example results for $n = 20$ from 10 random initializations:



5. Objectives

- approximation
- maximum likelihood estimation
- nonparametric distribution estimation
- discrimination

Residuals and approximation

consider a set of m (possibly nonlinear) equations in n variables $x \in \mathbf{R}^n$:

$$f_i(x) = b_i, \quad i = 1, \ldots, m$$

- $f_i \colon \mathbf{R}^n \to \mathbf{R}$, $b \in \mathbf{R}^m$ are given
- sometimes written as $f(x) = b$, where $f(x) = (f_1(x), \ldots, f_m(x))$

our goal is to

- find x such that $f(x) = b$, *i.e.*, solve the system, or
- find x such that $f(x) \approx b$ in the sense that the *residual*

$$r = f(x) - b$$

  is small under some metric

examples:

- when $f(x) = Ax$ for some $A \in \mathbf{R}^{m\times n}$, we have the *linear model*

$$Ax = b$$

  - residual $r = Ax - b$
  - solvable if $b \in \mathcal{R}(A)$, *i.e.*, can be expressed as a linear combination of the columns of A

- *matrix factorization* problem

$$XY = B$$

  - $f(X, Y) = XY$, $X \in \mathbf{R}^{m\times k}$, $Y \in \mathbf{R}^{k\times n}$, $B \in \mathbf{R}^{m\times n}$
  - residual $R = XY - B$
  - solvable if $k \geq \mathbf{rank}\, B$
  - reduces to linear model if either X or Y is fixed

Norm approximation

$$\text{minimize} \quad \|r\| = \|f(x) - b\|$$

- variable: $x \in \mathbf{R}^n$
- $\|\cdot\|$ is a norm on $\mathbf{R}^m$
- optimal value $\|r^\star\| = 0$ if $f(x) = b$ is solvable
- $x^\star$ called an *approximate solution* of $f(x) = b$ if $\|r^\star\| > 0$

interpretation when $f(x) = Ax$ ($A \in \mathbf{R}^{m\times n}$) is linear:

- **approximation**: find coefficient $x \in \mathbf{R}^n$ so the linear combination

$$Ax = x_1 a_1 + \cdots + x_n a_n$$

  of the columns $a_1, \ldots, a_n \in \mathbf{R}^m$ of A is as close as possible to b

- **model fitting**: express residual in the rows $\tilde{a}_i^T$ of A:

$$r_i = \tilde{a}_i^T x - b_i, \quad i = 1, \ldots, m$$

  - data points $(\tilde{a}_i, b_i)$, $i = 1, \ldots, m$
  - *feature vector*: $\tilde{a}_i \in \mathbf{R}^n$ and label $b_i \in \mathbf{R}$
  - *model prediction*: $\tilde{a}_i^T x$ (for label b_i)

  find model parameter $x \in \mathbf{R}^n$ so the prediction error $r_i = \tilde{a}_i^T x - b_i$ is small for all data points

- **geometric interpretation**:

$$\begin{array}{ll} \text{minimize} & \|z - b\| \\ \text{subject to} & z \in \mathcal{R}(A) = \{Ax \mid x \in \mathbf{R}^n\} \end{array}$$

  find point in $\mathcal{R}(A)$ that is closest to b

- **estimation**: assume *linear measurement model* $b = Ax + r$ with r as (random) noise, find x minimizes the noise r so best explains b

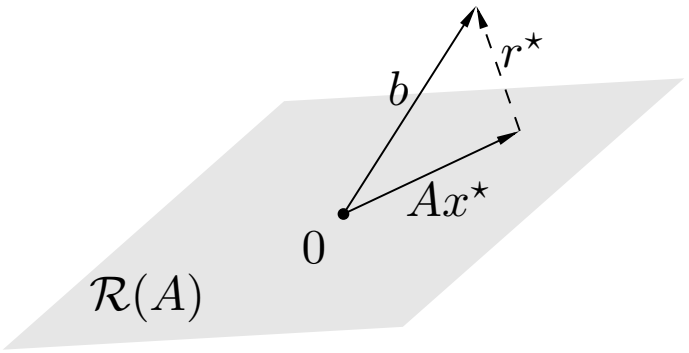
Least squares approximation

$$\text{minimize} \quad \|Ax - b\|_2^2 = r_1^2 + \cdots + r_m^2$$

- squared ℓ_2-*norm* (or *Euclidean norm*) of the residual
- rewrite objective as a convex quadratic function of x:

$$\|Ax - b\|_2^2 = x^T A^T A x - 2b^T A x + b^T b$$

- optimality condition: $\nabla \|Ax - b\|_2^2 = 2A^T A x - 2A^T b = 0$
- (unique) solution $x^\star = (A^T A)^{-1} A^T b$ (assume A tall, full rank)
- **geometric interpretation**:
  $x^\star$ is the projection of b onto $\mathcal{R}(A)$



- generalizes to nonlinear residual $f(x) - b$, but may not be convex

Least absolute residuals approximation

$$\text{minimize} \quad \|f(x) - b\|_1 = |r_1| + |r_2| + \cdots + |r_m|$$

- $\|u\|_1 = |u_1| + \cdots + |u_m|$ is the ℓ_1-*norm* of $u \in \mathbf{R}^m$
- convex optimization problem if f convex
- for linear $f(x) = Ax$ ($A \in \mathbf{R}^{m \times n}$),

$$\text{minimize} \quad \|Ax - b\|_1 = \sum_{i=1}^m |a_i^T x - b_i|$$

(a_i^T are rows of A) can be formulated as an LP:

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T t \\ \text{subject to} & -t \preceq Ax - b \preceq t \end{array}$$

(with variables $x \in \mathbf{R}^n$, $t \in \mathbf{R}^m$)

Least maximum residuals approximation

$$\text{minimize} \quad \|f(x) - b\|_\infty = \max\{|r_1|, \ldots, |r_m|\},$$

- $\|\cdot\|_\infty$: ℓ_∞-*norm* (or *Chebyshev norm*), maximum absolute value of the components
- also called *minimax* or *Chebyshev approximation*; convex if f convex
- for linear $f(x) = Ax$ ($A \in \mathbf{R}^{m\times n}$),

$$\text{minimize} \quad \|Ax - b\|_\infty = \max_{i=1,\ldots,m} |a_i^T x - b_i|$$

(a_i^T are rows of A) can be formulated as an LP:

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -t\mathbf{1} \preceq Ax - b \preceq t\mathbf{1}, \end{array}$$

(with variables $x \in \mathbf{R}^n$, $t \in \mathbf{R}$)

Penalty function approximation

$$\text{minimize} \quad \phi(r_1) + \cdots + \phi(r_m) = \sum_{i=1}^m \phi(f_i(x) - b_i)$$

- ϕ: $\mathbf{R} \to \mathbf{R}$ is a (scalar) *penalty function*
- sometimes written as

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \phi(r_i) \\ \text{subject to} & r = f(x) - b \end{array}$$

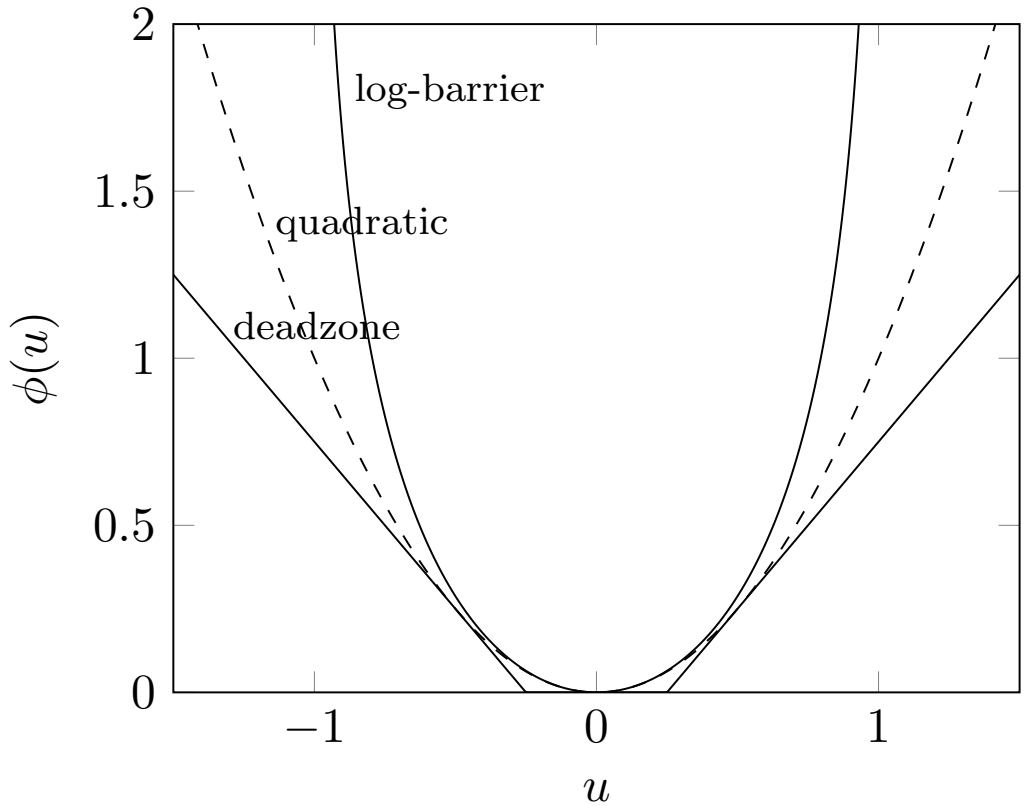
examples:

- ϕ associated with norms
  - ℓ_p-norm $\|r\|_p = \left(\sum_{i=1}^m |r_i|^p\right)^{1/p}$ $(p \in [1, \infty)) \iff \phi(r_i) = |r_i|^p$
    ($p = 1$: *absolute penalty*; $p = 2$: *quadratic penalty*)
  - ℓ_∞-norm $\|r\|_\infty \iff \phi(r_i) = \begin{cases} |r_i|, & |r_i| = \max\{|r_1|, \ldots, |r_m|\} \\ 0, & \text{otherwise.} \end{cases}$

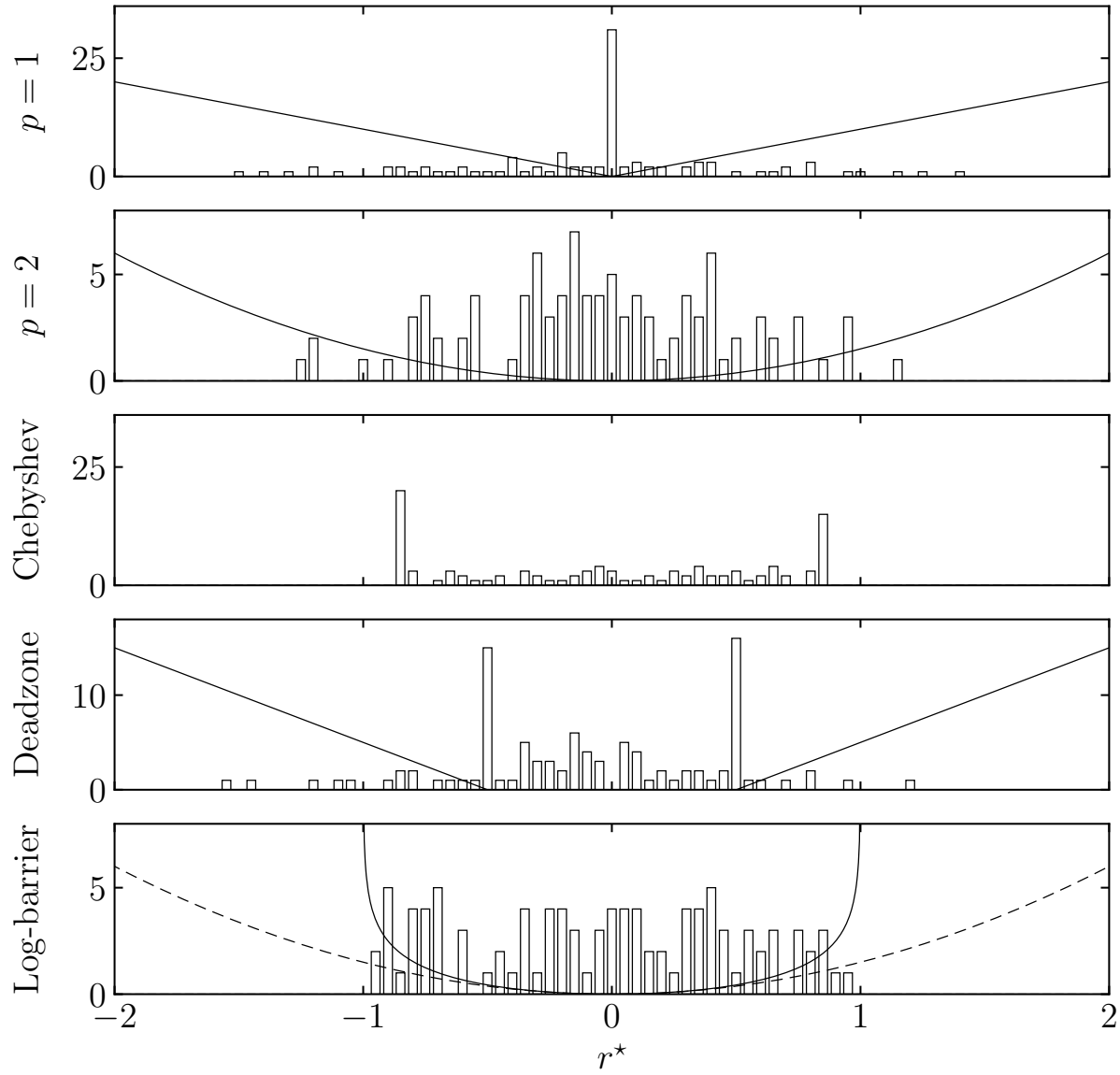
    (sometimes called *Chebyshev penalty*)

- *deadzone-linear penalty*: $\phi(u) = \begin{cases} |u| - \alpha, & |u| > \alpha \\ 0, & \text{otherwise} \end{cases}$

- *log-barrier penalty*: $\phi(u) = \begin{cases} -\alpha^2 \log(1 - (u/\alpha)^2), & |u| < \alpha \\ \infty, & \text{otherwise} \end{cases}$



numerical example: linear approximation of 100 data points in $\mathbf{R}^{30}$

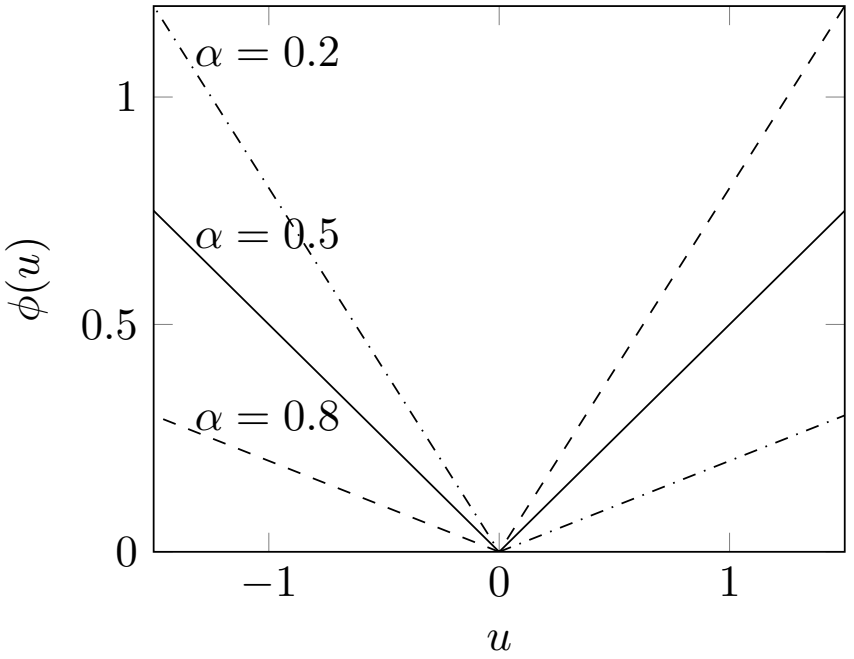


Quantile regression

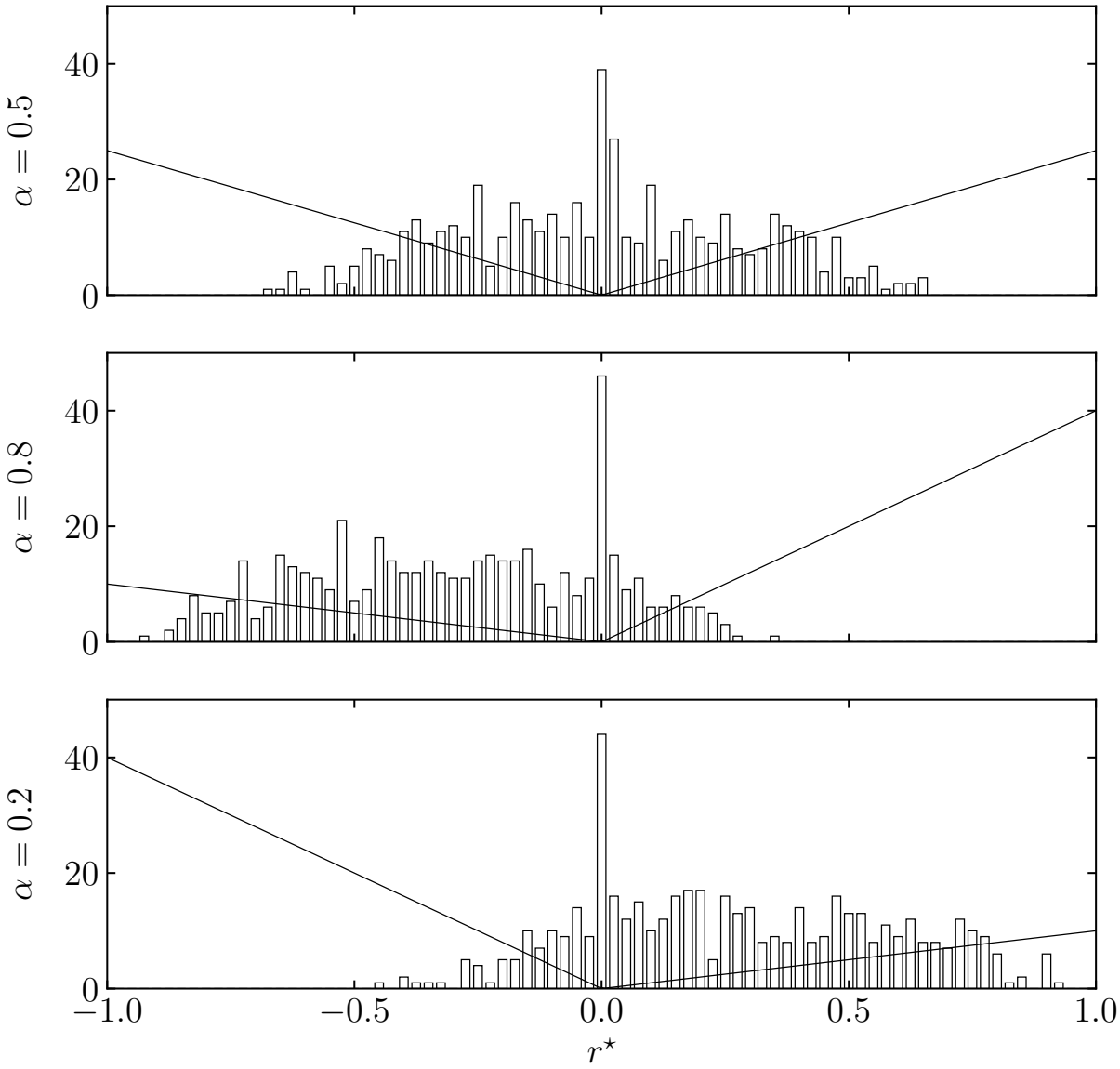
quantile penalty function:

$$\begin{aligned}\phi(u) &= \max\{\alpha u, (\alpha-1)u\} \\ &= \begin{cases} \alpha|u|, & u \geq 0 \\ (1-\alpha)|u|, & \text{otherwise} \end{cases}\end{aligned}$$

- parameter $\alpha \in (0,1)$
- $\alpha = 0.5$: absolute value penalty; otherwise tilted
- associated approximation problem called **quantile regression**



example: quantile regression of 500 data points in $\mathbf{R}^{30}$

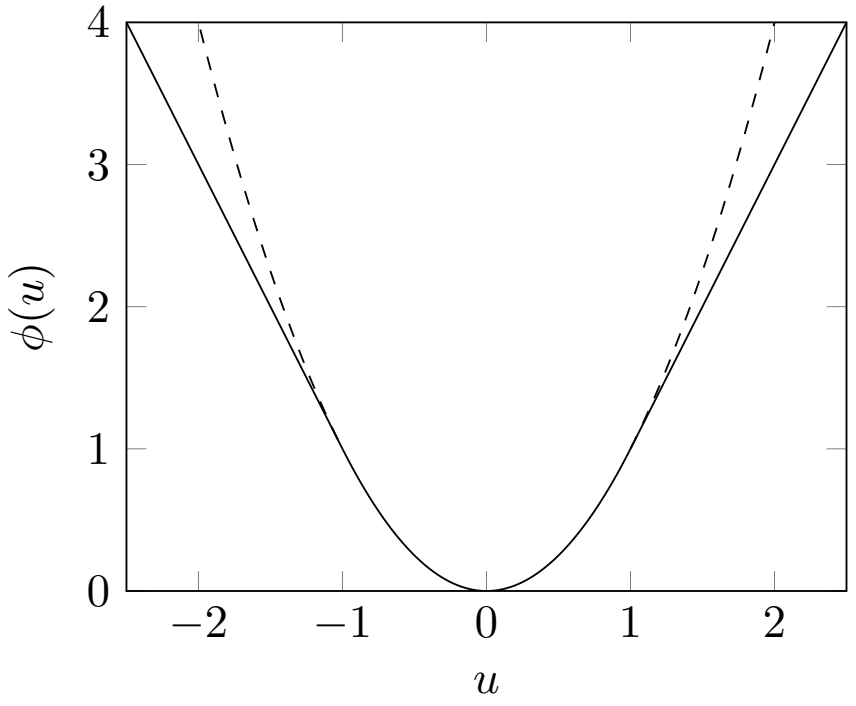


Robust regression

Huber penalty function (or *robust least squares penalty*):

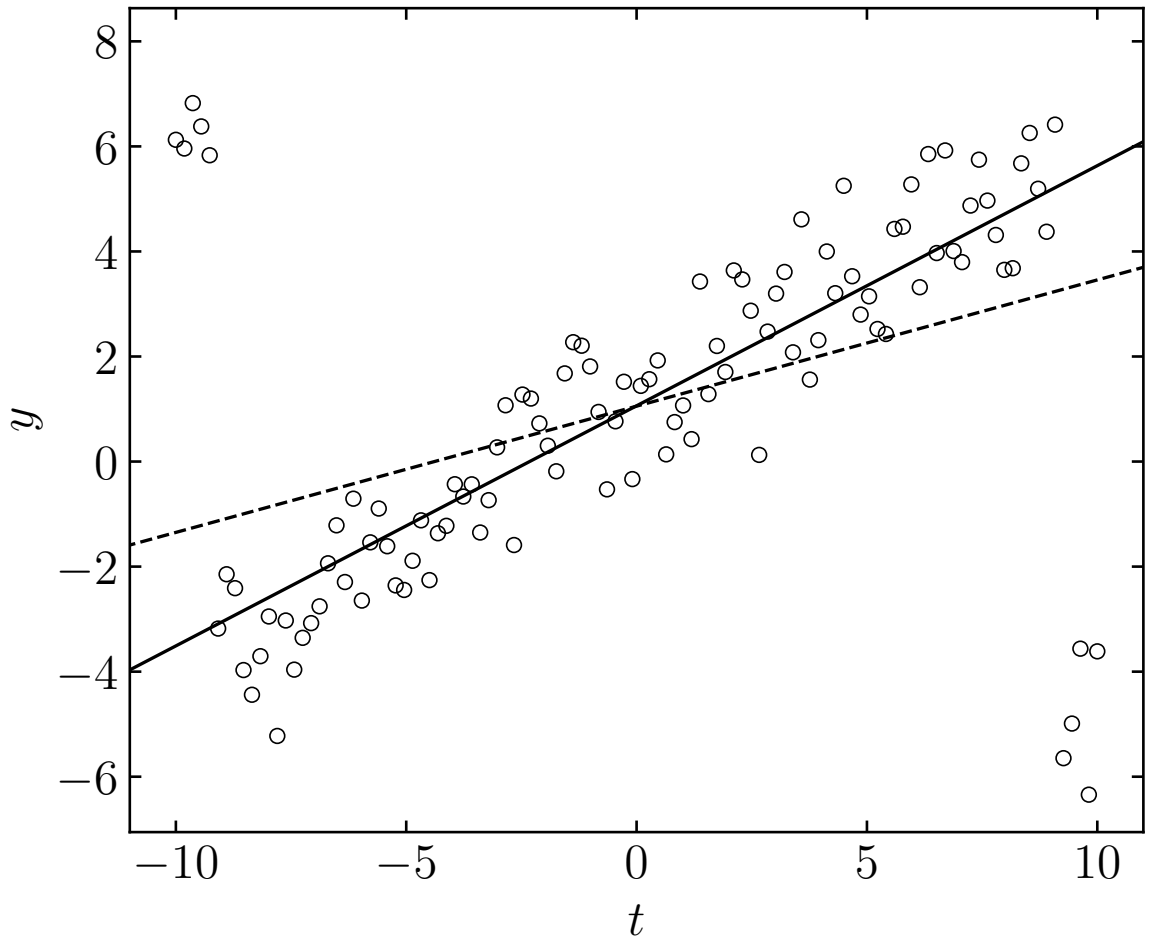
$$\phi(u) = \begin{cases} u^2, & |u| \leq M \\ M(2|u| - M), & \text{otherwise} \end{cases}$$

- parameter $M > 0$
- small residuals: quadratic penalty; large residuals: linear penalty
- hence not sensitive to outliers
- associated approximation problem called **robust regression**



example: robust regression in $\mathbf{R}^2$

- 100 points linearly distributed with random noise; 10 outliers
- dashed line shows ordinary least squares solution



Maximum likelihood estimation (MLE)

$$\text{maximize} \quad l(x) = \log p_x(y)$$

- variable $x \in \mathbf{R}^n$; data $y \in \mathbf{R}^m$ is given
- $p_x \colon \mathbf{R}^m \to \mathbf{R}_+$: density function of some random variable $y \in \mathbf{R}^m$ parameterized by $x \in \mathbf{R}^n$ (sometimes called *likelihood function*)
- $l \colon \mathbf{R}^n \to \mathbf{R}$ is the associated *log-likelihood function*

- find parameter x that maximizes the likelihood of observing the data y
- equivalent to minimizing the *negative log-likelihood* $-l(x)$
- convex program if l concave, $i.e.$, p_x log-concave in x

Linear approximation as MLE

linear measurement model: given dataset (a_i, y_i), $i = 1, \ldots, m$, assume

$$y_i = a_i^T x + v_i, \quad i = 1, \ldots, m$$

- $a_i \in \mathbf{R}^n$ are known *feature vectors*, $y_i \in \mathbf{R}$ are *responses*
- $x \in \mathbf{R}^n$ is unknown parameter, $v_i \in \mathbf{R}$ is random noise

assuming v_i are *independent and identically distributed* ($i.e.$, IID):

$$p_x(y) = \prod_{i=1}^m p(y_i - a_i^T x), \quad \text{so } l(x) = \sum_{i=1}^m \log p(y_i - a_i^T x)$$

- p: $\mathbf{R} \to \mathbf{R}_+$ is the density function of the noise v_i
- multiple possible choices of p leads to different MLE problems
- convex if p log-concave in x

Gaussian noise: $v_i \sim \mathcal{N}(0, \sigma^2)$ (with known $\sigma^2 > 0$), *i.e.*,

$$p(v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{v^2}{2\sigma^2}\right)$$

let $A = \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}^T \in \mathbf{R}^{m\times n}$,

$$\begin{aligned} l(x) &= \sum_{i=1}^{m} \log p(y_i - a_i^T x) \\ &= \sum_{i=1}^{m} \left(-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(y_i - a_i^T x)^2}{2\sigma^2}\right) \\ &= -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|Ax - y\|_2^2. \end{aligned}$$

so the MLE problem is equivalent to the least squares approximation

$$\text{maximize} \quad -\|Ax - y\|_2^2$$

Laplace noise: IID v_i following Laplace distribution (zero mean, known parameter $b > 0$), *i.e.*,

$$p(v) = \frac{1}{2b} \exp\left(-\frac{|v|}{b}\right)$$

log-likelihood function given by:

$$\begin{aligned} l(x) &= \sum_{i=1}^{m} \log p(y_i - a_i^T x) \\ &= \sum_{i=1}^{m} \left(-\log(2b) - \frac{|y_i - a_i^T x|}{b}\right) \\ &= -m \log(2b) - \frac{1}{b}\|Ax - y\|_1 \end{aligned}$$

MLE problem is equivalent to the least absolute residuals approximation

$$\text{maximize} \quad -\|Ax - y\|_1$$

uniform noise: IID $v_i \sim \mathcal{U}(-b, b)$ (with known $b > 0$), *i.e.*,

$$p(v) = \begin{cases} 1/(2b), & |v| \leq b \\ 0, & \text{otherwise} \end{cases}$$

log-likelihood function given by:

$$l(x) = \sum_{i=1}^{m} \log p(y_i - a_i^T x) = \begin{cases} -m \log(2b), & \|Ax - y\|_\infty \leq b \\ -\infty, & \text{otherwise} \end{cases}$$

any x satisfying $\|Ax - y\|_\infty \leq b$ is an MLE solution

in the most general case, any penalty function approximation problem

$$\text{minimize} \quad \textstyle\sum_{i=1}^{m} \phi(y_i - a_i^T x)$$

can be interpreted as MLE with IID noise having density function

$$p(v) = \frac{\exp(-\phi(v))}{\int \exp(-\phi(z))\ dz}$$

Probabilistic classification

logistic model: given feature $a_i \in \mathbf{R}^n$ with labels $y_i \in \{0, 1\}$, assume

$$\begin{array}{l} \mathbf{prob}(y_i = 1) = p_i \\ \mathbf{prob}(y_i = 0) = 1 - p_i \end{array}, \qquad p_i = \frac{\exp(a_i^T x)}{1 + \exp(a_i^T x)} \in (0, 1)$$

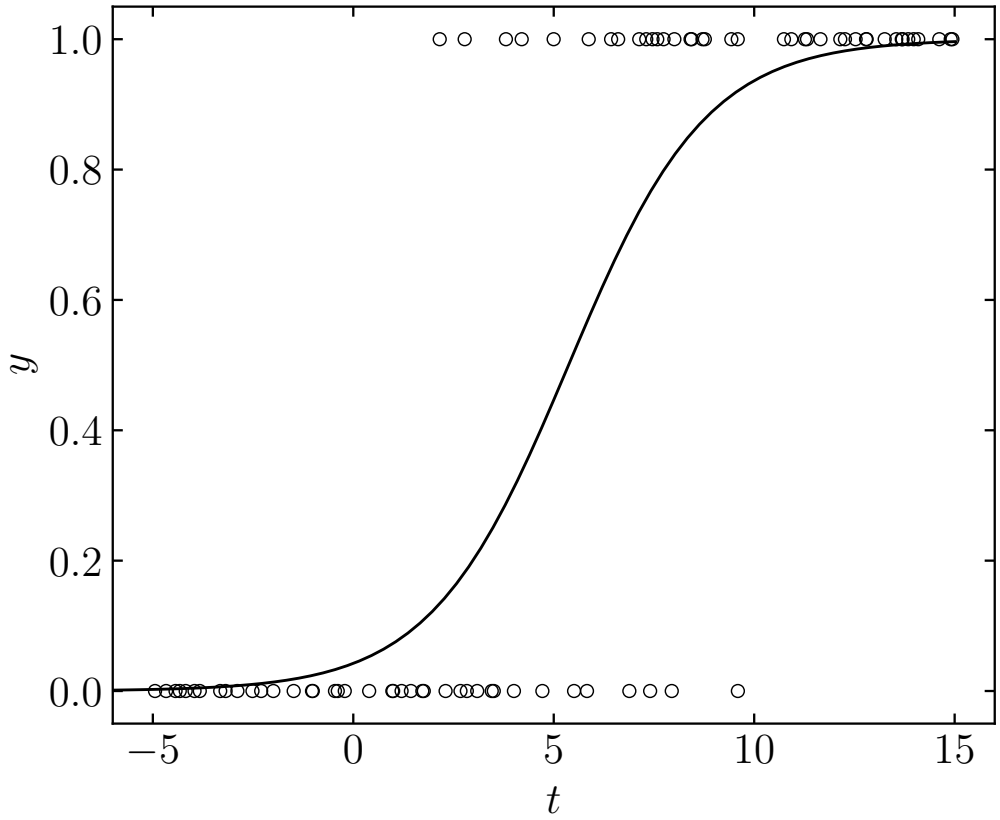
- estimate unknown parameter $x \in \mathbf{R}^n$

log-likelihood: assume $y_1, \ldots, y_k = 1$ and $y_{k+1}, \ldots, y_m = 0$,

$$\begin{aligned} l(x) &= \sum_{i=1}^{k} \log p_i + \sum_{i=k+1}^{m} \log(1 - p_i) \\ &= \sum_{i=1}^{k} \log \left(\frac{\exp(a_i^T x)}{1 + \exp(a_i^T x)} \right) + \sum_{i=k+1}^{m} \log \left(\frac{1}{1 + \exp(a_i^T x)} \right) \\ &= \sum_{i=1}^{m} \Big(y_i a_i^T x - \log(1 + \exp(a_i^T x)) \Big) \end{aligned}$$

- maximize $l(x)$ is a convex optimization problem

example: 80 data points in $\mathbf{R}^2$ with $a_i = (t_i, 1)$



solid curve shows $\mathbf{prob}(y = 1)$ for MLE solution $x^\star$

Counting problems

Poisson model: given feature $a_i \in \mathbf{R}^n$ with count label $y_i \in \mathbf{Z}_+$, assume

$$\mathbf{prob}(y_i = k) = \frac{\lambda_i^k \exp(-\lambda_i)}{k!}, \quad \log \lambda_i = a_i^T x$$

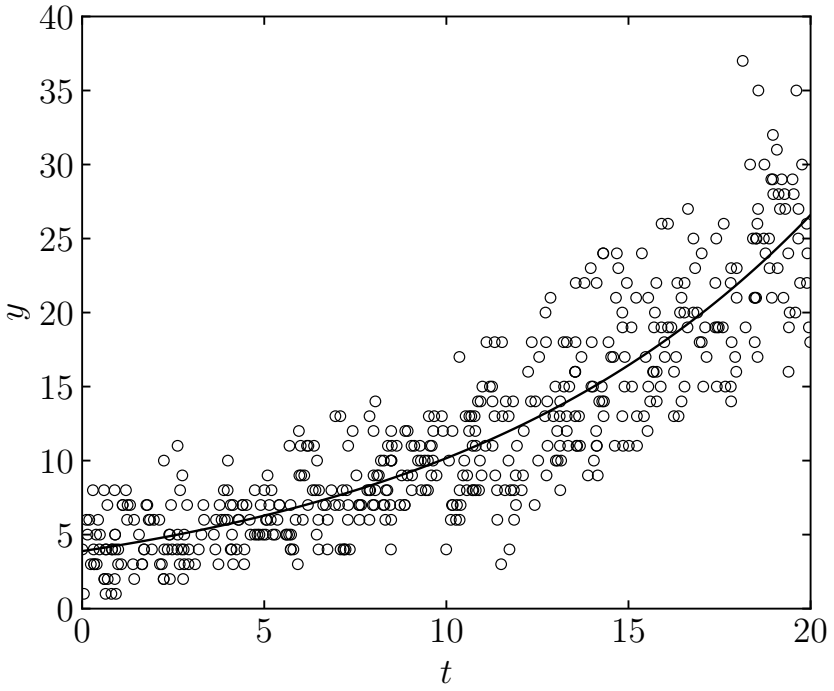
log-likelihood of parameter $x \in \mathbf{R}^n$:

$$l(x) = \log \prod_{i=1}^m \frac{\lambda_i^{y_i} \exp(-\lambda_i)}{y_i!}$$

$$= \sum_{i=1}^m (y_i \log \lambda_i - \lambda_i - \log(y_i!))$$

$$= \sum_{i=1}^m \left(y_i a_i^T x - \exp(a_i^T x) - \log(y_i!) \right)$$

- maximize $l(x)$ is convex



Gaussian covariance estimation

given Gaussian IID samples $y_1, \ldots, y_m \in \mathbf{R}^n$ for random $y \in \mathbf{R}^n$ with mean zero and covariance $X = \mathbf{E}\, yy^T \succ 0$, log-likelihood of X is

$$\begin{aligned} l(X) &= \sum_{i=1}^m \log p_X(y_i) = \sum_{i=1}^m \left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det X - \frac{1}{2} y_i^T X^{-1} y_i \right) \\ &= -\frac{mn}{2}\log(2\pi) - \frac{m}{2}\log\det X - \frac{1}{2}\mathbf{tr}\left(X^{-1} \sum_{i=1}^m y_i y_i^T \right) \end{aligned}$$

- $p_X : \mathbf{R}^n \to \mathbf{R}_+$ is Gaussian density function with covariance X

$l(X)$ nonconcave, but let $S = X^{-1}$ gives convex MLE problem:

$$\begin{array}{ll} \text{maximize} & \log\det S - \mathbf{tr}(SY) \\ \text{subject to} & S \succ 0 \end{array}$$

- $Y = (1/m)\sum_{i=1}^m y_i y_i^T \succeq 0$ is the sample covariance matrix
- if $Y \succ 0$, unique solution $S^\star = Y^{-1}$, otherwise unbounded above

Nonparametric distribution estimation

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

- variable $x \in \mathbf{R}^n$ represents a discrete distribution over n outcomes
- convex optimization problem if f convex

maximum entropy estimation:

$$f(x) = \sum_{i=1}^{n} x_i \log x_i$$

- solution $x^\star = (1/n)\mathbf{1}$ is the uniform distribution over n outcomes

minimum KL-divergence estimation:

$$f(x) = D_{\mathrm{kl}}(x, q) = \sum_{i=1}^{n} x_i \log(x_i/q_i)$$

- $q \in \mathbf{R}^n$ ($q \succeq 0$, $\mathbf{1}^T q = 1$) is a given reference distribution
- solution $x^\star = q$

maximum likelihood estimation given IID samples

$$z_i \sim x, \quad z_i \in \{1, \ldots, n\}, \quad i = 1, \ldots, m$$

from unknown distribution $x \in \mathbf{R}^n$:

$$\begin{array}{ll} \text{minimize} & -l(x) = -\sum_{i=1}^m \log x_{z_i} \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

- solution $x_i^\star = c_i/m$, where $c \in \mathbf{R}^n$ is the sample count vector

bounding problems:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n g(c_i) x_i \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

- find a lower bound on expected value of g: $\mathbf{R} \to \mathbf{R}$ over $\{c_1, \ldots, c_n\}$ under distribution x
- similar for lower bounding, *e.g.*, probability of events, sum of k largest probabilities, . . .

Discrimination

given $\{x_1, \ldots, x_M\}, \{y_1, \ldots, y_N\} \subseteq \mathbf{R}^n$, find $f : \mathbf{R}^n \to \mathbf{R}$ such that

$$f(x_i) > 0 \text{ for all } i = 1, \ldots, M, \qquad f(y_i) < 0 \text{ for all } i = 1, \ldots, N$$

i.e., the 0-sublevel set of f separates the two sets

linear discrimination:

$$\begin{array}{ll} \text{find} & (a, b) \\ \text{subject to} & a^T x_i - b > 0, \quad i = 1, \ldots, M \\ & a^T y_i - b < 0, \quad i = 1, \ldots, N \end{array}$$

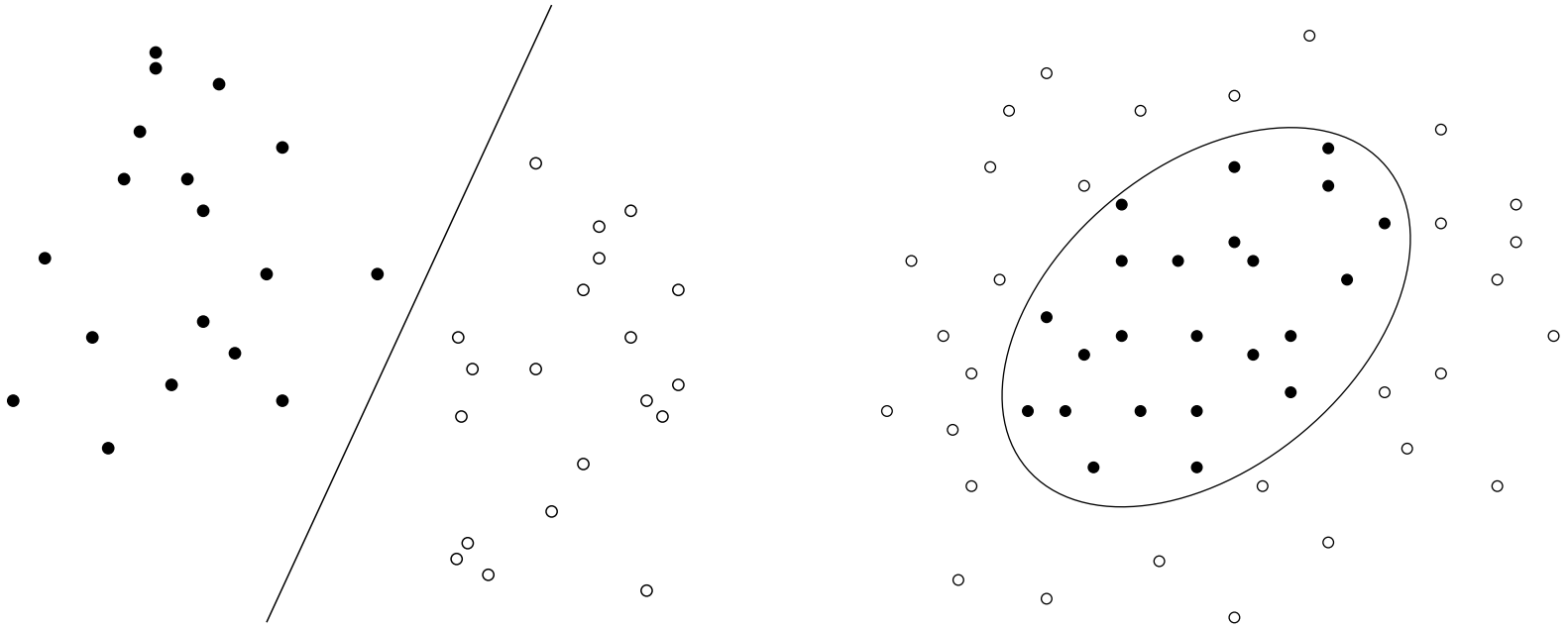
- constraints homogeneous in (a, b), so equivalent to

$$\begin{array}{ll} \text{find} & (a, b) \\ \text{subject to} & a^T x_i - b \geq 1, \quad i = 1, \ldots, M \\ & a^T y_i - b \leq -1, \quad i = 1, \ldots, N \end{array}$$

quadratic discrimination:

$$\begin{array}{ll} \text{find} & (P, q, r) \\ \text{subject to} & x_i^T P x_i + q^T x_i + r \geq 1, \quad i = 1, \ldots, M \\ & y_i^T P y_i + q^T y_i + r \leq -1, \quad i = 1, \ldots, N \end{array}$$

- LP of variables $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, $r \in \mathbf{R}$
- $P \succ 0$ gives an ellipsoidal separator



6. Regularization functions

- multiobjective optimization problems
- optimal and Pareto optimal
- salarization and trade-off analysis
- regularized approximation
- Tikhonov and sparsity regularization
- smoothing
- maximum a posteriori estimation
- matrix regularizers

Multiobjective optimization

$$\begin{array}{ll} \text{minimize} & f_0(x) = (F_1(x), \ldots, F_k(x)) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- $f_0 \colon \mathbf{R}^n \to \mathbf{R}^k$ is vector-valued objective; ith component $F_i \colon \mathbf{R}^n \to \mathbf{R}$
- also called *vector optimization* or *multicriteria optimization*

convex multiobjective problem if

- for all $x, y \in \mathbf{dom}\, f_0$ and $\theta \in [0, 1]$,

$$f_0(\theta x + (1-\theta)y) \preceq \theta f_0(x) + (1-\theta) f_0(y)$$

  i.e., F_i convex for all $i = 1, \ldots, k$

- f_i convex for $i = 1, \ldots, m$ and h_i affine for $i = 1, \ldots, p$

Interpretations

scalar optimization ($k = 1$):

- $f_0(x) \leq f_0(y)$: x no worse than y
- $f_0(x) \geq f_0(y)$: x no better than y
- $f_0(x) = f_0(y)$: x and y are equally good

find $x \in \mathbf{dom}\, f_0$ such that $f_0(x) \leq f_0(y)$ for all $y \in \mathbf{dom}\, f_0$

vector optimization:

- $f_0(x) \preceq f_0(y)$: x no worse than y in all objectives F_i
- $f_0(x) \succeq f_0(y)$: x no better than y in all objectives F_i
- $f_0(x) = f_0(y)$: x and y are equally good in all objectives F_i
- $f_0(x) \not\preceq f_0(y)$ and $f_0(x) \not\succeq f_0(y)$: x better than y in some objectives F_i and worse in others, *i.e.*, x and y are *incomparable*

'minimize' is ambiguous in the last case (which is, sadly, mostly the case)

Optimal value and points

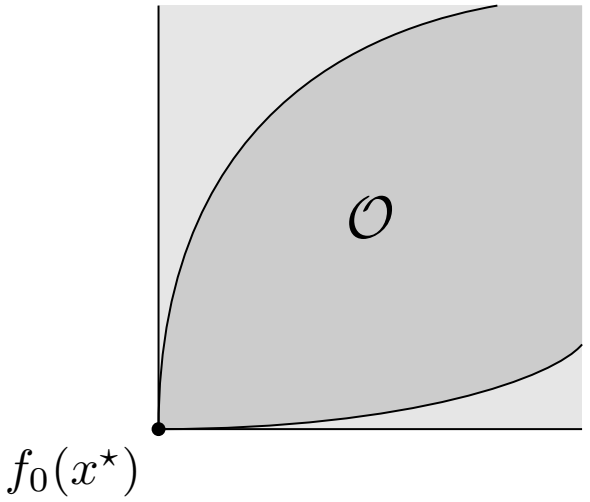
feasible point $x^\star$ is **optimal** if for *all* feasible x

$$f_0(x^\star) \preceq f_0(x) \quad \iff \quad F_i(x^\star) \leq F_i(x), \quad i = 1, \ldots, k$$

- $f_0(x^\star)$ is called **optimal value**
- implicit requirement: $f_0(x^\star)$ is comparable to $f_0(x)$ for all feasible x

set $\mathcal{O} \subseteq \mathbf{R}^k$ of **achievable objective values**:

$$\mathcal{O} = \left\{ f_0(x) \;\middle|\; \begin{array}{ll} x \in \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i \\ f_i(x) \leq 0, & i = 1, \ldots, m \\ h_i(x) = 0, & i = 1, \ldots, p \end{array} \right\}$$



so $x^\star$ optimal $\iff$ $\mathcal{O} \subseteq f_0(x^\star) + \mathbf{R}^k_+$

most multiobjective problems do not have an optimal point

Pareto optimal values and points

feasible point $x^{\rm po}$ is **Pareto optimal** or *efficient* if for all feasible x

$$f_0(x) \preceq f_0(x^{\rm po}) \quad \Longrightarrow \quad f_0(x) = f_0(x^{\rm po})$$

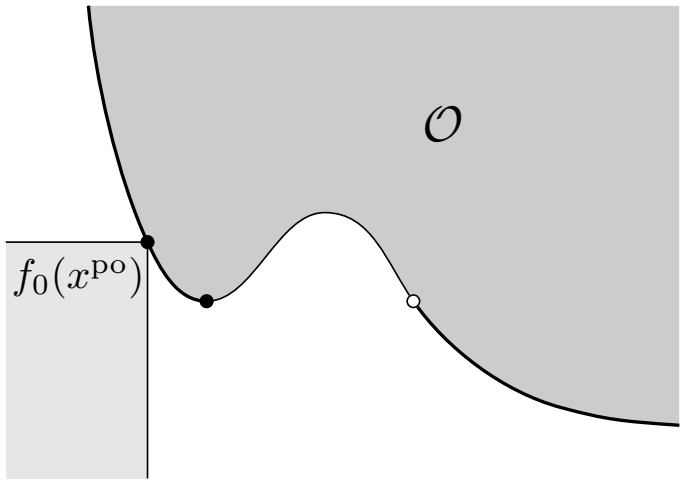
i.e., $x^{\rm po}$ is the best among all comparable feasible points

- $f_0(x^{\rm po})$ is called **Pareto optimal value**
- usually many Pareto optimal points
- feasible $x^{\rm po}$ Pareto optimal $\iff$

$$\mathcal{O} \cap (f_0(x^{\rm po}) - \mathbf{R}^k_+) = \{f_0(x^{\rm po})\}$$

- **Pareto front** is the set of values

$$\{f_0(x^{\rm po}) \mid x^{\rm po} \text{ Pareto optimal}\}$$



Scalarization

consider the (possibly nonconvex) scalar optimization problem

$$\begin{array}{ll} \text{minimize} & \lambda^T f_0(x) = \sum_{i=1}^k \lambda_i F_i(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- variable $x \in \mathbf{R}^n$; vector-valued objective $f_0 \colon \mathbf{R}^n \to \mathbf{R}^k$
- parameter $\lambda \in \mathbf{R}^k_{++}$ is a given **weight vector**
- optimal point is Pareto optimal for the original multiobjective problem
  - to show this, suppose optimal x is not Pareto optimal, *i.e.*, for some y

$$f_0(y) \preceq f_0(x), \ f_0(y) \neq f_0(x) \implies \lambda^T (f_0(x) - f_0(y)) > 0$$

then $\lambda^T f_0(y) < \lambda^T f_0(x)$, contradicting optimality of x

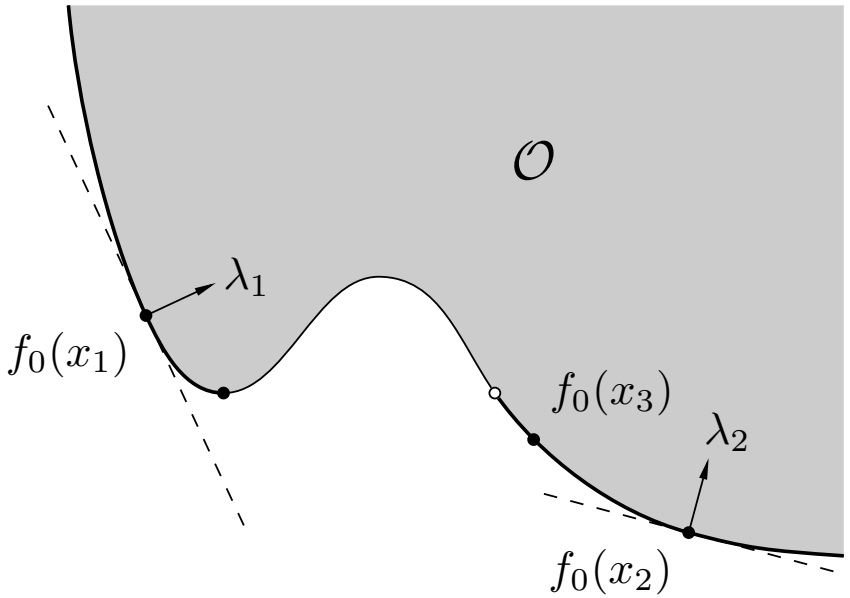
- varying $\lambda \succ 0$ gives (possibly) different Pareto optimal points
- but not all Pareto optimal points can be obtained this way

geometric interpretation:

$$x \text{ optimal} \iff \lambda^T(z - f_0(x)) \geq 0 \text{ for all } z \in \mathcal{O}$$

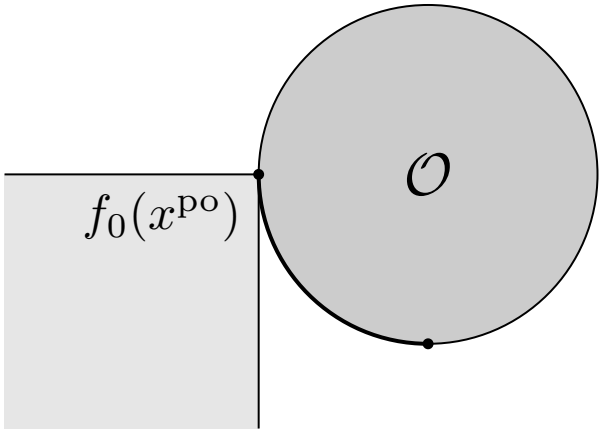
i.e., hyperplane $\{z \mid \lambda^T(z - f_0(x)) = 0\}$ supports $\mathcal{O}$ at $f_0(x)$

- $f_0(x_1)$, $f_0(x_2)$, $f_0(x_3)$ are Pareto optimal values
- $f_0(x_1)$ and $f_0(x_2)$ can be obtained by scalarization with $\lambda \succ 0$
- there is no supporting hyperplane of $\mathcal{O}$ at $f_0(x_3)$

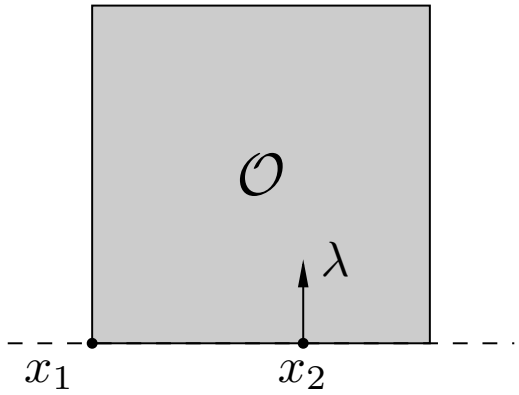


scalarization of convex problems:

- convex multiobjective problem has convex scalarization for any $\lambda \succeq 0$
- a partial converse holds: *all* Pareto optimal points can be obtained by scalarization with $\lambda \succeq 0$ $(\lambda \neq 0)$
- *not* all optimal points x from scalarization $\lambda \succeq 0$ are Pareto optimal



Pareto optimal value from $\lambda \succeq 0$
(but $\lambda \not\succeq 0$)



x_2 optimal from $\lambda \succeq 0$
but not Pareto optimal

- in practice: find all optimal x with $\lambda \succeq 0$, and then check Pareto optimality for $\lambda \not\succeq 0$

Trade-off analysis

consider biobjective problem with no optimal

$$\begin{array}{ll} \text{minimize} & (F_1(x), F_2(x)) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{array}$$

- to achieve $F_1(x) - \delta$, how much worse do we have to accept in F_2?
- if we accept worsen $F_1(x) + \delta$, how much can we improve F_2?

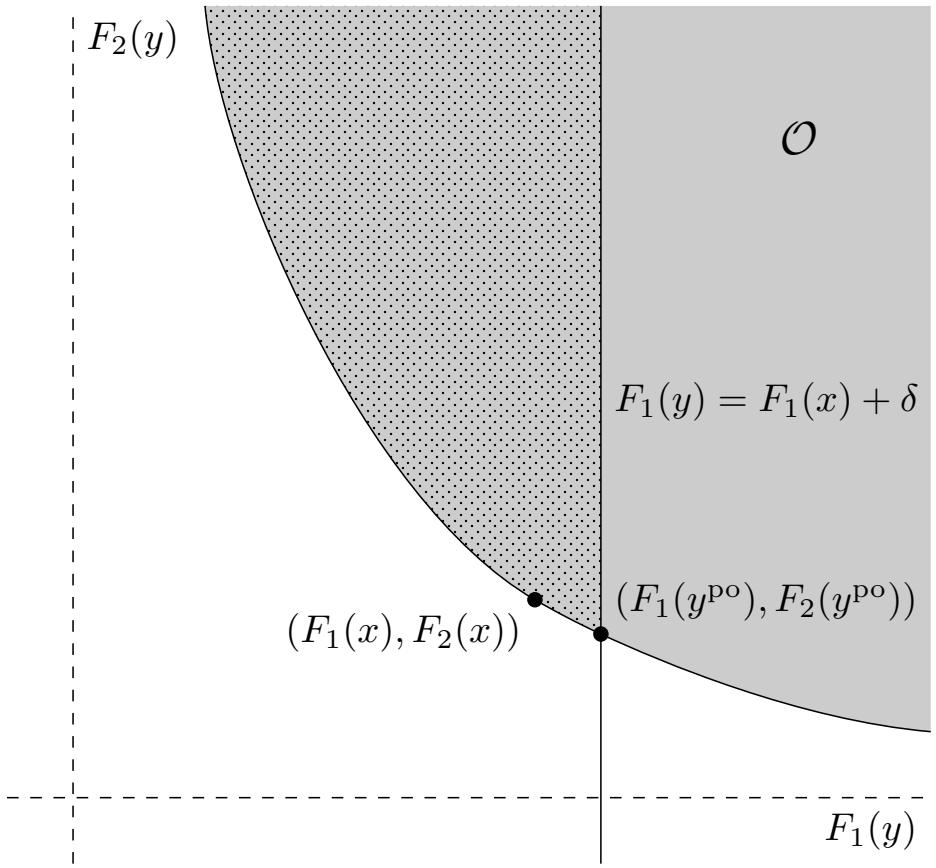
to answer these questions, we solve

$$\begin{array}{ll} \text{minimize} & F_2(y) \\ \text{subject to} & F_1(y) \leq F_1(x) + \delta \\ & f_i(y) \leq 0, \quad i = 1, \ldots, m \\ & h_i(y) = 0, \quad i = 1, \ldots, p \end{array}$$

- $y \in \mathbf{R}^n$ is the variable; $F_1(x), \delta \in \mathbf{R}$ are given

geometric interpretation: trade-off analysis problem equivalent to

$$\begin{array}{ll} \text{minimize} & F_2(y) \\ \text{subject to} & F_1(y) \leq F_1(x) + \delta \\ & (F_1(y), F_2(y)) \in \mathcal{O} \end{array}$$

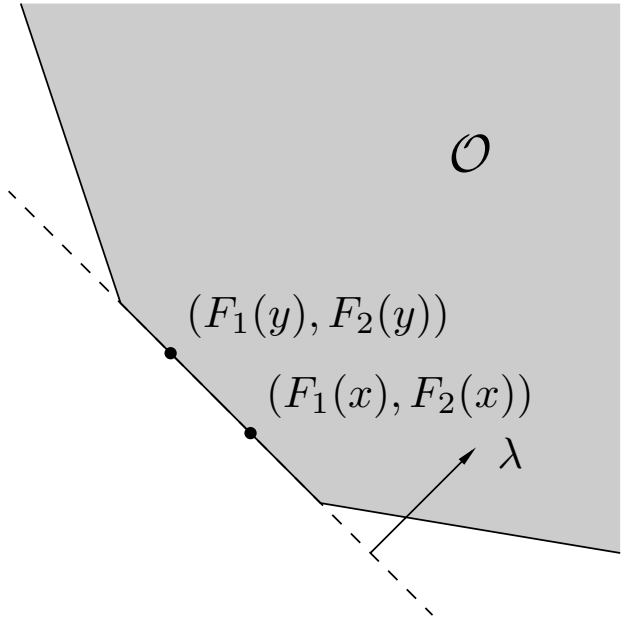


scalarization interpretation: for $\lambda^T f_0(x) = \sum_{i=1}^k \lambda_i F_i(x)$

- $\lambda_i > 0$ is the *importance*, *concern*, or *irritation* of objective F_i
- so adjusting $\lambda \succ 0$ we are trading-off between the objectives
- ratio λ_i/λ_j is the *relative importance* or *exchange rate* of F_i to F_j

example: Pareto optimal x, y with

$$\lambda_1 F_1(x) + \lambda_2 F_2(x) = \lambda_1 F_1(y) + \lambda_2 F_2(y)$$



suppose $\delta = F_1(x) - F_1(y) > 0$, then

$$F_2(y) = F_2(x) + (\lambda_1/\lambda_2)\delta$$

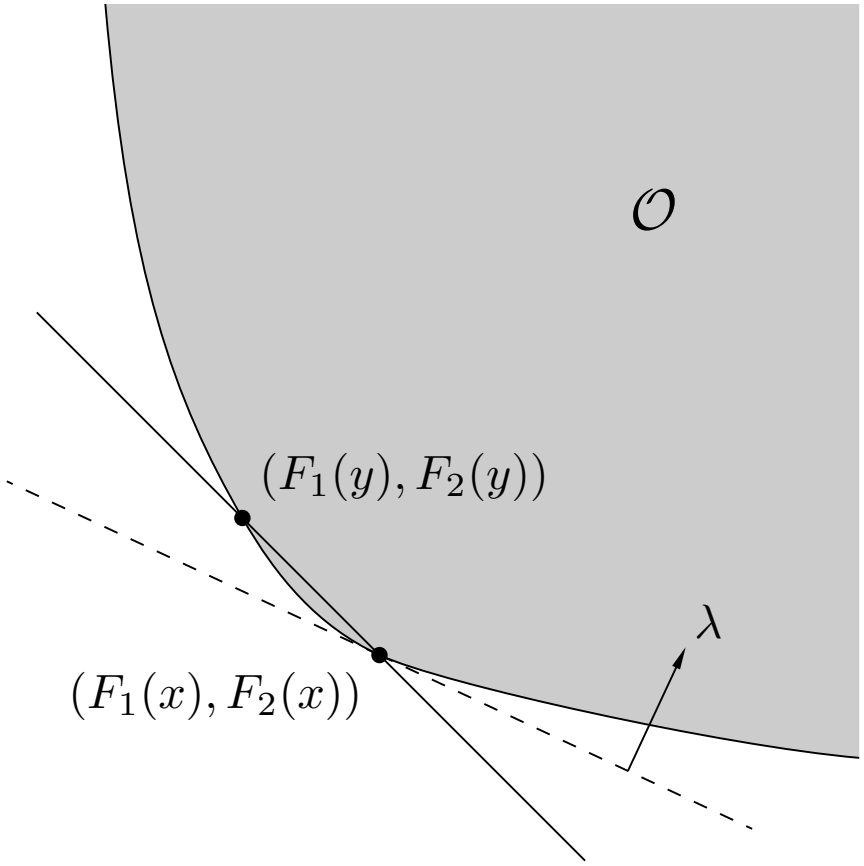
in other words,

$$\frac{F_2(y) - F_2(x)}{F_1(y) - F_1(x)} = -\frac{\lambda_1}{\lambda_2}$$

in the most general case, as $\delta = F_1(x) - F_1(y) \to 0$, we have

$$\lim_{\delta \to 0} \frac{F_2(y) - F_2(x)}{F_1(y) - F_1(x)} = -\frac{\lambda_1}{\lambda_2}$$

so λ gives the slope of the supporting hyperplane of $\mathcal{O}$ at $(F_1(x), F_2(x))$



Regularized approximation

$$\text{minimize} \quad (\phi_1(f(x) - b),\ \phi_2(x))$$

- $x \in \mathbf{R}^n$ is the variable; $f(x) - b \in \mathbf{R}^m$ is the residual
- $\phi_1 : \mathbf{R}^m \to \mathbf{R}$: residual penalty function (*primary objective*)
- $\phi_2 : \mathbf{R}^n \to \mathbf{R}$: regularization function (*secondary objective*)

relative scalarization:

$$\text{minimize} \quad \phi_1(f(x) - b) + \gamma\phi_2(x)$$

- $\gamma > 0$ is the *regularization parameter*
- varying γ gives different Pareto optimal points
- two extreme cases:
  - $\gamma \to 0$: ignore regularization, minimize residual
  - $\gamma \to \infty$: ignore residual, minimize regularization

Tikhonov regularization

$$\text{minimize} \quad (\phi_1(f(x) - b),\ \|x\|_2^2)$$

- $\|x\|_2^2$ called *Tikhonov regularization*; induce small solution x

example: Tikhonov regularization least squares

$$\text{minimize} \quad (\|Ax - b\|_2^2,\ \|x\|_2^2)$$

scalarize with $\lambda \succ 0$:

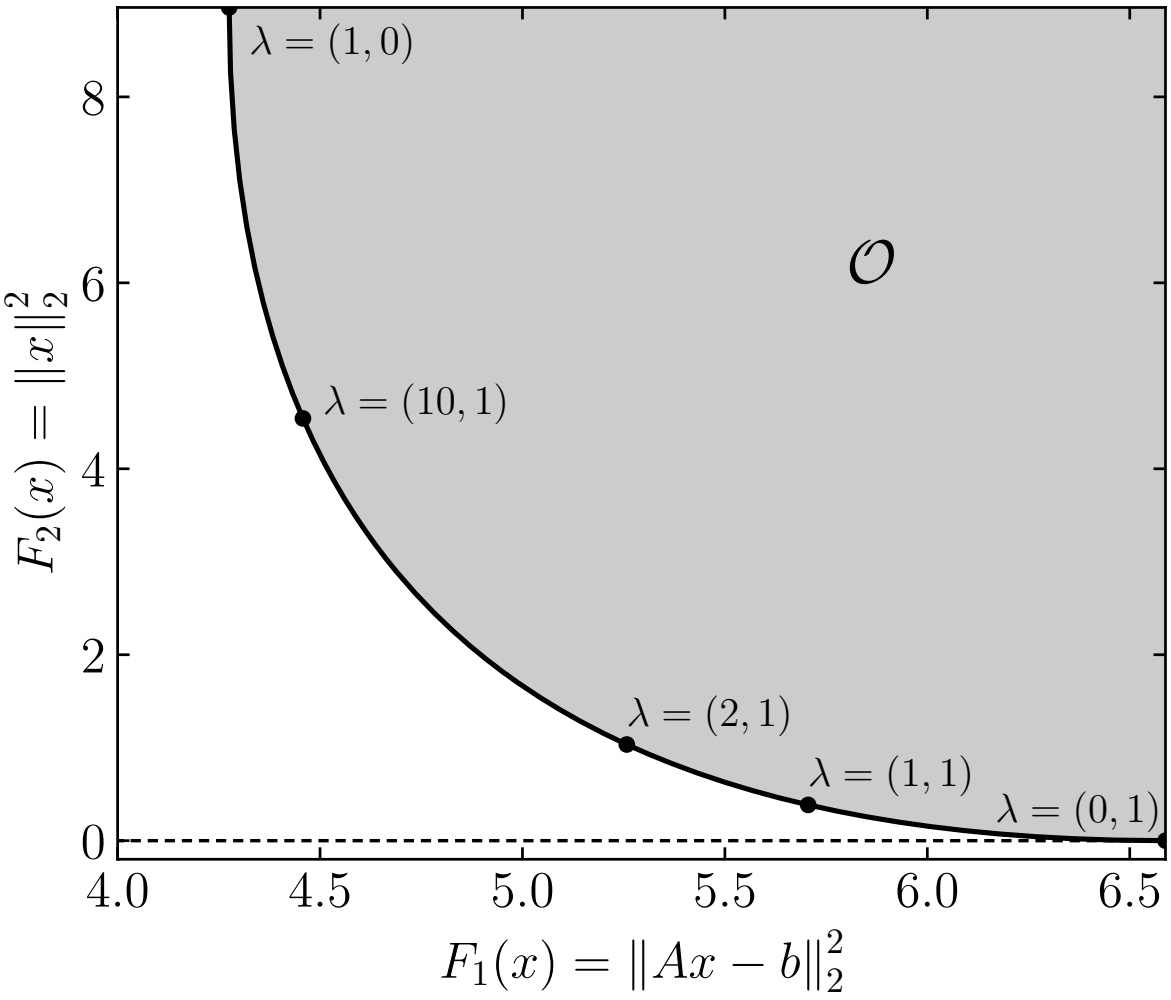
$$\lambda_1\|Ax - b\|_2^2 + \lambda_2\|x\|_2^2 = x^T(\lambda_1 A^T A + \lambda_2 I)x - 2\lambda_1 b^T Ax + \lambda_1 b^T b$$

is a convex quadratic in x, so the optimal point is given by

$$x = (A^T A + \gamma I)^{-1} A^T b, \quad \gamma = \lambda_2/\lambda_1$$

- removes rank requirement on A for least squares
- sometimes called *ridge regression*

- numerical results for $A \in \mathbf{R}^{100\times 10}$:



Sparsity regularization

$$\text{minimize} \quad (\phi_1(f(x) - b),\ \mathbf{card}\, x)$$

- $\mathbf{card}\, x = |\{i \mid x_i \neq 0,\ i = 1, \ldots, n\}|$: number of nonzero components
- induces sparse solution x, *i.e.*, many components of x are zero
- to find the Pareto front, solve for each $\mathbf{card}\, x = 0, 1, \ldots, n$:
  - if n is small, can enumerate all possible sparsity patterns of x
  - very difficult for large n

ℓ_1-norm regularization:

$$\text{minimize} \quad (\phi_1(f(x) - b),\ \|x\|_1)$$

- very useful convex approximation of $\mathbf{card}\, x$

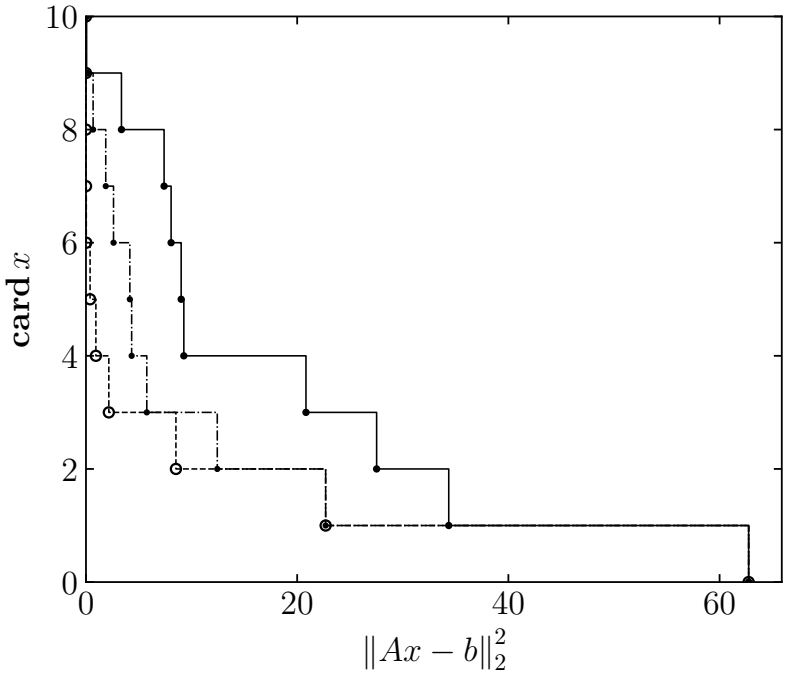
example: ℓ_1-norm regularized least squares, scalarized with $\gamma > 0$:

$$\text{minimize} \quad \|Ax - b\|_2^2 + \gamma \|x\|_1$$

- larger γ gives sparser solution x
- solution with $\mathbf{card}\, x = k$ from smallest γ approximates Pareto optimal
- **polishing**: fix sparsity pattern of x according to ℓ_1-norm heuristic solution and minimize $\|Ax - b\|_2^2$

numerical example for $A \in \mathbf{R}^{10\times 20}$:

- dashes: true Pareto optimal
- solid line: ℓ_1-norm approximation
- dashdots: ℓ_1-norm approximation and polishing



Smoothing

find estimation $\hat{x}$ of true signal $x \in \mathbf{R}^n$ from corrupted observation

$$x_{\rm obs} = x + v, \quad v \in \mathbf{R}^n \text{ is noise}$$

assumption: x is smooth, *i.e.*, $x_i \approx x_{i+1}$ for most $i = 1, \ldots, n-1$

regularized approximation formulation:

$$\text{minimize} \quad (\|\hat{x} - x_{\rm obs}\|_2^2, \ \phi(\hat{x}))$$

- **quadratic smoothing**: $\phi(\hat{x}) = \sum_{i=1}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2$

  assumes x does not change too much from i to $i+1$

- **total variation smoothing**: $\phi(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$

  assumes x is piecewise constant, *i.e.*, only a few changes for all i

scalarized smoothing problem with $\gamma > 0$:

$$\text{minimize} \quad \|\hat{x} - x_{\text{obs}}\|_2^2 + \gamma \|D\hat{x}\|$$

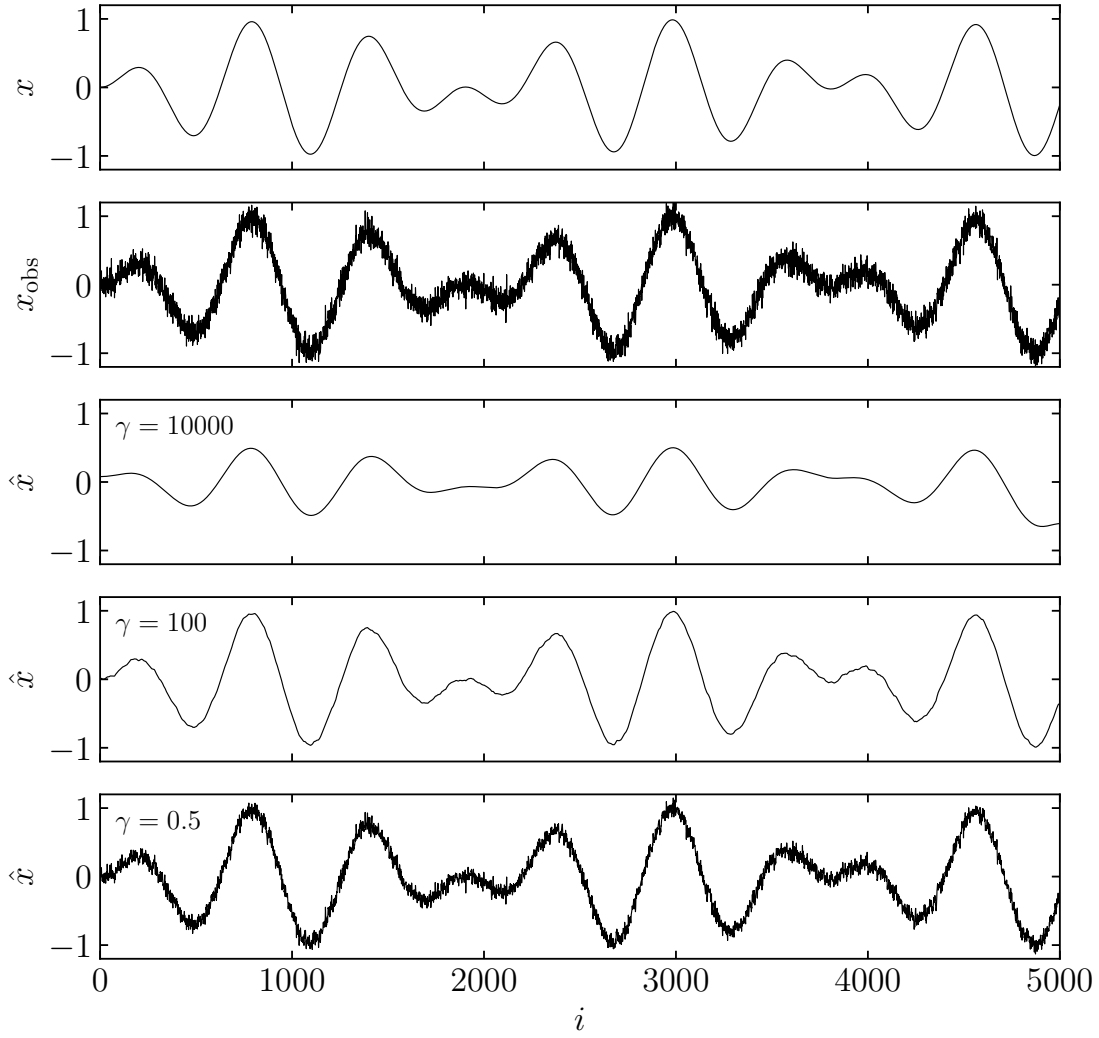
where $D \in \mathbf{R}^{(n-1)\times n}$ is the *first-order difference matrix*, given by

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

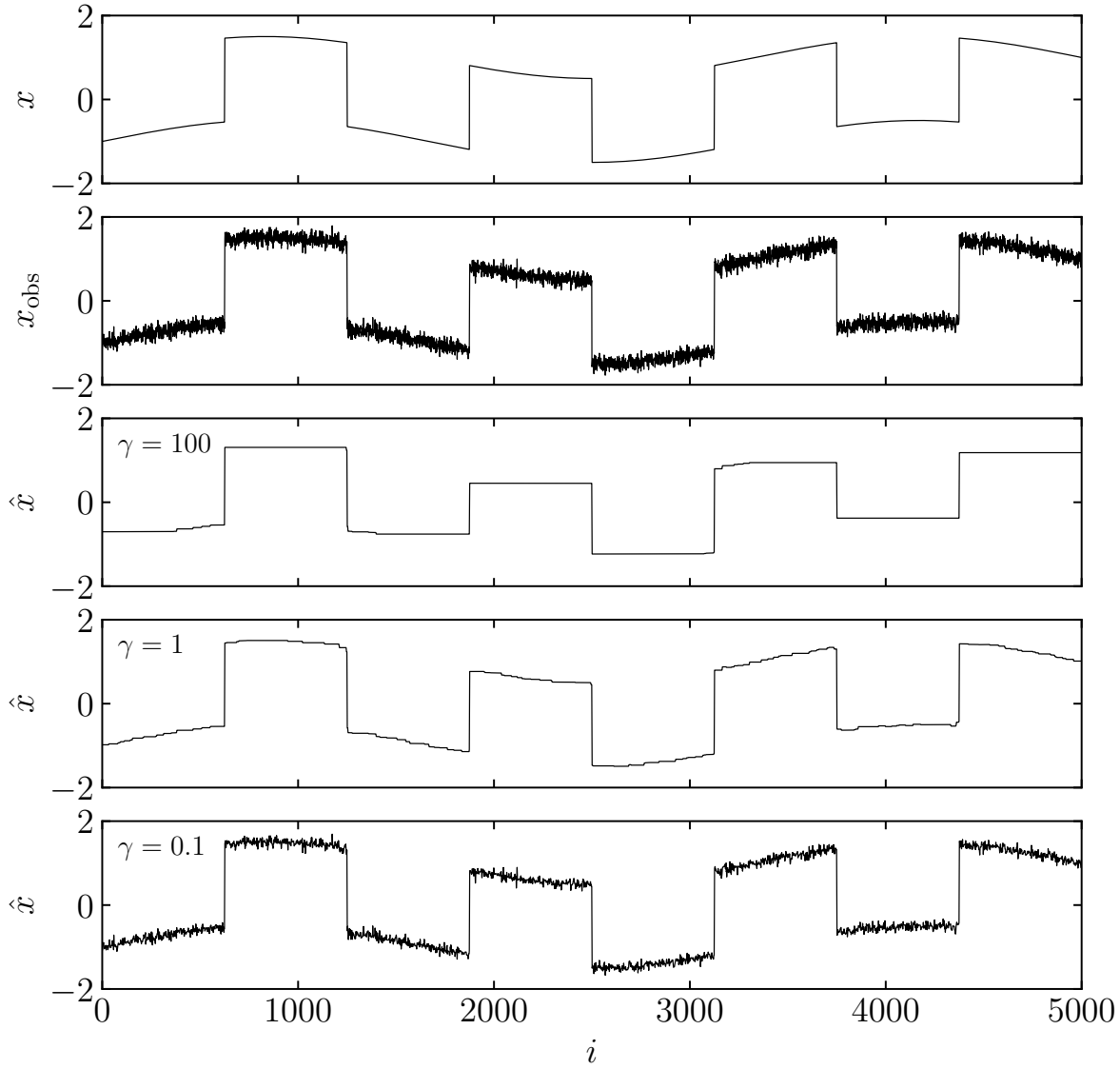
- $\|\cdot\|_1$ gives total variation smoothing
- $\|\cdot\|_2$ gives quadratic smoothing

examples:

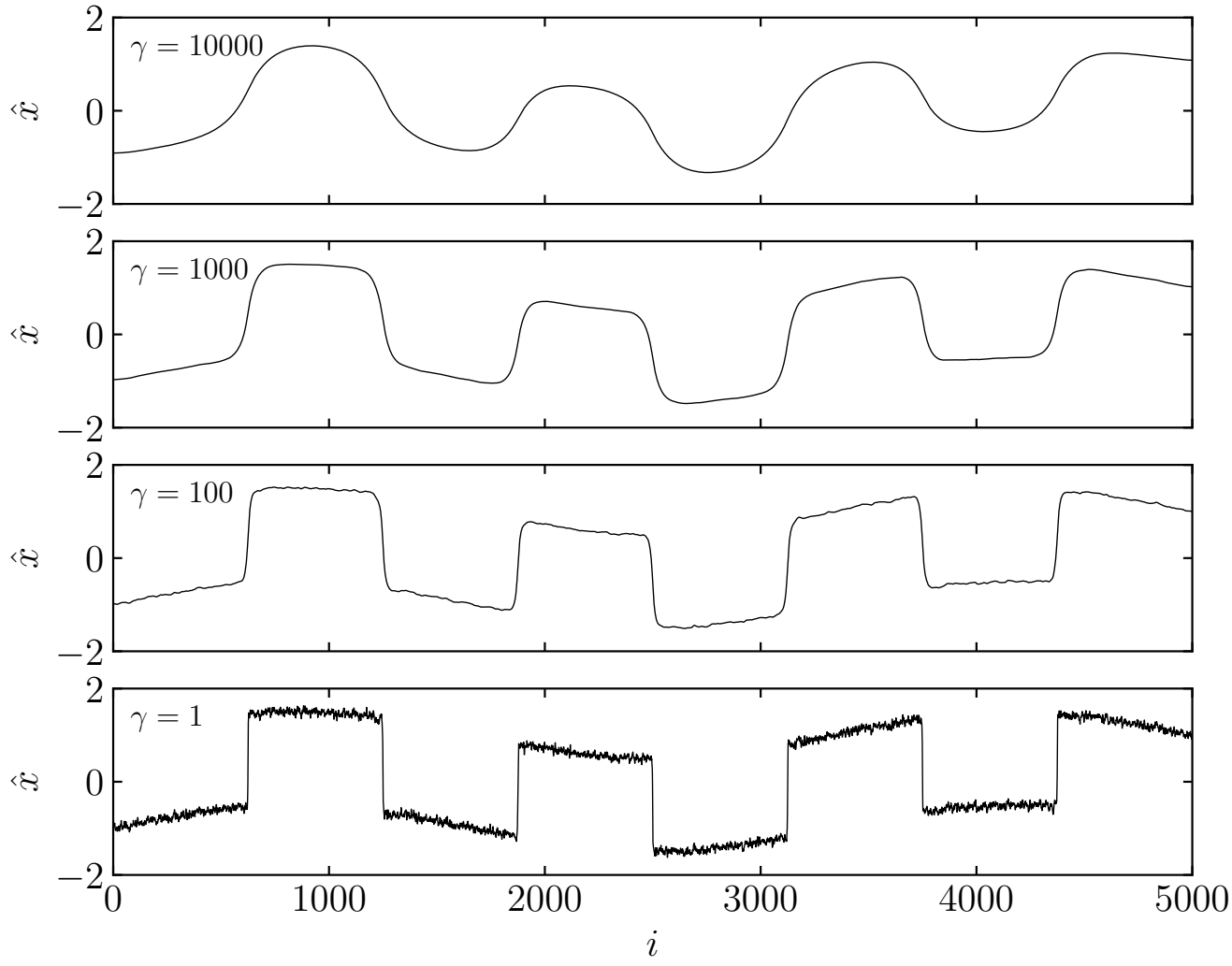
- quadratic smoothing for $x \in \mathbf{R}^{5000}$:



- total variation smoothing for $x \in \mathbf{R}^{5000}$:



- previous example under quadratic smoothing:



Maximum a posteriori (MAP) estimation

$$\text{maximize} \quad \log p_{y|x}(x, y) + \log p_x(x)$$

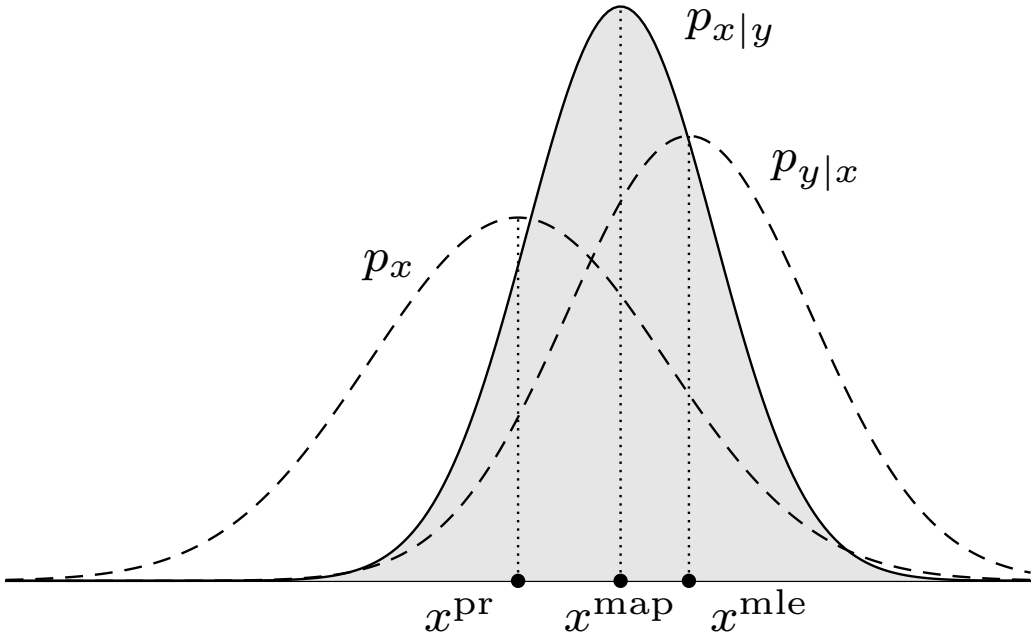
- variable $x \in \mathbf{R}^n$; observation $y \in \mathbf{R}^m$ is given
- $p_{y|x}$: $\mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}_+$: likelihood of y given x
- p_x: $\mathbf{R}^n \to \mathbf{R}_+$: prior probability of x
- according to Bayes' theorem: *posterior probability* of x given y is

$$p_{x|y}(x, y) = \frac{p_{y|x}(x, y)p_x(x)}{p_y(y)}$$

find x that maximizes posterior probability $p_{x|y}(x, y)$, *i.e.*, the most likely x given observation y and prior knowledge of x

(contrast with MLE, which only maximizes likelihood $p_{y|x}(x, y)$)

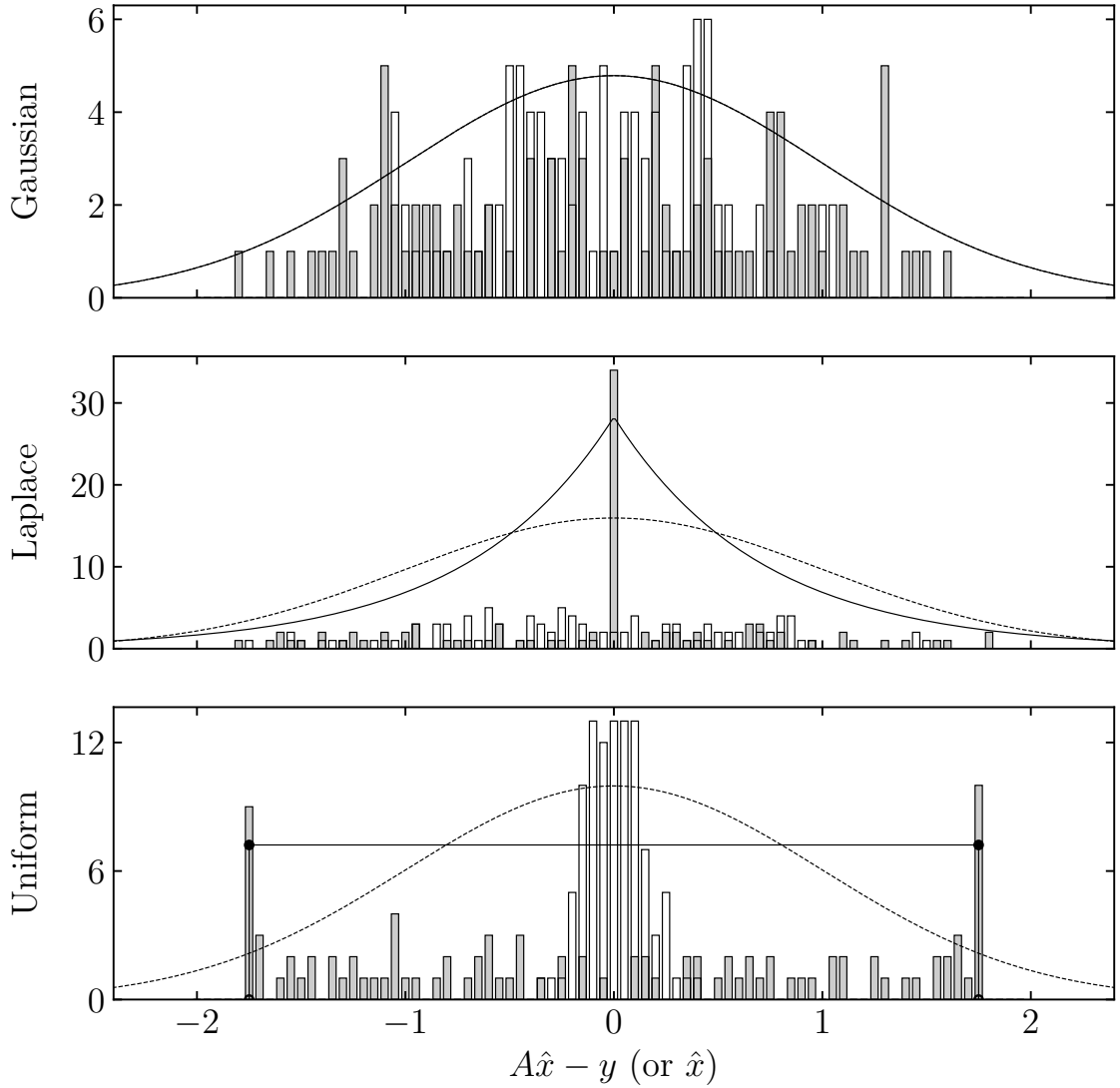
Bayesian interpretation:



regularization interpretation:

- MAP estimation is MLE with some kind of regularization function
- any regularized approximation can be interpreted as an MAP estimation by properly choosing the prior and likelihood densities

example: linear model $y = Ax + v$ ($A \in \mathbf{R}^{100\times 100}$) with IID Gaussian noise v, under Gaussian, Laplace, and uniform priors for x



(residual $A\hat{x} - y$ shown white; estimation $\hat{x}$ shown shaded)

Trade-off in MAP estimation

for linear model $y = Ax + v$ with IID noise $v_i \sim \mathcal{N}(0, 1)$, log-likelihood is

$$\log p_{y|x}(x, y) = -\frac{m}{2}\log(2\pi) - \frac{1}{2}\|Ax - y\|_2^2$$

suppose IID Gaussian prior $x \sim \mathcal{N}(0, \sigma_x^2 I)$, then

$$\log p_x(x) = -\frac{n}{2}\log(2\pi\sigma_x^2) - \frac{1}{2\sigma_x^2}\|x\|_2^2$$

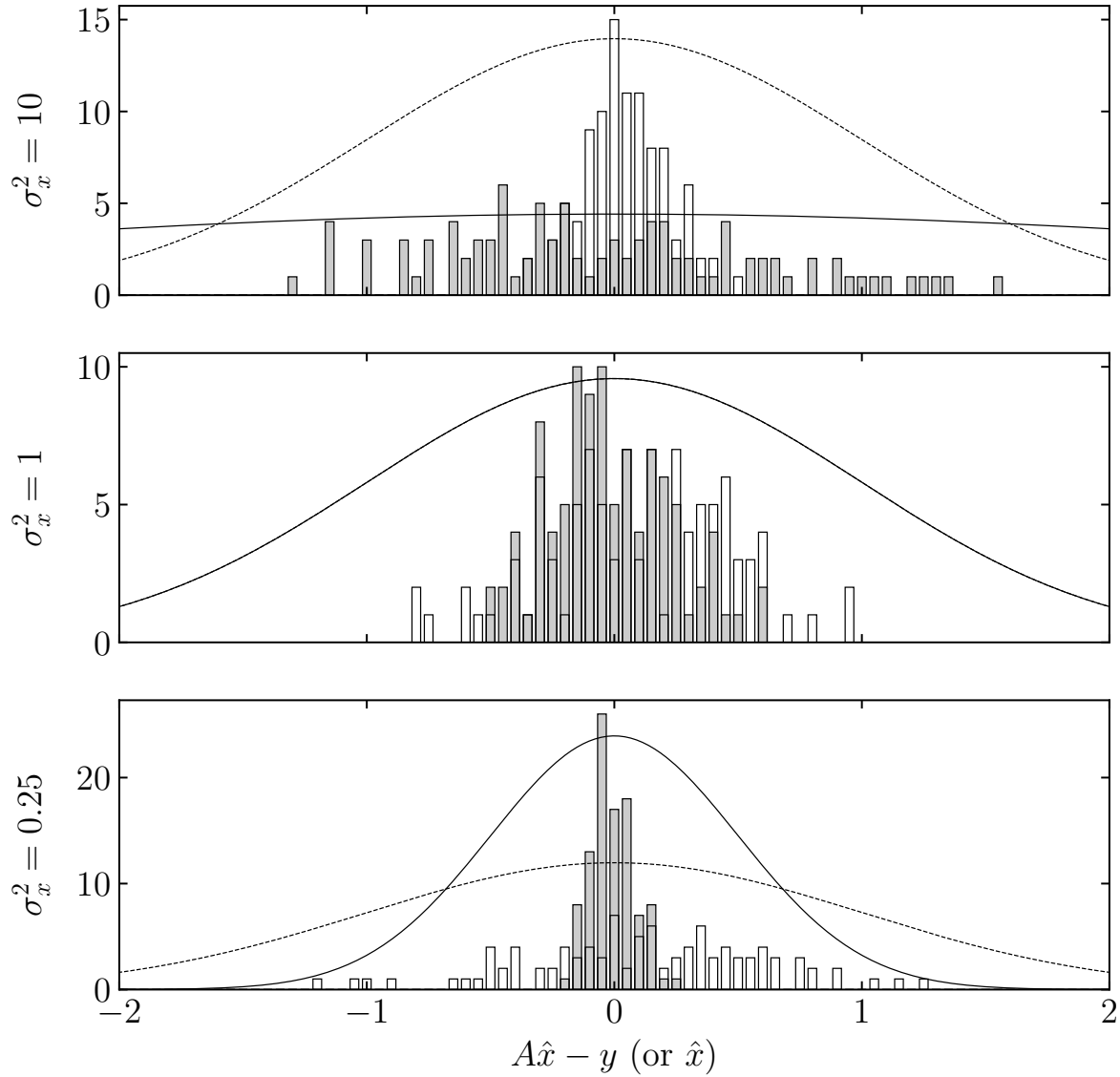
so MAP estimation solves

$$\text{minimize} \quad \|Ax - y\|_2^2 + (1/\sigma_x^2)\|x\|_2^2$$

- variable x, data y, and (hyper)parameter σ_x^2
- σ_x^2 controls the trade-off between likelihood and prior

trade-off parameter in MAP estimation usually implicit in density functions

example for $A \in \mathbf{R}^{100\times 100}$: (estimation $\hat{x}$ shown shaded)



larger σ_x^2 means larger variance in prior, so less regularization effect

Matrix regularizers

$$\text{minimize} \quad f_0(X) + \gamma\phi(X)$$

- variable $X \in \mathbf{R}^{m\times n}$; primary objective $f_0 \colon \mathbf{R}^{m\times n} \to \mathbf{R}$
- regularization function $\phi \colon \mathbf{R}^{m\times n} \to \mathbf{R}$, weighted by $\gamma > 0$

componentwise regularization:

- Tikhonov: $\phi(X) = \|X\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n X_{ij}^2$
- sparsity: $\phi(X) = \|X\|_{\text{sav}} = \sum_{i=1}^m \sum_{j=1}^n |X_{ij}|$

suppose $X = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$, **columnwise sparsity** regularizer is

$$\phi(X) = \sum_{i=1}^n \|x_i\|_2$$

(note that there is no square on the ℓ_2-norm)

example: group lasso approximately solves the linear equations

$$\mathbf{tr}(A_i^T X) = b_i, \quad i = 1, \ldots, k$$

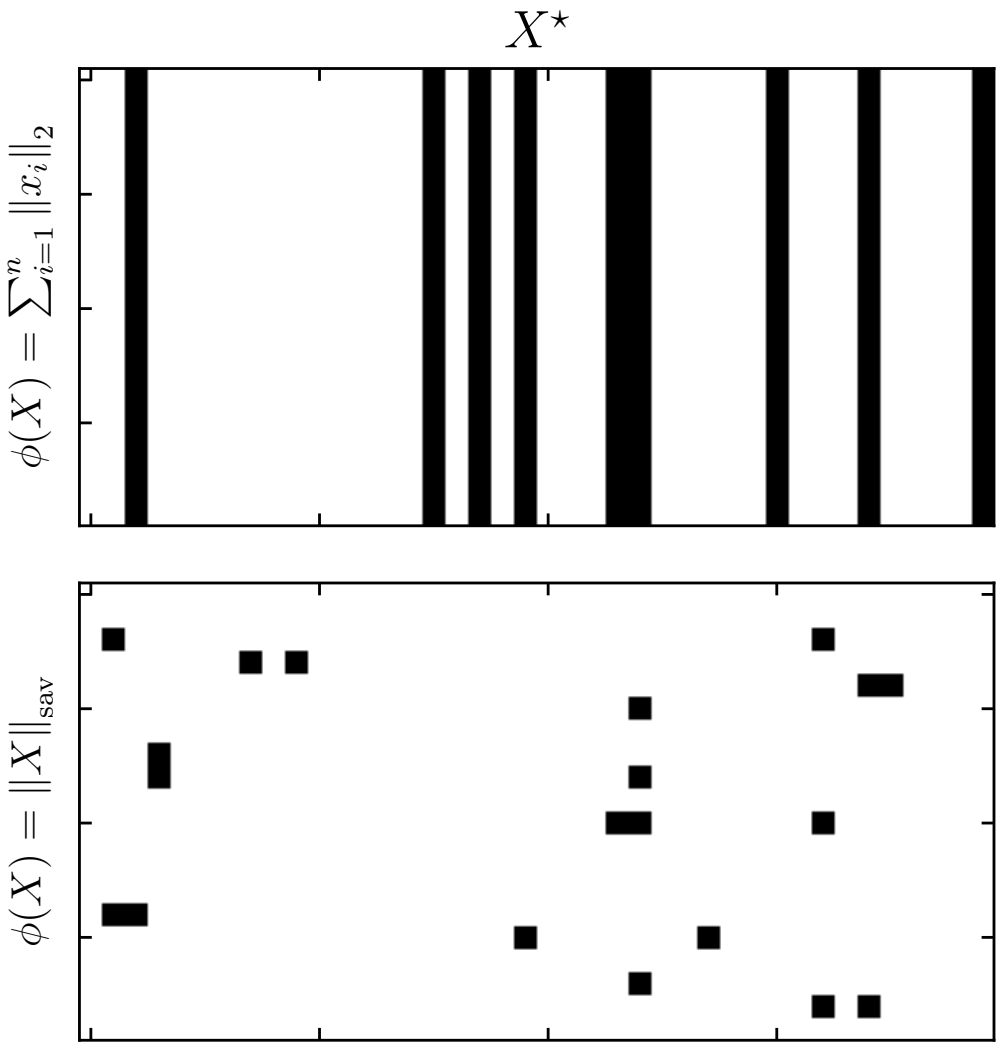
with columnwise sparsity regularization:

$$\text{minimize} \quad \sum_{i=1}^k \left(\sum_{j=1}^n a_{ij}^T x_j - b_i \right)^2 + \gamma \sum_{i=1}^n \|x_i\|_2$$

- variable $X \in \mathbf{R}^{m \times n}$; data $A_i \in \mathbf{R}^{m \times n}$, $b_i \in \mathbf{R}$
- $a_{ij} \in \mathbf{R}^m$ is the jth column of A_i (called *feature group*)
- $x_i \in \mathbf{R}^m$ is the ith column of X

- sparsity regularization encourages some x_i to be zero, $i.e.$, some feature groups are not used

numerical results for $m = 20$, $n = 40$, $k = 20$:
(compared with componentwise sparsity regularizer; zeros shown black)



Rank regularization

$$\text{minimize} \quad f_0(X) + \gamma \, \mathbf{rank}\, X$$

- variable $X \in \mathbf{R}^{m \times n}$; primary objective $f_0 \colon \mathbf{R}^{m \times n} \to \mathbf{R}$
- $\mathbf{rank}\, X$ counts the rank (number of nonzero singular values) of X, weighted by $\gamma > 0$
- generalizes cardinality sparsity regularization to matrices
- nonconvex so very hard to solve

usually solved via heuristic:

$$\text{minimize} \quad f_0(X) + \gamma \|X\|_*$$

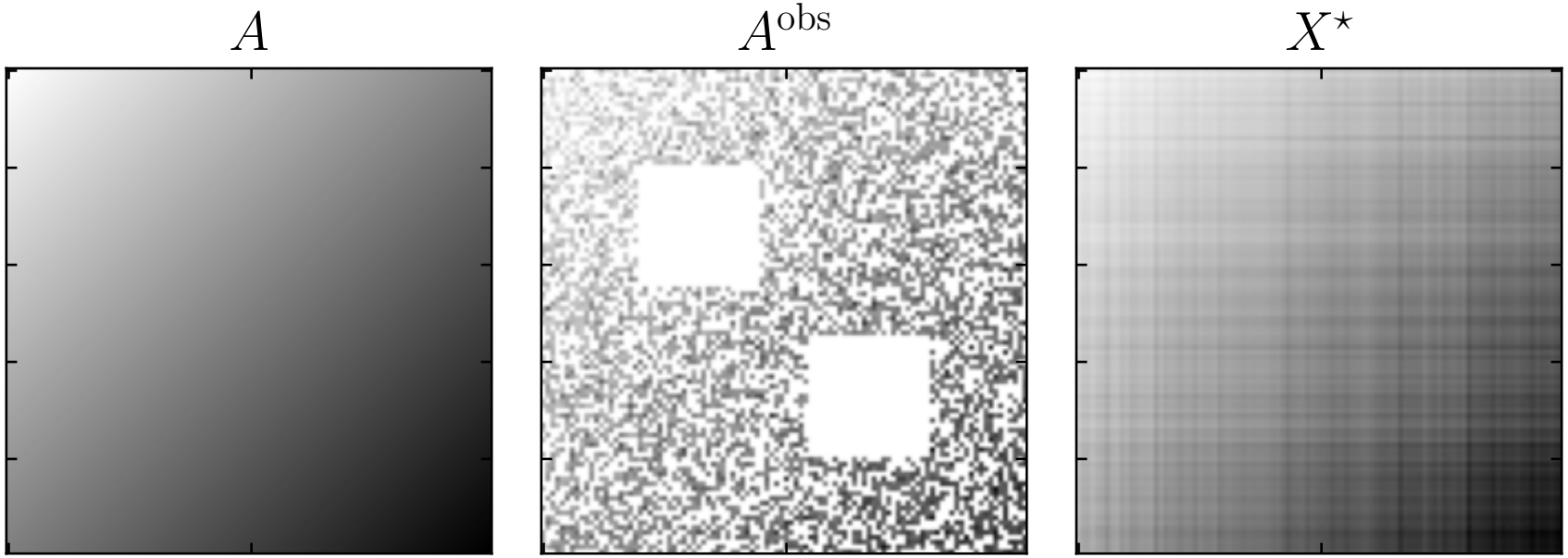
- $\|X\|_* = \sum_{i=1}^{\min\{m,n\}} \sigma_i(X)$ is the nuclear norm of X
- replace the number of nonzero singular values with their sum

example: low-rank matrix completion for $A \in \mathbf{R}^{m\times n}$

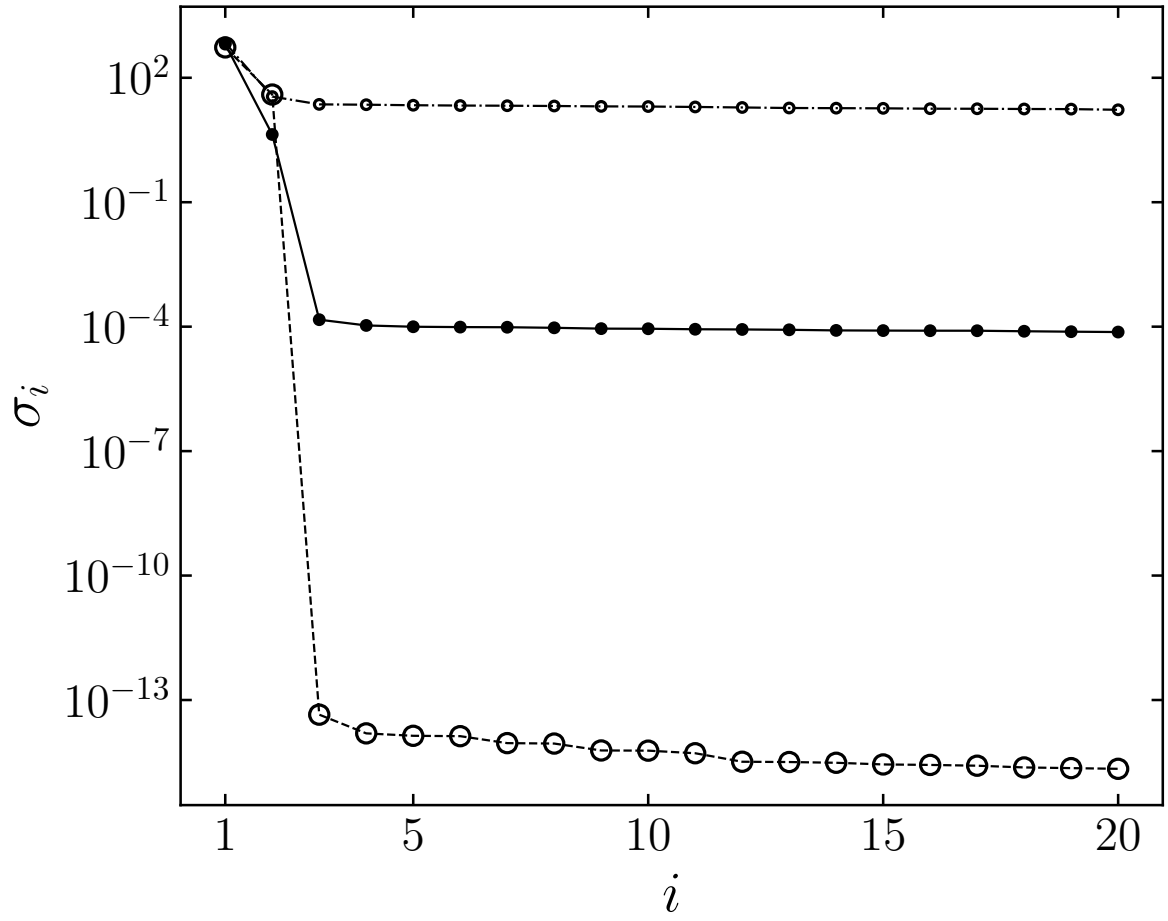
$$\text{minimize} \quad \sum_{(i,j)\in\mathcal{I}} \left(X_{ij} - A^{\text{obs}}_{ij}\right)^2 + \gamma\|X\|_*$$

- variable $X \in \mathbf{R}^{m\times n}$; observation A^{obs}_{ij} of A for $(i,j) \in \mathcal{I}$ is given

results for $A \in \mathbf{R}^{100\times 100}$ with rank 2; observations are corrupted with Gaussian noise and 60% of entries are missing:



singular values of A (circles bottom), A^{obs} (circles top), and $X^\star$ (dots):



7. Constraints

- constrained optimization problems
- examples of constraints
- least norm problems
- probabilities and distributions
- functional constraints
- relaxations and sensitivity analysis
- the Lagrangian, Lagrange dual function, and dual problem
- weak and strong duality
- complementary slackness and KKT optimality conditions
- infeasible problems and exact penalty methods

Constrained optimization problems

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

- $f_1, \ldots, f_m : \mathbf{R}^n \to \mathbf{R}$ are inequality constraint functions
- $h_1, \ldots, h_p : \mathbf{R}^n \to \mathbf{R}$ are equality constraint functions
- physical properties, safety requirements, resource limitations, etc.

sometimes written in **abstract form**

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathcal{X} \end{array}$$

with **constraint set** $\mathcal{X} = \left\{ x \in \mathbf{R}^n \;\middle|\; \begin{array}{l} f_i(x) \leq 0, \; i = 1, \ldots, m \\ h_i(x) = 0, \; i = 1, \ldots, p \end{array} \right\}$

Regularization and constraints

- both represent prior knowledge about the solution
- regularization: soft constraints, penalize violation
- constraints: hard constraints, violation unacceptable

example: solution x should be small in $\|\cdot\|_2$

- Tikhonov regularization:

$$\text{minimize} \quad \|Ax-b\|_2^2 + \lambda\|x\|_2^2$$

- Euclidean ball constraint:

$$\begin{array}{ll} \text{minimize} & \|Ax-b\|_2^2 \\ \text{subject to} & \|x\|_2^2 \leq t \end{array}$$

Examples of constraints

- **nonnegativity**:
  - $x \succeq 0$ for nonnegative least squares
  - componentwise $X_{ij} \geq 0$ for nonnegative matrix factorization
- **box constraints**:

$$\begin{array}{ll} \text{minimize} & \|Ax - b\| \\ \text{subject to} & l \preceq x \preceq u \end{array}$$

  projects b onto the image of the box $\{x \mid l \preceq x \preceq u\}$ under A

  - equivalent to $\|x\|_\infty \leq t$ for equally bounded components
- **probability simplex**: $x \succeq 0$, $\mathbf{1}^T x = 1$ for multinomial distributions
- **linear constraints**: $Cx \preceq d$, $Gx = h$; include, *e.g.*, box constraints by

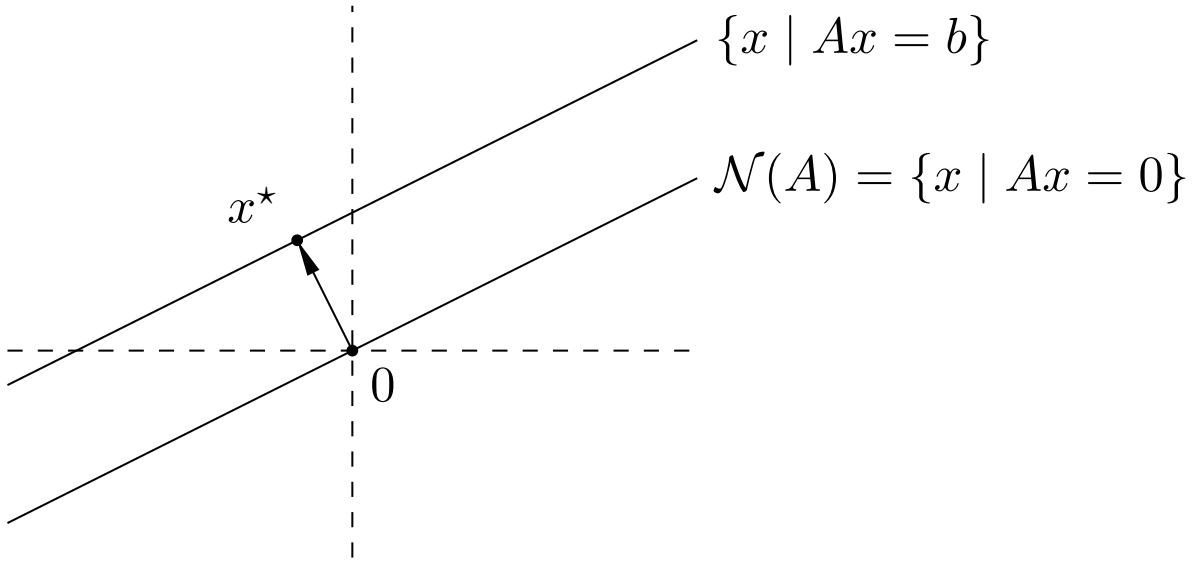
$$C = \begin{bmatrix} I \\ -I \end{bmatrix}, \quad d = \begin{bmatrix} u \\ -l \end{bmatrix}, \quad G = 0, \quad h = 0$$

- **trust region constraints**: $\|x - x_0\| \leq t$ for some $x_0 \in \mathbf{R}^n$

Least norm problems

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

- $Ax = b$ underdetermined, *e.g.*, $A \in \mathbf{R}^{m \times n}$ is fat with $m < n$
  - affine solution set $\{x_0 + z \mid z \in \mathcal{N}(A)\}$ (x_0 a particular solution)
- **geometric interpretation**: find point in the solution set closest to zero



Probabilities and distributions

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \\ & x \in \mathcal{X} \end{array}$$

- find nonparametric distribution x minimizing some objective $f_0(x)$
- many useful types of constraint set $\mathcal{X}$

moment constraints:

$$\mathbf{E}\, Z^k = \sum_{i=1}^n c_i^k x_i$$

is kth moment of random variable $Z \in \{c_1, \ldots, c_n\}$

- linear equality constraints: $\mathbf{E}\, Z = c^T x = \mu$; $\mathbf{E}\, Z^k = \sum_{i=1}^n c_i^k x_i = \eta$
- linear inequality constraints: $\alpha \leq \sum_{i=1}^n c_i^k x_i \leq \beta$
- generalization: expectation $\mathbf{E}\, g(Z) = \sum_{i=1}^n g(c_i) x_i$ of some function g

probability constraints:

$$\mathbf{prob}(Z \in S) = p^T x, \qquad p_i = \begin{cases} 1, & c_i \in S \\ 0, & \text{otherwise,} \end{cases} \qquad i = 1, \ldots, n$$

- probability of $Z \in \{c_1, \ldots, c_n\}$ in set S is linear function of x
- known or bounded probability: $p^T x = \mu$ or $\alpha \leq p^T x \leq \beta$
- bounded conditional probability:

$$\mathbf{prob}(Z \in A \mid Z \in B) = \frac{\mathbf{prob}(Z \in A \cap B)}{\mathbf{prob}(Z \in B)} = \frac{p^T x}{q^T x}$$

with $p_i = 1$ if $c_i \in A \cap B$, $q_i = 1$ if $c_i \in B$, and 0 otherwise, so

$$\alpha \leq \frac{p^T x}{q^T x} \leq \beta \quad \Longleftrightarrow \quad (p - \alpha q)^T x \geq 0 \quad \text{and} \quad (\beta q - p)^T x \geq 0$$

variance constraints: for random variable $Z \in \{c_1, \ldots, c_n\}$

$$\mathbf{var}\, Z = \mathbf{E}\, Z^2 - (\mathbf{E}\, Z)^2 = \sum_{i=1}^n c_i^2 x_i - \left(\sum_{i=1}^n c_i x_i\right)^2$$

$$= \sum_{i=1}^n c_i^2 x_i - x^T (cc^T) x$$

- difference of linear and convex quadratic functions of x
- lower bound $\mathbf{var}\, Z \geq \mu$ is convex constraint

$$x^T (cc^T) x - \sum_{i=1}^n c_i^2 x_i \leq -\mu$$

- upper bound $\mathbf{var}\, Z \leq \eta$ is DC constraint

$$\sum_{i=1}^n c_i^2 x_i - x^T (cc^T) x \leq \eta$$

entropy constraints:

- distribution x has at least μ level of uncertainty:

$$\sum_{i=1}^{n} x_i \log x_i \leq -\mu$$

- KL-divergence between x and given reference q is at most η:

$$\sum_{i=1}^{n} x_i \log \frac{x_i}{q_i} \leq \eta$$

both are convex constraints in x

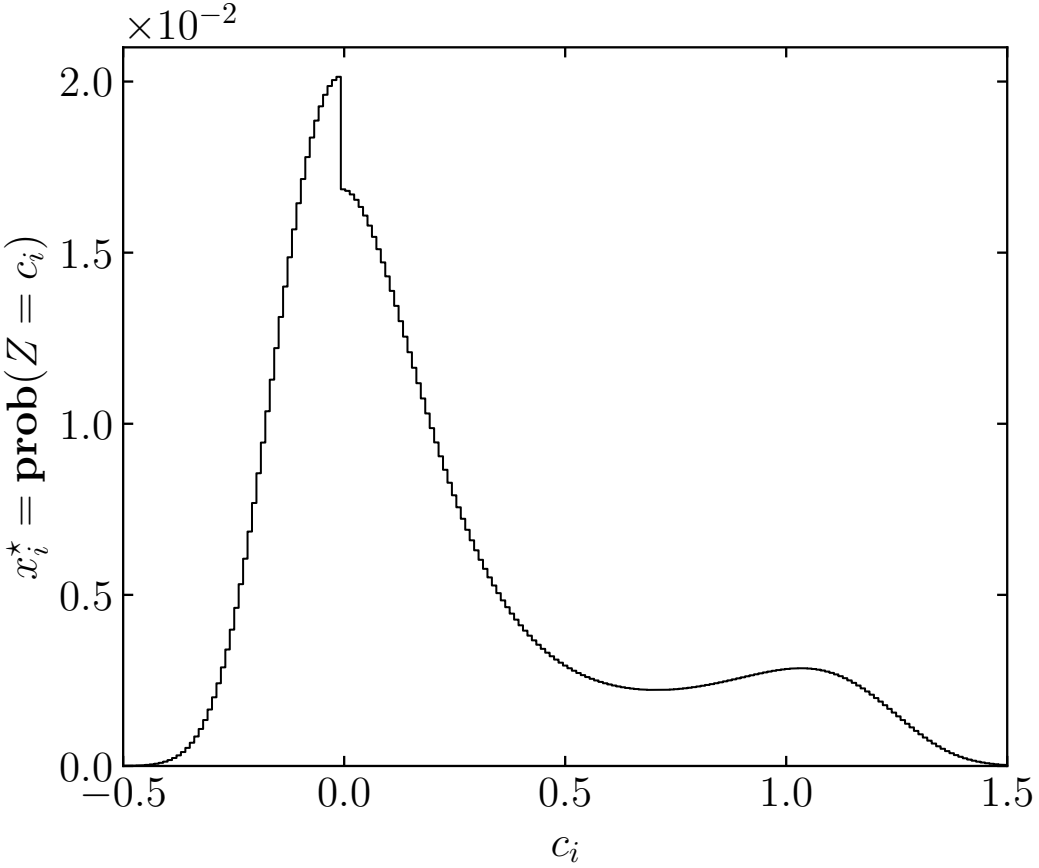
example:

- maximum entropy estimation:

  minimize $\sum_{i=1}^n x_i \log x_i$

  subject to $x \succeq 0,\ \mathbf{1}^T x = 1,\ \mathbf{prob}(Z < 0) \geq 0.35,\ \mathbf{var}\, Z \geq 0.1$

  $-0.2 \leq \mathbf{E}\, Z \leq 0.5,\ \mathbf{E}\, Z^2 \leq 0.2,\ \mathbf{E}(3Z^4 - 7Z^3) \leq -0.7$

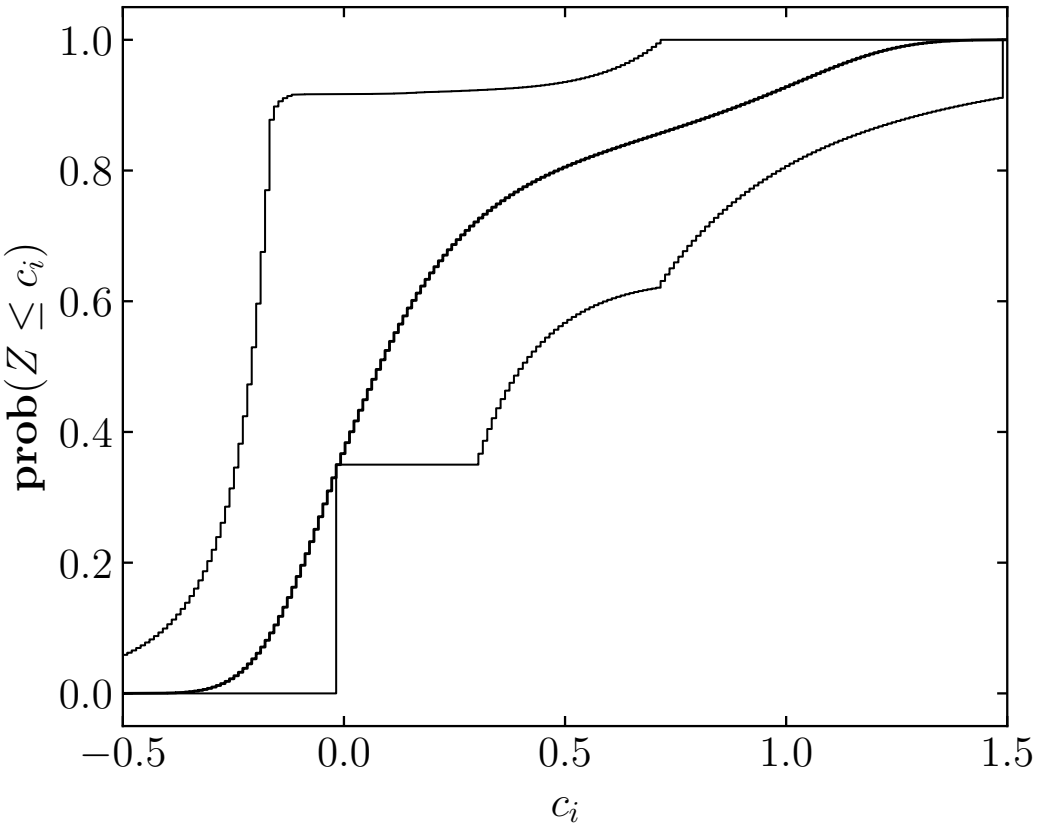


- bounding cumulative density function:

  min./max. $\mathbf{prob}(Z \leq c_i)$

  subject to $x \succeq 0,\ \mathbf{1}^T x = 1,\ \mathbf{prob}(Z < 0) \geq 0.35,\ \mathbf{var}\, Z \geq 0.1$

  $-0.2 \leq \mathbf{E}\, Z \leq 0.5,\ \mathbf{E}\, Z^2 \leq 0.2,\ \mathbf{E}(3Z^4 - 7Z^3) \leq -0.7$



(maximum entropy distribution shown thicker)

Functional constraints

function fitting problems: given $(z_i, y_i) \in \mathbf{R}^k \times \mathbf{R}$, $i = 1, \ldots, m$

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \phi(f(z_i) - y_i) \\ \text{subject to} & x \in \mathcal{X} \end{array}$$

fits $f : \mathbf{R}^k \to \mathbf{R}$ with $f(z) = x_1 f_1(z) + \cdots + x_n f_n(z)$ to the data points

- $x \in \mathbf{R}^n$ are unknown coefficients of *basis functions* $f_1, \ldots, f_n : \mathbf{R}^k \to \mathbf{R}$
- $\phi : \mathbf{R} \to \mathbf{R}$ is penalty for misfit
- approximation interpretation: $a_i = (f_1(z_i), \ldots, f_n(z_i))$ so $f(z_i) = a_i^T x$
- example: polynomial fitting for $(z_i, y_i) \in \mathbf{R}^2$ with
  - $f(z) = x_1 + x_2 z + \cdots + x_n z^{n-1}$
  - $\phi(u) = u^2$

interpolation constraints: for (s_i, t_i), $i = 1, \ldots, p$, require

$$f(s_i) = t_i, \ i = 1, \ldots, p \quad \Longleftrightarrow \quad Ax = t$$

- $A \in \mathbf{R}^{p \times n}$ with rows $a_i^T = (f_1(s_i), \ldots, f_n(s_i))^T$
- feasible if and only if $t \in \mathcal{R}(A)$

bounding constraints: for (s_i, t_i), $i = 1, \ldots, p$, require

$$f(s_i) \in [\alpha_i, \beta_i], \ i = 1, \ldots, p \quad \Longleftrightarrow \quad \alpha \preceq Ax \preceq \beta$$

- nonnegativity: $f(s_i) = a_i^T x \geq 0$
- monotonicity: $f(s_{i+1}) - f(s_i) = (a_{i+1} - a_i)^T x \geq 0$
- lower bounding data points: $f(s_i) = a_i^T x \leq t_i$

(generalize to full $\mathbf{dom}\, f$ under some assumptions, *e.g.*, f piecewise affine)

Lipschitz continuity constraints: at s_i, $i = 1, \ldots, p$, require

$$|f(s_i) - f(s_j)| \leq L\|s_i - s_j\|, \quad i, j = 1, \ldots, p$$

$$\iff \quad -L\|s_i - s_j\| \leq (a_i - a_j)^T x \leq L\|s_i - s_j\|, \quad i, j = 1, \ldots, p$$

derivative constraints:

- bounding $\nabla f(s) = x_1 \nabla f_1(s) + \cdots + x_n \nabla f_n(s)$:

$$\|\nabla f(s)\| \leq M \quad \iff \quad \|Ax\| \leq M$$

  are convex, where $A = \begin{bmatrix} \nabla f_1(s) & \cdots & \nabla f_n(s) \end{bmatrix}$

- bounding $\nabla^2 f(s) = x_1 \nabla^2 f_1(s) + \cdots + x_n \nabla^2 f_n(s)$:

$$\alpha I \preceq \nabla^2 f(s) \preceq \beta I$$

  are LMIs in x

example: quadratic basis functions

$$f_i(z) = \frac{1}{2} z^T P_i z + q_i^T z + r_i, \quad i = 1, \ldots, n$$

so $f(z) = \frac{1}{2} z^T P z + q^T z + r$ with

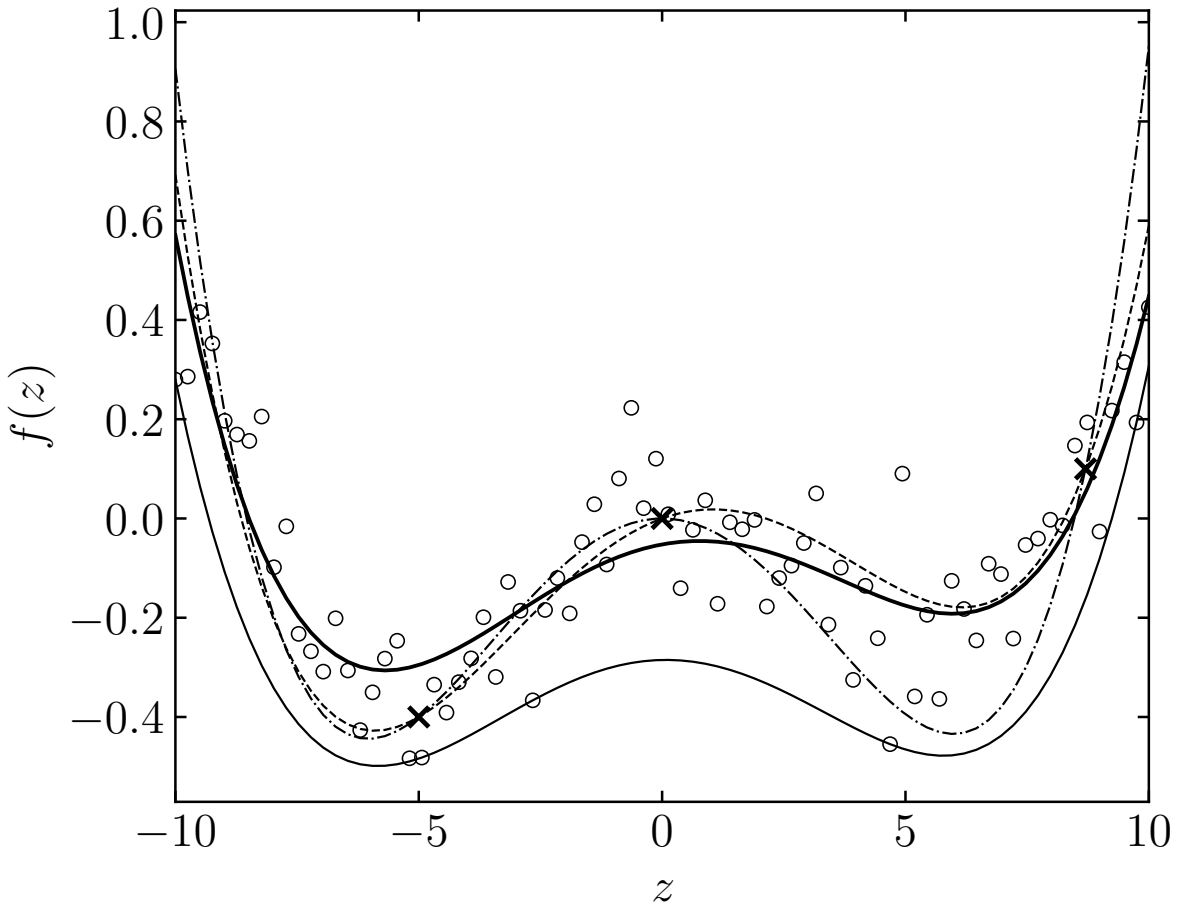
$$P = \sum_{i=1}^{n} x_i P_i, \quad q = \sum_{i=1}^{n} x_i q_i, \quad r = \sum_{i=1}^{n} x_i r_i$$

convexity constraints on f

$$\nabla^2 f(z) = P = \sum_{i=1}^{n} x_i P_i \succeq 0$$

are LMIs in x

numerical example:



- thick: unconstrained
- solid: lower bounding
- dashed: interpolation; dashdotted: interpolation + derivative

Relaxations

for two optimization problems

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathcal{P} \end{array} \quad \text{and} \quad \begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in \mathcal{Q} \end{array}$$

call the latter a **relaxation** of the former if $\mathcal{P} \subseteq \mathcal{Q}$

lower bounding property: let $p^\star, q^\star$ be optimal values of each problem,

$$q^\star = \inf_{x \in \mathcal{Q}} f_0(x) \leq \inf_{x \in \mathcal{P}} f_0(x) = p^\star$$

- especially useful if $\mathcal{Q}$ is convex and $\mathcal{P}$ is not
- relaxation infeasible $\implies$ original problem infeasible
- if relaxation optimal $x^\star \in \mathcal{P}$, then $x^\star$ is also original problem optimal

objective relaxation: if $g_0(x) \leq f_0(x)$ for all $x \in \mathcal{P}$, then

$$\text{minimize} \quad g_0(x) \qquad \text{relax} \qquad \text{minimize} \quad f_0(x)$$

- often seen as $g_0 = \mathbf{conv}\, f_0$
- first problem sometimes called *surrogate problem* of the second
- relaxation follows from epigraph formulation
- lower bounding property of optimal values holds
- optimal points relationship depends on shape of g_0

Sensitivity analysis

for (possibly nonconvex) optimization problem

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

relax the inequality constraints by $u \in \mathbf{R}^m$:

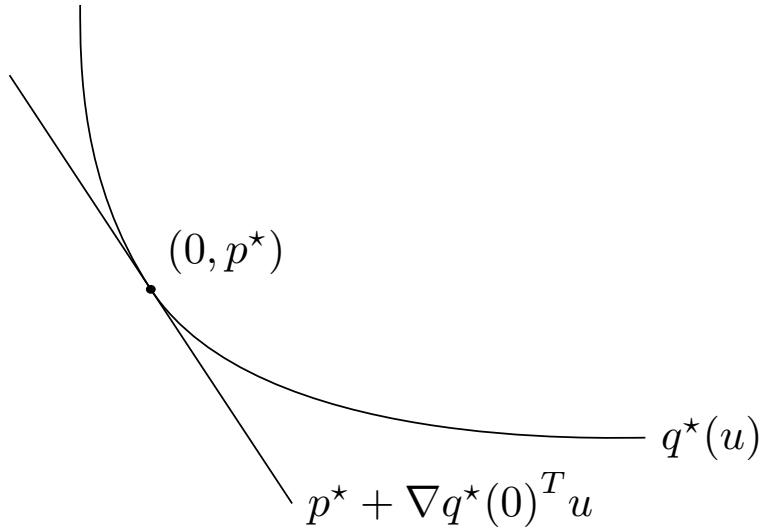
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq u_i, \quad i = 1, \ldots, m \end{array}$$

- second problem sometimes called *perturbation* of the first
- $u = 0$ is original problem
- $u_i > 0$: the ith constraint is relaxed
- $u_i < 0$: the ith constraint is tightened

let $p^\star$ be optimal value of original problem, define $q^\star\colon \mathbf{R}^m \to \mathbf{R}$

$$q^\star(u) = \inf \left\{ f_0(x) \;\middle|\; \begin{array}{l} x \in \bigcap_{i=0}^m \mathbf{dom}\, f_i \\ f_i(x) \le u_i, \quad i = 1, \ldots, m \end{array} \right\}$$

- gives optimal value of perturbed problem, as function of perturbation u
- for some u, $q^\star(u) = \infty \implies$ perturbation by u is infeasible
- $q^\star$ is convex if original problem is convex
- $q^\star(u) - q^\star(0)$ gives sensitivity to constraint perturbation



Cardinality and rank constraints

- cardinality constrained problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \mathbf{card}\, x \leq k, \quad \|x\|_\infty \leq 1 \end{array}$$

  has convex relaxation

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & \|x\|_1 \leq k, \quad \|x\|_\infty \leq 1 \end{array}$$

- similarly for rank constrained problem

$$\begin{array}{ll} \text{minimize} & f_0(X) \\ \text{subject to} & \mathbf{rank}\, X \leq k, \quad \|X\|_2 \leq 1 \end{array}$$

  we have convex relaxation

$$\begin{array}{ll} \text{minimize} & f_0(X) \\ \text{subject to} & \|X\|_* \leq k, \quad \|X\|_2 \leq 1 \end{array}$$

Boolean LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0,1\}, \quad i = 1, \ldots, n \end{array}$$

is often relaxed to LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & 0 \preceq x \preceq 1 \end{array}$$

- boolean LP is NP-hard, but the relaxation can be solved very efficiently
- if relaxation solution $x^\star$ happens to be boolean, then it is also optimal for the original problem
- similar ideas apply to other problems with integer constraints

Two-way partitioning

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i \in \{-1, 1\}, \quad i = 1, \ldots, n \end{array}$$

- feasible x consists in partitioning n objects into two groups -1 and 1:

$$\{1, \ldots, n\} = \{i \mid x_i = 1\} \cup \{i \mid x_i = -1\}$$

- $W \in \mathbf{S}^n$: weight matrix encoding the pairwise relationships
  - $W_{ij}/-W_{ij}$: cost of putting i and j in the same/different group
- $x_i \in \{-1, 1\} \implies x_i^2 = 1$, so we have equivalent formulation

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X \succeq 0, \quad \mathbf{rank}\, X = 1 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{array}$$

where $X = xx^T \in \mathbf{S}_+^n$ is a rank-one matrix

- relaxation: drop rank constraint, get SDP

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n \end{array}$$

The Lagrangian

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

- variable $x \in \mathbf{R}^n$; no assumptions on $f_i, h_i \colon \mathbf{R}^n \to \mathbf{R}$

Lagrangian $L \colon \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ of the problem is

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

- $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$, where $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$
- λ_i, ν_i are **Lagrange multipliers** for the ith inequality and equality constraints, respectively
- $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$ are **dual variables**

Lagrange dual function

Lagrange dual function (or just *dual function*) $g\colon \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ is

$$g(\lambda,\nu) = \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) = \inf_{x\in\mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

- partial minimization of Lagrangian L over x for fixed λ, ν
- $g(\lambda,\nu) = -\infty$ if $L(x,\lambda,\nu)$ is unbounded below on $\mathcal{D}$

g is **concave** (in λ,ν) even if original problem is not convex:

- for each fixed x, $L(x,\lambda,\nu)$ is affine in (λ,ν)
- pointwise infimum of affine functions is concave

g provides **lower bound** on optimal value $p^\star$ of original problem, *i.e.*,

$$g(\lambda, \nu) \leq p^\star \text{ for all } \lambda \succeq 0,\ \nu$$

- holds trivially if $g(\lambda, \nu) = -\infty$ for some $\lambda \succeq 0,\ \nu$
- informative only when $\lambda \succeq 0$, $(\lambda, \nu) \in \mathbf{dom}\, g$ (*i.e.*, $g(\lambda, \nu) > -\infty$)
  such (λ, ν) are called **dual feasible** points

proof of lower bound:

- let $\tilde{x} \in \mathcal{D}$ be feasible, so

$$f_i(\tilde{x}) \leq 0, \quad i = 1, \ldots, m, \qquad h_i(\tilde{x}) = 0, \quad i = 1, \ldots, p$$

- then for all $\lambda \in \mathbf{R}^m_+$, $\nu \in \mathbf{R}^p$,

$$\sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \leq 0$$

- hence, we have upper bound on Lagrangian

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x})$$

- taking infimum over $x \in \mathcal{D}$ gives

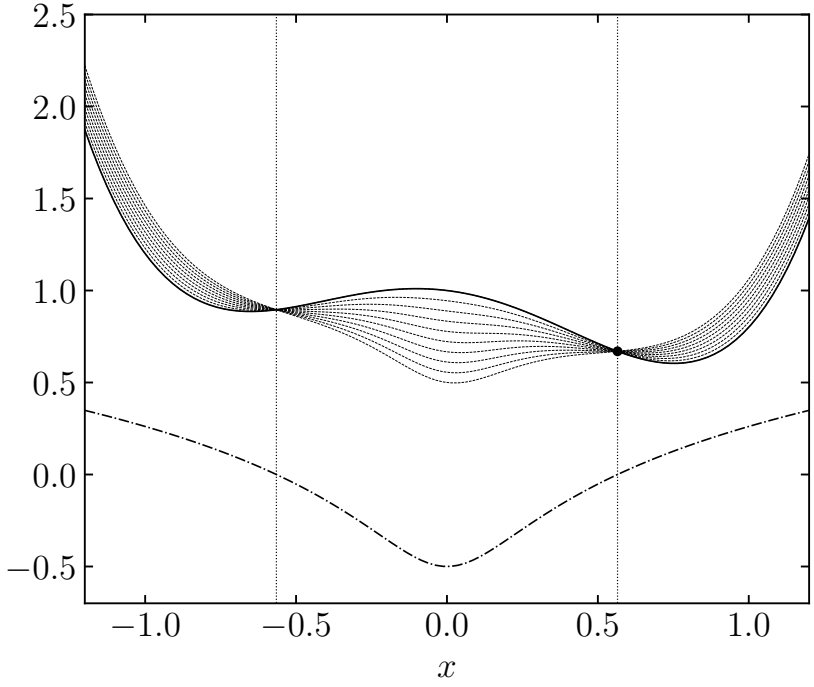
$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$$

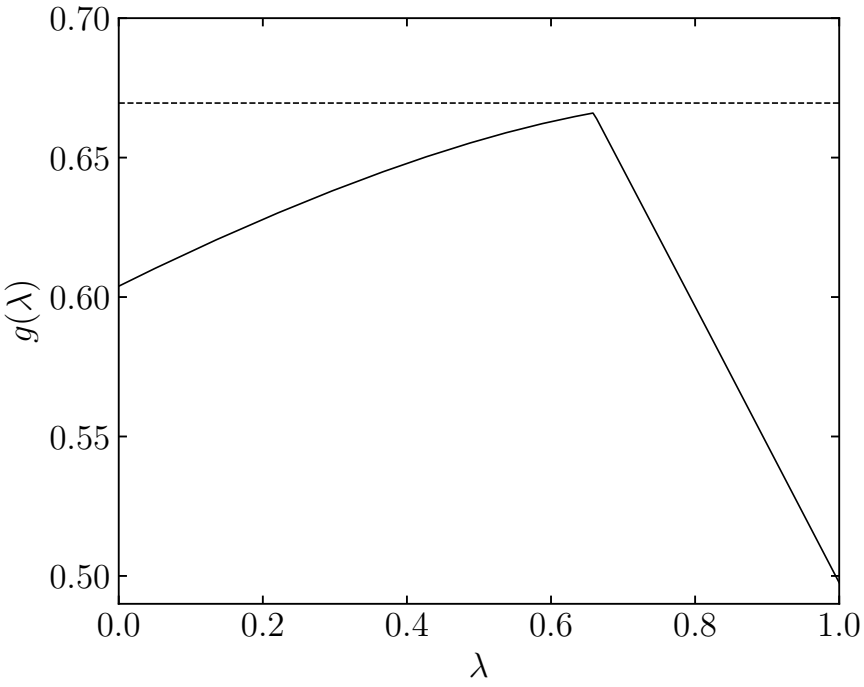
- since the inequality holds for all feasible points $\tilde{x} \in \mathcal{D}$, we have

$$g(\lambda, \nu) \leq \inf_{x \in \mathcal{D}} \left\{ f_0(x) \;\middle|\; \begin{array}{ll} f_i(x) \leq 0, & i = 1, \ldots, m \\ h_i(x) = 0, & i = 1, \ldots, p \end{array} \right\} = p^\star$$

example:

$$\begin{array}{ll} \text{minimize} & f_0(x) = x^4 - x^2 - (1/5)x + 1 \\ \text{subject to} & f_1(x) = (1/4)\log(1 + 20x^2) - 0.5 \leq 0 \end{array}$$





- $f_0(x)$ solid; $f_1(x)$ dashdotted; $L(x, \lambda)$ dashed for different $\lambda > 0$
- dashed line on right shows optimal value $p^\star$; $g(\lambda)$ shown solid

Interpretation of dual function

relaxation interpretation: $g(\lambda, \nu)$ is optimal value of

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \end{array}$$

$$\begin{array}{ll} \text{minimize} & f_0(x) + \lambda^T u + \nu^T v \\ \iff \quad \text{subject to} & f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{array}$$

- $x \in \mathbf{R}^n$, $u \in \mathbf{R}^m$, $v \in \mathbf{R}^p$ are variables
- $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$ are fixed parameters
- always feasible for any λ, ν, so original problem is relaxed to an unconstrained problem
- constraint violation is penalized by $\lambda^T u + \nu^T v$

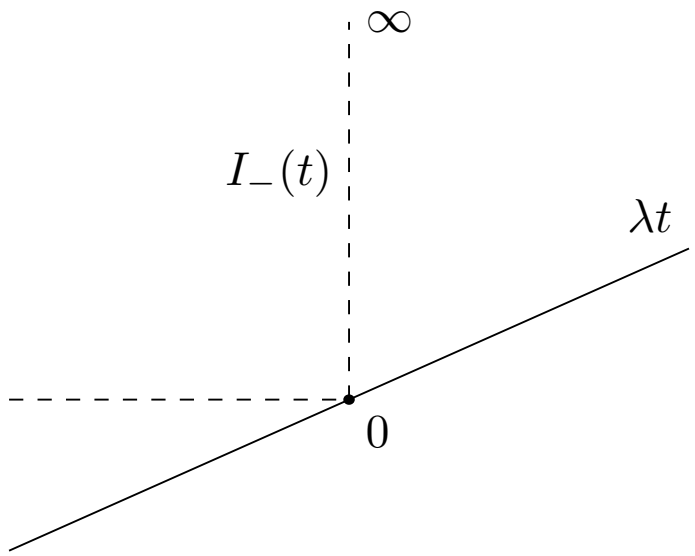
linear approximation interpretation: original problem equals to

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^m I_-(f_i(x)) + \sum_{i=1}^p I_0(h_i(x))$$

where $I_-, I_0 : \mathbf{R} \to \mathbf{R}$ are *indicator functions* of $-\mathbf{R}_+$ and $\{0\}$, *i.e.*,

$$I_-(t) = \begin{cases} 0, & t \leq 0 \\ \infty, & \text{otherwise} \end{cases} \quad \text{and} \quad I_0(t) = \begin{cases} 0, & t = 0 \\ \infty, & \text{otherwise} \end{cases}$$

so $g(\lambda, \nu)$ consists in replacing I_-, I_0 by their linear approximations



- objective relaxation: $\lambda_i f_i(x) \leq I_-(f_i(x))$, $\nu_i h_i(x) \leq I_0(h_i(x))$ for all x

Lagrange dual problem

call original problem **primal problem**

then its associated **(Lagrange) dual problem** is

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

- variables: $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$
- find best lower bound on optimal value of primal problem
- dual problem always convex, even if primal is not
- feasible set: $\{(\lambda, \nu) \in \mathbf{dom}\, g \mid \lambda \succeq 0\}$
- optimal value $d^\star$ called **dual optimal** and $p^\star$ is **primal optimal**

Weak duality

for *all* primal and dual problem pairs, we have

$$d^\star \leq p^\star$$

- follows from $g(\lambda, \nu) \leq f_0(x)$ for all feasible x and dual feasible (λ, ν)
- $p^\star - d^\star \geq 0$ is **(optimal) duality gap**
- some useful consequences:
  - dual problem unbounded above $\implies$ primal problem infeasible
  - primal unbounded below $\implies$ dual infeasible

  serves as a *certificate* of problem feasibility and boundedness

example: two-way partitioning ($W \in \mathbf{S}^n$)

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n \end{array}$$

$$L(x, \nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1) = x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu$$

$$g(\nu) = \inf_x L(x, \nu) = \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu$$

$$= \begin{cases} -\mathbf{1}^T \nu, & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

hence, dual problem is an SDP (efficiently solved; lower bounding primal)

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array}$$

Strong duality

for *some* primal and dual problem pairs, we have

$$d^\star = p^\star$$

- condition for strong duality called **constraint qualification**
  - many types exist, depending on problem structure
  - example: for LP, strong duality holds if primal feasible
- *roughly speaking*, strong duality *usually* holds for convex problems

certificate of optimality:

- for $d^\star$ with strong duality and primal feasible $x^\star$:

$$f_0(x^\star) = d^\star \quad \Longrightarrow \quad x^\star \text{ primal optimal}$$

- in general, for primal feasible $\tilde{x}$ and dual feasible $(\tilde{\lambda}, \tilde{\nu})$,

$$p^\star \in [g(\tilde{\lambda}, \tilde{\nu}), f_0(\tilde{x})], \qquad d^\star \in [g(\tilde{\lambda}, \tilde{\nu}), f_0(\tilde{x})]$$

  - $f_0(\tilde{x}) - g(\tilde{\lambda}, \tilde{\nu})$ is **duality gap** and gives suboptimality of $\tilde{x}$
  - if $f_0(\tilde{x}) = g(\tilde{\lambda}, \tilde{\nu})$, then $\tilde{x}$ is primal optimal and $(\tilde{\lambda}, \tilde{\nu})$ is dual optimal

sensitivity analysis: perturbed primal optimal value $q^\star : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$,

$$q^\star(u, v) = \inf \left\{ f_0(x) \;\middle|\; \begin{array}{l} x \in \mathcal{D} \\ f_i(x) \le u_i, \;\; i = 1, \ldots, m \\ h_i(x) = v_i, \;\; i = 1, \ldots, p \end{array} \right\}$$

assume strong duality, we have

$$\begin{aligned} & q^\star(0,0) = p^\star = d^\star = g(\lambda^\star, \nu^\star) = \inf_{x \in \mathcal{D}} L(x, \lambda^\star, \nu^\star) \\ \Longrightarrow \quad & q^\star(0,0) \le f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x), \quad x \in \mathcal{D} \end{aligned}$$

since for $x \in \mathcal{D}$, $f_i(x) \leq u_i$, $h_i(x) = v_i$, we have

$$f_0(x) \geq q^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v$$

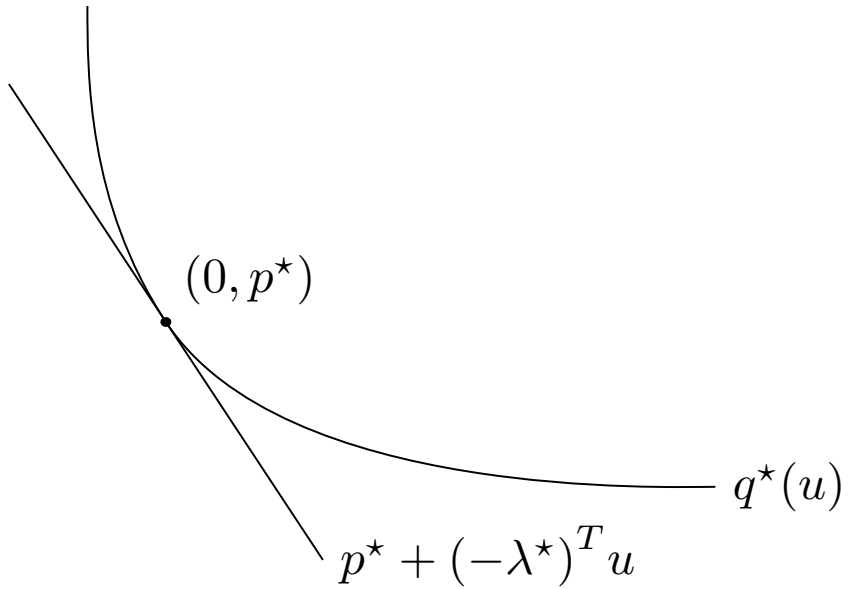
$$\Longrightarrow \quad q^\star(u,v) \geq q^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v$$

provides *global* linear underestimator of $q^\star$ at primal optimal

for differentiable $q^\star$:

$$\frac{\partial q^\star(0,0)}{\partial u_i} = -\lambda_i^\star, \quad i = 1, \ldots, m$$

$$\frac{\partial q^\star(0,0)}{\partial v_i} = -\nu_i^\star, \quad i = 1, \ldots, p$$



dual optimal $\lambda^\star$ and $\nu^\star$ give local sensitivity of primal optimal value to perturbations in constraints

Complementary slackness

suppose strong duality holds, $x^\star$ primal optimal, $(\lambda^\star, \nu^\star)$ dual optimal,

$$\begin{aligned} f_0(x^\star) = p^\star = d^\star &= g(\lambda^\star, \nu^\star) \\ &= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

so inequality must hold with equality

therefore we have the following *necessary* conditions:

- $x^\star$ primal optimal $\implies$ $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ (but not necessarily unique)

- **complementary slackness**:

$$\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0 \quad \Longrightarrow \quad \lambda_i^\star f_i(x^\star) = 0,\ i = 1, \ldots, m$$

  in other words,

  – if $\lambda_i^\star > 0$, then $f_i(x^\star) = 0$ (constraint is active)

  – if $f_i(x^\star) < 0$, then $\lambda_i^\star = 0$ (constraint inactive)

  so either the ith constraint is active or its dual variable is zero

KKT optimality conditions

assume differentiable objective, strong duality holds,

necessary conditions for $x^\star$ primal optimal, $(\lambda^\star, \nu^\star)$ dual optimal:

$$\begin{aligned}
f_i(x^\star) &\leq 0, \quad i = 1, \ldots, m \\
h_i(x^\star) &= 0, \quad i = 1, \ldots, p \\
\lambda_i^\star &\geq 0, \quad i = 1, \ldots, m \\
\lambda_i^\star f_i(x^\star) &= 0, \quad i = 1, \ldots, m \\
\nabla f_0(x^\star) + \textstyle\sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) &= 0
\end{aligned}$$

- first two: primal feasibility; third: dual feasibility
- forth: complementary slackness
- last: stationarity of Lagrangian
- no convexity assumption needed

if primal problem is convex, KKT conditions are **sufficient** for optimality

to see this, suppose $\tilde{x}$ primal feasible, $(\tilde{\lambda}, \tilde{\nu})$ dual feasible, and

$$\begin{array}{rcll} \tilde{\lambda}_i f_i(\tilde{x}) & = & 0, & i = 1, \ldots, m \\ \nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) & = & 0 & \end{array}$$

for convex primal problem, Lagrangian $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in x, so

$$g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i h_i(\tilde{x}) = f_0(\tilde{x})$$

- first equality follows from stationarity
- last equality from complementary slackness and feasibility
- implies zero duality gap, so $\tilde{x}$ primal optimal and $(\tilde{\lambda}, \tilde{\nu})$ dual optimal
- no strong duality assumption needed

Infeasible problems

consider *infeasible* feasibility problem

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

to find the largest subset of constraints that is feasible,

$$\begin{array}{ll} \text{minimize (over } x, u, v) & \sum_{i=1}^m u_i + \sum_{i=1}^p |v_i| \\ \text{subject to} & u \succeq 0, \quad f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{array}$$

- relaxation variables $u \in \mathbf{R}^m$, $v \in \mathbf{R}^p$ measure constraint violations
- objective equivalent to $\|u\|_1 + \|v\|_1$
- so u and v are expected to be sparse

Exact penalty methods

exact penalty method for constrained optimization

$$\begin{array}{ll} \text{minimize} & f_0(x) + \gamma\left(\sum_{i=1}^m u_i + \sum_{i=1}^p |v_i|\right) \\ \text{subject to} & u \succeq 0, \quad f_i(x) \leq u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p \end{array}$$

- variable $x \in \mathbf{R}^n$, $u \in \mathbf{R}^m$, $v \in \mathbf{R}^p$; penalty parameter $\gamma > 0$

regularization interpretation: transform constraints into regularization terms in objective; certainty of constraint satisfaction depends on γ

- if original problem feasible, sufficiently large γ guaranteed to give originally feasible solution

- if original problem infeasible, solution gives sparsest constraint violation

example: infeasible least norm problem

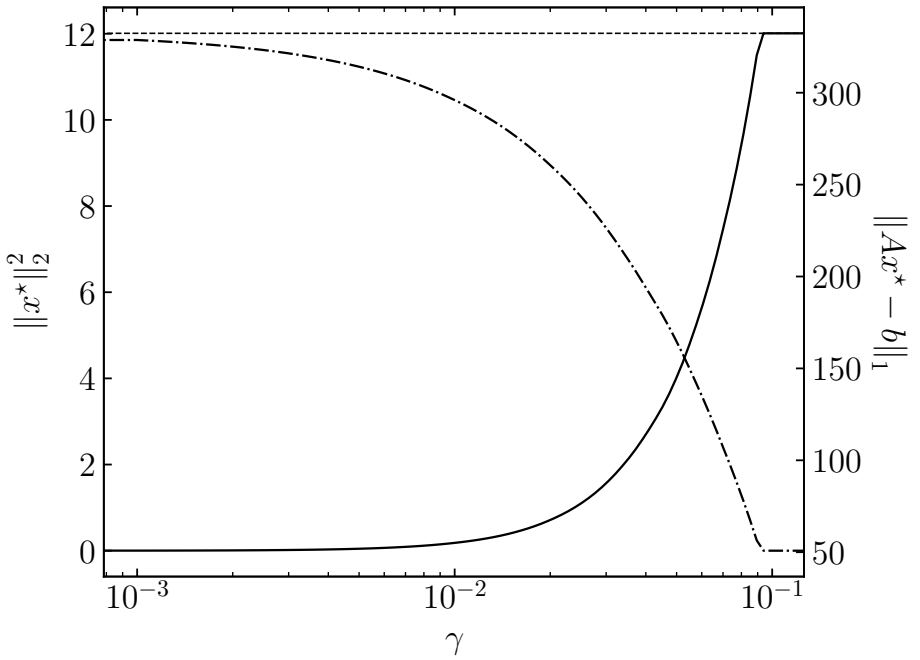
$$\begin{array}{ll} \text{minimize} & \|x\|_2^2 \\ \text{subject to} & Ax = b \end{array}$$

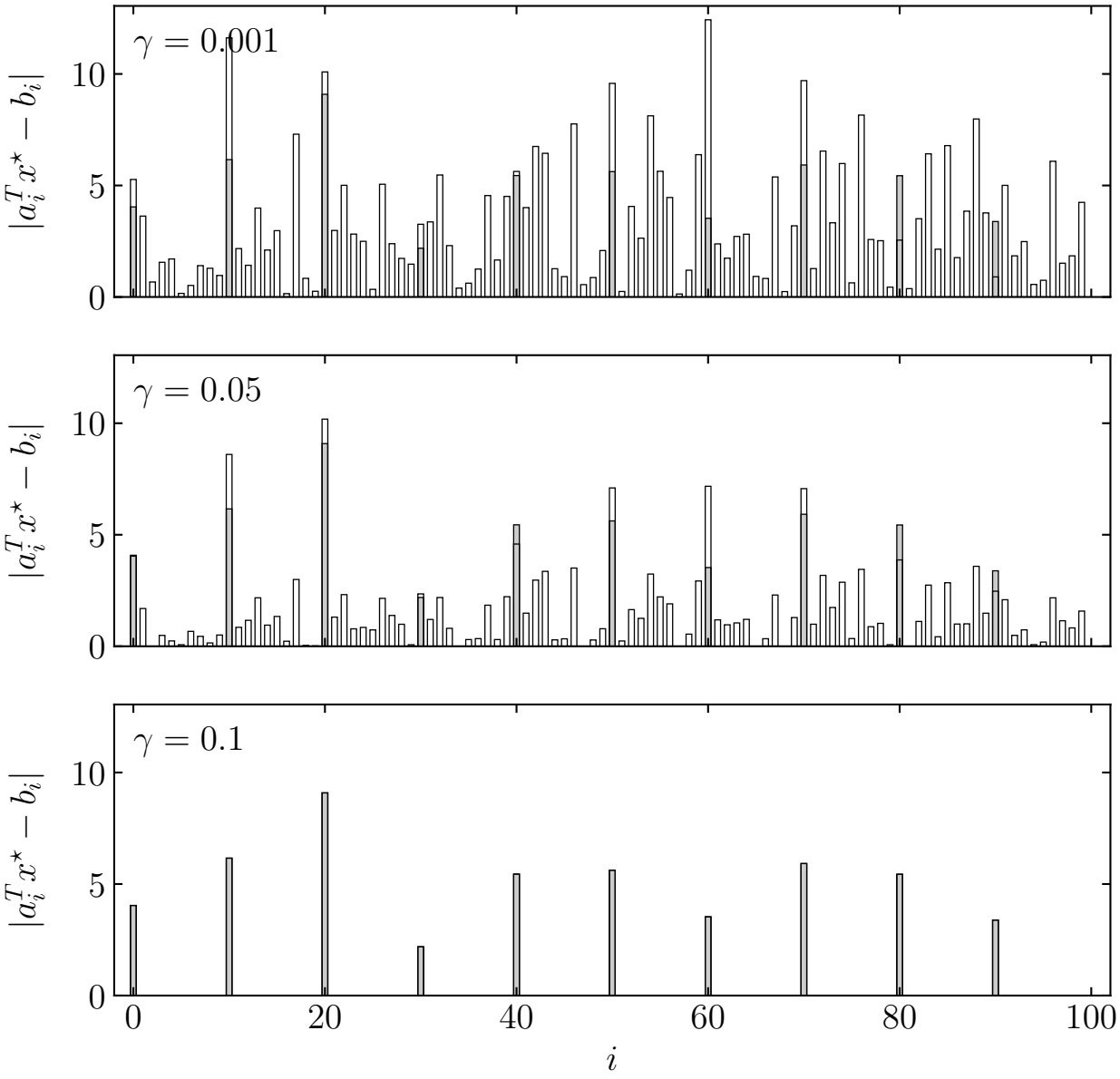
where linear system $Ax = b$ is inconsistent, exact penalty method gives

$$\text{minimize} \quad \|x\|_2^2 + \gamma \|Ax - b\|_1$$

for $A \in \mathbf{R}^{100\times 10}$ with different γ:

- $\|x^\star\|_2^2$ shown solid
- $\|Ax^\star - b\|_1$ dashdotted
- top dashed line shows $\|x^\star\|_2^2$ on largest feasible subset





(violation for solution on largest feasible subset shown darker)

8. Robust models

- stochastic programming
- mean field approximation
- finite event set and sample average approximation
- worst-case optimization
- worst-case robust approximation
- robust linear discrimination
- geometric interpretation and Lagrange duality
- support vector classifiers

Stochastic programming

a **stochastic program** has the form

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

- variable $x \in \mathbf{R}^n$; parameter $\omega \in \mathbf{R}^q$ is random variable
- expectation is taken over ω
- constraints called **stochastic constraints**

convex stochastic program if f_0 and f_i are convex in x for each ω

- a few can be solved efficiently
- in general, evaluating the expectation is difficult

Stochactic least squares

$$\text{minimize} \quad \mathbf{E}\,\|Ax - b\|_2^2$$

- problem data $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$ are random variables
- objective expanded as

$$\begin{aligned}
\mathbf{E}\,\|Ax - b\|_2^2 &= \mathbf{E}\,(Ax - b)^T(Ax - b) \\
&= \mathbf{E}(x^T A^T Ax - 2b^T Ax + b^T b) \\
&= x^T(\mathbf{E}\,A^T A)x - 2(\mathbf{E}\,b^T A)x + \mathbf{E}\,b^T b
\end{aligned}$$

so equivalent to a convex QP if second moments of A and b are known

example: $A = \bar{A} + U$ with U zero mean random, $\bar{A}$, b deterministic

$$\text{minimize} \quad \mathbf{E}\,\|Ax - b\|_2^2 = \|\bar{A}x - b\|_2^2 + x^T(\mathbf{E}\,U^T U)x$$

let $P = \mathbf{E}\, U^T U$ be the covariance of U, then problem equivalent to

$$\text{minimize} \quad \|\bar{A}x - b\|_2^2 + \|P^{1/2}x\|_2^2$$

has analytical solution $x^\star = (\bar{A}^T\bar{A} + P)^{-1}\bar{A}^T b$

- for $U \in \mathbf{R}^{m\times n}$ with IID $\mathcal{N}(0, \sigma^2)$ entries, $P = m\sigma^2 I$

$$x^\star = (\bar{A}^T\bar{A} + m\sigma^2 I)^{-1}\bar{A}^T b$$

  equivalent to Tikhonov regularization with weight $\gamma = m\sigma^2$

Mean field approximation

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

- ignore the randomness in ω and replace it with its mean
- sometimes called *certainty equivalent problem*
- convex if f_i convex in x for each ω

relaxation property: if f_i convex in ω for each x, by Jensen's inequality

$$f_i(x, \mathbf{E}\,\omega) \leq \mathbf{E}\, f_i(x, \omega)$$

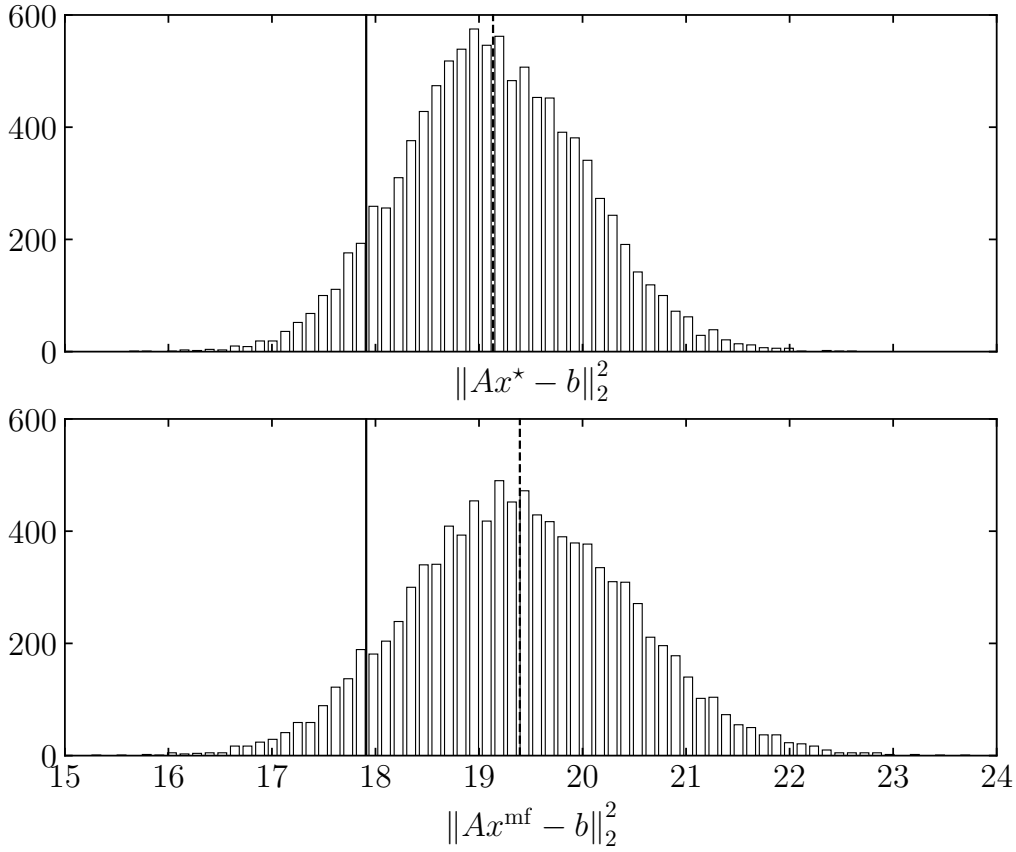
- objective lower bound: $f_0(x, \mathbf{E}\,\omega) \leq \mathbf{E}\, f_0(x, \omega)$
- constraint relaxation: $\mathbf{E}\, f_i(x, \omega) \leq 0 \implies f_i(x, \mathbf{E}\,\omega) \leq 0$, $i = 1, \ldots, m$

example: stochastic least squares

$$\text{minimize} \quad \mathbf{E}\,\|Ax-b\|_2^2 \;, \quad A=\bar{A}+U, \quad U\sim\mathcal{N}(0,m\sigma^2 I)$$

mean field approximation minimizes $\|\bar{A}x-b\|_2^2$

- $x^\star$: stochastic solution
- x^{mf}: mean field solution
- solid line shows mean field objective lower bound



Finite event set

for random parameter $\omega \in \{\omega_1, \ldots, \omega_k\} \subseteq \mathbf{R}^q$ with

$$\mathbf{prob}(\omega = \omega_i) = p_i \quad (p \succeq 0,\ \mathbf{1}^T p = 1)$$

stochastic program reduces to deterministic program

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^k p_i f_0(x, \omega_i) \\ \text{subject to} & \sum_{i=1}^k p_i f_j(x, \omega_i) \leq 0, \quad j = 1, \ldots, m \end{array}$$

example: stochastic linear approximation

$$\text{minimize} \quad \mathbf{E}\, \|Ax - b\|$$

if $A \in \{A_1, \ldots, A_k\}$ with $\mathbf{prob}(A = A_i) = p_i$, reduces to

$$\text{minimize} \quad \sum_{i=1}^k p_i \|A_i x - b\|$$

- for ℓ_1-norm, reduces to LP

$$\begin{array}{ll} \text{minimize (over } x,\, t) & p^T t \\ \text{subject to} & \|A_i x - b\|_1 \leq t_i, \quad i = 1, \ldots, k \end{array}$$

$$\iff \quad \begin{array}{ll} \text{minimize (over } x,\, u_i) & \sum_{i=1}^k p_i \mathbf{1}^T u_i \\ \text{subject to} & -u_i \preceq A_i x - b \preceq u_i, \quad i = 1, \ldots, k \end{array}$$

- for ℓ_∞-norm, reduces to LP

$$\begin{array}{ll} \text{minimize (over } x,\, t) & p^T t \\ \text{subject to} & -t_i \mathbf{1} \preceq A_i x - b \preceq t_i \mathbf{1}, \quad i = 1, \ldots, k \end{array}$$

Sample average approximation

approximate expectation by finite sample average

$$\begin{array}{ll} \text{minimize} & (1/N)\sum_{i=1}^{N} f_0(x,\omega_i) \\ \text{subject to} & (1/N)\sum_{i=1}^{N} f_j(x,\omega_i) \leq 0, \quad j=1,\ldots,m \end{array}$$

- $\omega_1,\ldots,\omega_N$ are IID samples of ω
- solution x^{saa} converges to stochastic solution $x^\star$ as $N \to \infty$
- optimal value p^{saa} lower bounds $p^\star$, tight as $N \to \infty$

example: stochastic ℓ_1-norm approximation

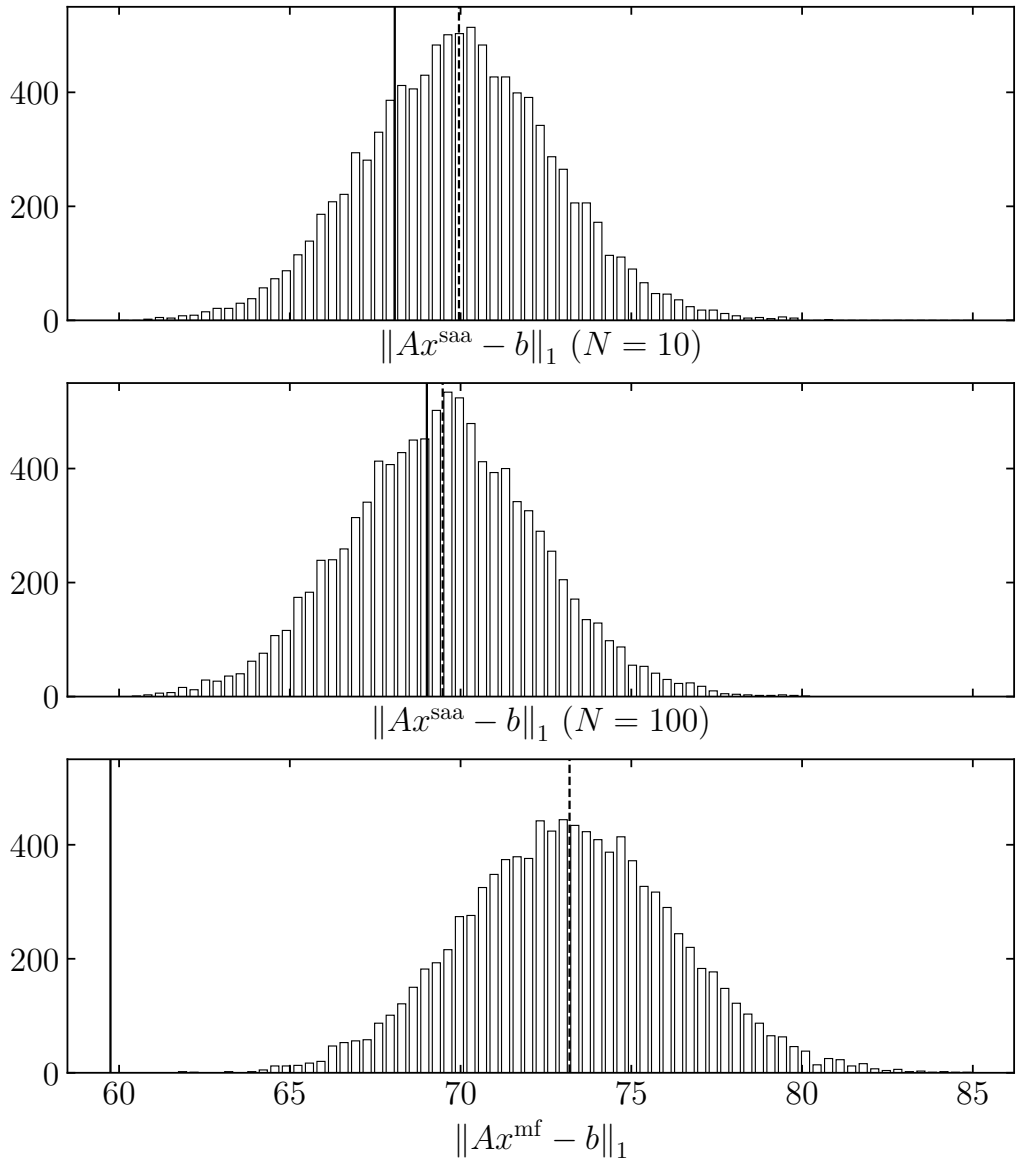
$$\text{minimize} \quad \mathbf{E}\,\|Ax-b\|_1$$

for $A = \bar{A} + U$, $b = \bar{b} + v$ with U, v zero mean IID Gaussian

sample average approximation:

$$\text{minimize} \quad (1/N)\sum_{i=1}^{N} \|A_i x - b_i\|_1$$

optimal value shown solid; mean field solution x^{mf} shown bottom:



Worst-case optimization

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

- variable $x \in \mathbf{R}^n$
- $f_i : \mathbf{R}^n \times \mathbf{R}^q \to \mathbf{R}$, parameterized by $\omega \in U \subseteq \mathbf{R}^q$
- U sometimes called the *uncertainty set*

interpretation: choose x so that for *all* $\omega \in U$

- $f_0(x, \omega)$ is minimized
- $f_i(x, \omega) \leq 0$ for $i = 1, \ldots, m$ are satisfied

convex worst-case optimization if f_i convex in x for each $\omega \in U$

- evaluating the supremum is often difficult; with a few exceptions

example: for finite uncertainty set $U = \{\omega_1, \ldots, \omega_k\}$,

$$\begin{array}{lll} & \text{minimize} & \max_{i=1,\ldots,k} f_0(x, \omega_i) \\ & \text{subject to} & \max_{i=1,\ldots,k} f_j(x, \omega_i) \leq 0, \quad j = 1, \ldots, m \end{array}$$

$$\begin{array}{lll} & \text{minimize} & t \\ \iff & \text{subject to} & f_0(x, \omega_i) \leq t, \quad i = 1, \ldots, k \\ & & f_j(x, \omega_i) \leq 0, \quad i = 1, \ldots, k, \ j = 1, \ldots, m \end{array}$$

- convex program (in x and t) if f_i convex in x for each $\omega \in U$

Worst-case robust approximation

$$\text{minimize} \quad \sup\{\|Ax - b\|_2^2 \mid A \in \mathcal{A}\}$$

- variable $x \in \mathbf{R}^n$
- $A \in \mathcal{A} \subseteq \mathbf{R}^{m \times n}$ is uncertain; assume $b \in \mathbf{R}^m$ fixed

example: for A on a line segment between $A_0 - A_1$ and $A_0 + A_1$, *i.e.*,

$$\mathcal{A} = \{A_0 + uA_1 \mid |u| \leq 1\}$$

- nominal solution ignores uncertainty in A, *i.e.*,

$$\text{minimize} \quad \|A_0 x - b\|_2^2$$

- stochastic solution minimizes expected objective, *i.e.*,

$$\text{minimize} \quad \mathbf{E}\,\|(A_0 + uA_1)x - b\|_2^2 \;, \quad u \sim \mathcal{U}(-1, 1)$$

which is equivalent to

$$\text{minimize} \quad \|A_0 x - b\|_2^2 + \|P^{1/2} x\|_2^2,$$

where $P = \mathbf{E}\,(uA_1)^T(uA_1) = \mathbf{E}\,u^2 A_1^T A_1 = (1/3)A_1^T A_1$

- worst-case objective is

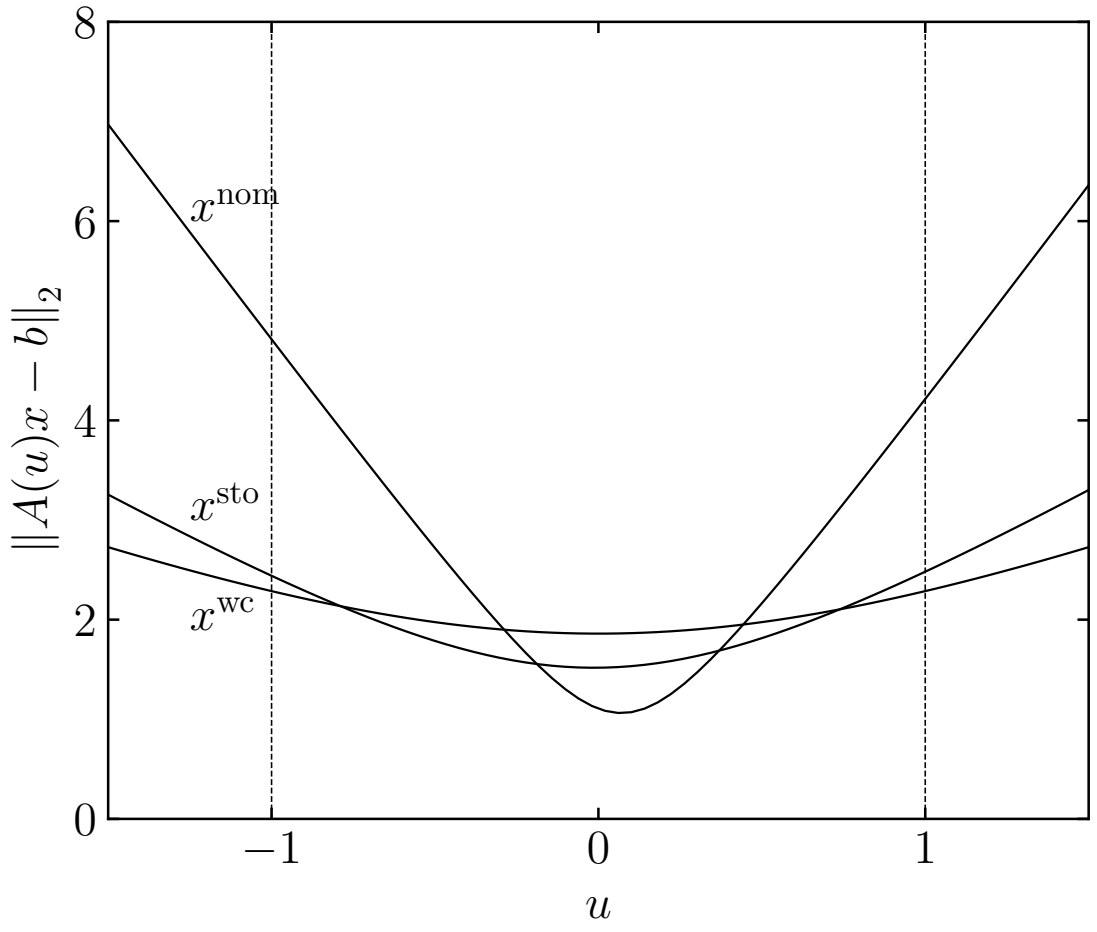
$$\begin{aligned} \sup_{A \in \mathcal{A}} \|Ax - b\|_2^2 &= \sup_{|u| \leq 1} \|(A_0 + uA_1)x - b\|_2^2 \\ &= \max\{\|(A_0 - A_1)x - b\|_2^2, \|(A_0 + A_1)x - b\|_2^2\} \end{aligned}$$

(since $\|(A_0 + uA_1)x - b\|_2^2$ is convex in u), so we have the problem

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \|(A_0 - A_1)x - b\|_2^2 \leq t \\ & \|(A_0 + A_1)x - b\|_2^2 \leq t \end{array}$$

x^{nom}, x^{sto}, x^{wc}: nominal, stochastic, and worst-case solutions

evaluate $\|A(u)x - b\|_2$ for $u \in [-1, 1]$:



Robust linear discrimination

linear discrimination problem: given two groups of points $\{x_1, \ldots, x_M\}$ and $\{y_1, \ldots, y_N\}$ in $\mathbf{R}^n$,

$$\begin{array}{ll} \text{find} & (a, b) \\ \text{subject to} & a^T x_i - b \geq 1, \quad i = 1, \ldots, M \\ & a^T y_i - b \leq -1, \quad i = 1, \ldots, N \end{array}$$

- solution set is polyhedron

$$\left\{ z \in \mathbf{R}^{n+1} \;\middle|\; \begin{array}{l} z^T \tilde{x}_i \geq 1, \quad i = 1, \ldots, M \\ z^T \tilde{y}_i \leq -1, \quad i = 1, \ldots, N \end{array} \right\}$$

where

$$z = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \tilde{x}_i = \begin{bmatrix} x_i \\ -1 \end{bmatrix}, \quad \tilde{y}_i = \begin{bmatrix} y_i \\ -1 \end{bmatrix}$$

robust linear discrimination:

$$\begin{array}{ll} \text{minimize} & \|a\|_2 \\ \text{subject to} & a^T x_i - b \geq 1, \quad i = 1, \ldots, M \\ & a^T y_i - b \leq -1, \quad i = 1, \ldots, N \end{array}$$

- find normal vector a of separating hyperplane with least norm

Geometric interpretation

change of variables $a = \tilde{a}/t$, $b = \tilde{b}/t$ with $t > 0$ gives

$$\begin{array}{ll} \text{minimize} & \|\tilde{a}\|_2/t \\ \text{subject to} & \tilde{a}^T x_i - \tilde{b} \geq t, \quad i = 1, \ldots, M \\ & \tilde{a}^T y_i - \tilde{b} \leq -t, \quad i = 1, \ldots, N \\ & t > 0 \end{array}$$

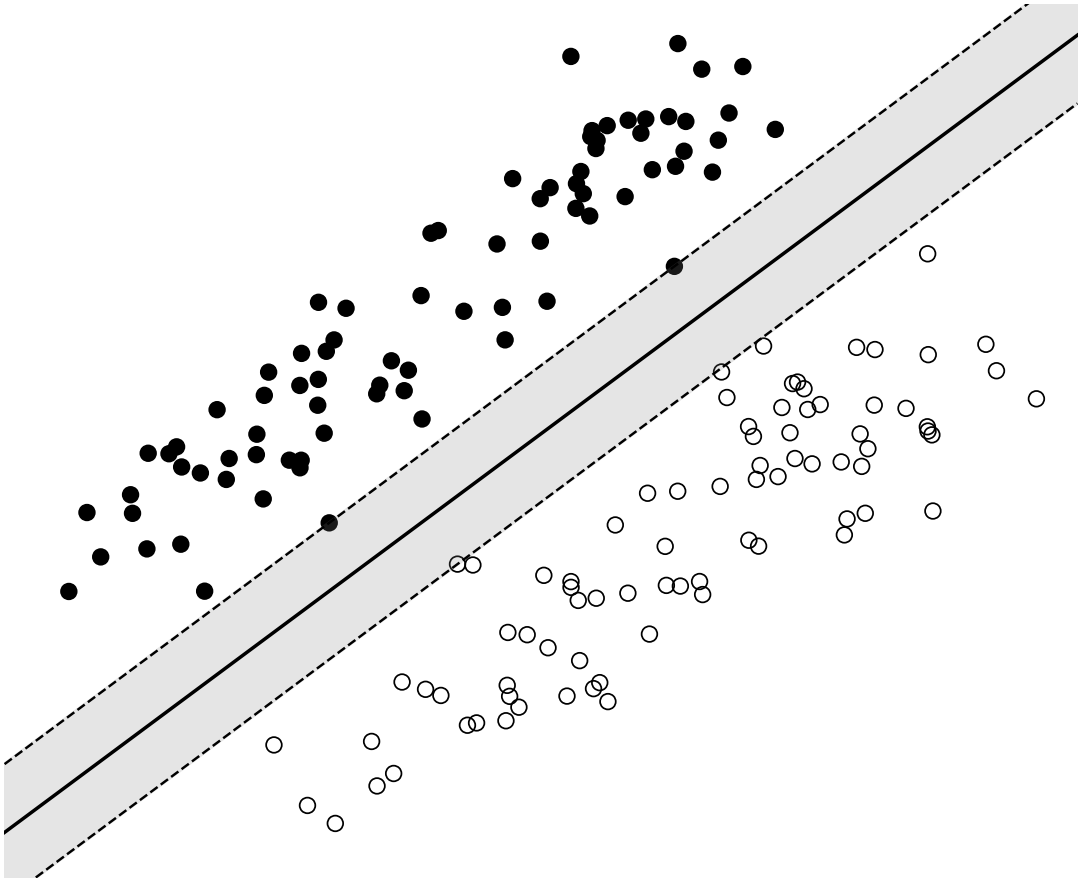
then to remove ambiguity, we bound $\|\tilde{a}\|_2 \leq 1$ and maximize t:

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & \tilde{a}^T x_i - \tilde{b} \geq t, \quad i = 1, \ldots, M \\ & \tilde{a}^T y_i - \tilde{b} \leq -t, \quad i = 1, \ldots, N \\ & t > 0, \quad \|\tilde{a}\|_2 \leq 1 \end{array}$$

interpretation: maximize the distance to the closest point in either group

for optimal point $(\tilde{a}^\star, \tilde{b}^\star, t^\star)$,

- $\|\tilde{a}^\star\|_2 = 1$ (by homogeneity of constraints)
- $2t^\star$ is the (thickest) separating slab width between two groups of points



Interpretation via Lagrange duality

primal problem of robust linear discrimination:

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & \tilde{a}^T x_i - \tilde{b} \geq t, \quad i = 1, \ldots, M \\ & \tilde{a}^T y_i - \tilde{b} \leq -t, \quad i = 1, \ldots, N \\ & \|\tilde{a}\|_2 \leq 1 \end{array}$$

- $t^\star > 0 \iff$ two groups of points are linearly separable

Lagrangian:

$$\begin{aligned} & L((\tilde{a}, \tilde{b}, t), (u, v, w)) \\ = \; & -t + \textstyle\sum_{i=1}^M u_i (t - \tilde{a}^T x_i + \tilde{b}) + \sum_{i=1}^N v_i (t + \tilde{a}^T y_i - \tilde{b}) + w(\|\tilde{a}\|_2 - 1) \\ = \; & (\mathbf{1}^T u + \mathbf{1}^T v - 1)t + (\mathbf{1}^T u - \mathbf{1}^T v)\tilde{b} \\ & -\tilde{a}^T \left(\textstyle\sum_{i=1}^M u_i x_i - \sum_{i=1}^N v_i y_i \right) + w(\|\tilde{a}\|_2 - 1) \end{aligned}$$

dual function $g(u,v,w) = \inf_{(\tilde{a},\tilde{b},t)} L((\tilde{a},\tilde{b},t),(u,v,w))$ bounded iff

$$\mathbf{1}^T u + \mathbf{1}^T v = 1, \quad \mathbf{1}^T u = \mathbf{1}^T v \quad \Longleftrightarrow \quad \mathbf{1}^T u = 1/2, \quad \mathbf{1}^T v = 1/2$$

in this case, we have

$$\begin{aligned} g(u,v,w) &= \inf_{\tilde{a}} \left(-\tilde{a}^T \left(\sum_{i=1}^M u_i x_i - \sum_{i=1}^N v_i y_i \right) + w(\|\tilde{a}\|_2 - 1) \right) \\ &= -w + \inf_{\tilde{a}} \left(-\tilde{a}^T \left(\sum_{i=1}^M u_i x_i - \sum_{i=1}^N v_i y_i \right) + w\|\tilde{a}\|_2 \right) \\ &\geq -w + \inf_{\tilde{a}} \left(\|\tilde{a}\|_2 \left(w - \left\| \sum_{i=1}^M u_i x_i - \sum_{i=1}^N v_i y_i \right\|_2 \right) \right) \end{aligned}$$

inequality follows from Cauchy-Schwarz, so

$$g(u,v,w) = \begin{cases} -w, & w \geq \left\| \sum_{i=1}^M u_i x_i - \sum_{i=1}^N v_i y_i \right\|_2 \\ -\infty, & \text{otherwise} \end{cases}$$

dual problem of robust linear discrimination:

$$\begin{array}{ll} \text{maximize} & -w \\ \text{subject to} & w \geq \left\| \sum_{i=1}^M u_i x_i - \sum_{i=1}^N v_i y_i \right\|_2 \\ & \mathbf{1}^T u = 1/2, \quad \mathbf{1}^T v = 1/2 \\ & u \succeq 0, \quad v \succeq 0 \end{array}$$

is the epigraph form of

$$\begin{array}{ll} \text{minimize} & \left\| \sum_{i=1}^M u_i x_i - \sum_{i=1}^N v_i y_i \right\|_2 \\ \text{subject to} & \mathbf{1}^T u = 1/2, \quad \mathbf{1}^T v = 1/2 \\ & u \succeq 0, \quad v \succeq 0 \end{array}$$

interpretation: find closest points in convex hulls of two groups, *i.e.*,

$$\begin{array}{ll} \text{minimize} & (1/2)\|x - y\|_2 \\ \text{subject to} & x \in \mathbf{conv}\{x_1, \ldots, x_M\} \\ & y \in \mathbf{conv}\{y_1, \ldots, y_N\} \end{array}$$

Support vector classifiers

for two groups of points $\{x_1, \ldots, x_M\}$ and $\{y_1, \ldots, y_N\}$ in $\mathbf{R}^n$ that are not linearly separable, consider relaxation

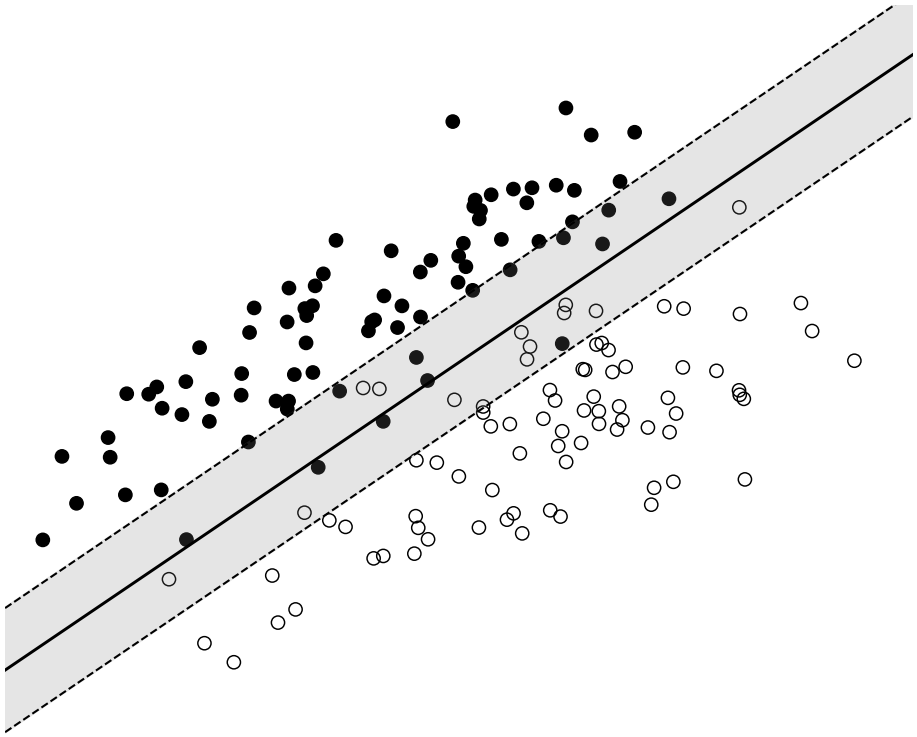
$$\begin{array}{ll} \text{find} & (a, b, u, v) \\ \text{subject to} & a^T x_i - b \geq 1 - u_i, \quad i = 1, \ldots, M \\ & a^T y_i - b \leq -1 + v_i, \quad i = 1, \ldots, N \\ & u \succeq 0, \quad v \succeq 0 \end{array}$$

- slack variables u, v represent margin violations for each point
- always feasible, but may have $u_i, v_i > 0$ for some i
  - within margin, but not misclassified if $0 < u_i, v_i < 1$
  - misclassification if $u_i, v_i \geq 1$
- linearly separable if exist optimal point with $u = v = 0$

Margin violation penalties

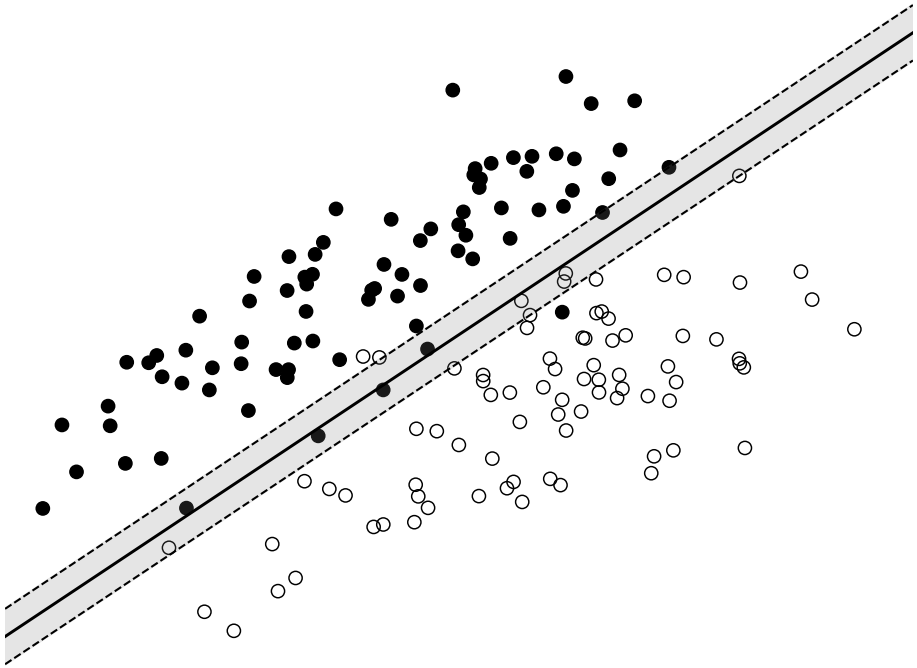
least squares penalty:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{M} u_i^2 + \sum_{i=1}^{N} v_i^2 \\ \text{subject to} & a^T x_i - b \geq 1 - u_i, \quad i = 1, \ldots, M \\ & a^T y_i - b \leq -1 + v_i, \quad i = 1, \ldots, N \\ & u \succeq 0, \quad v \succeq 0 \end{array}$$



sparse margin violations: heuristic for minimizing $\mathbf{card}\, u + \mathbf{card}\, v$,

$$
\begin{array}{ll}
\text{minimize} & \mathbf{1}^T u + \mathbf{1}^T v \\
\text{subject to} & a^T x_i - b \geq 1 - u_i, \quad i = 1, \ldots, M \\
& a^T y_i - b \leq -1 + v_i, \quad i = 1, \ldots, N \\
& u \succeq 0, \quad v \succeq 0
\end{array}
$$



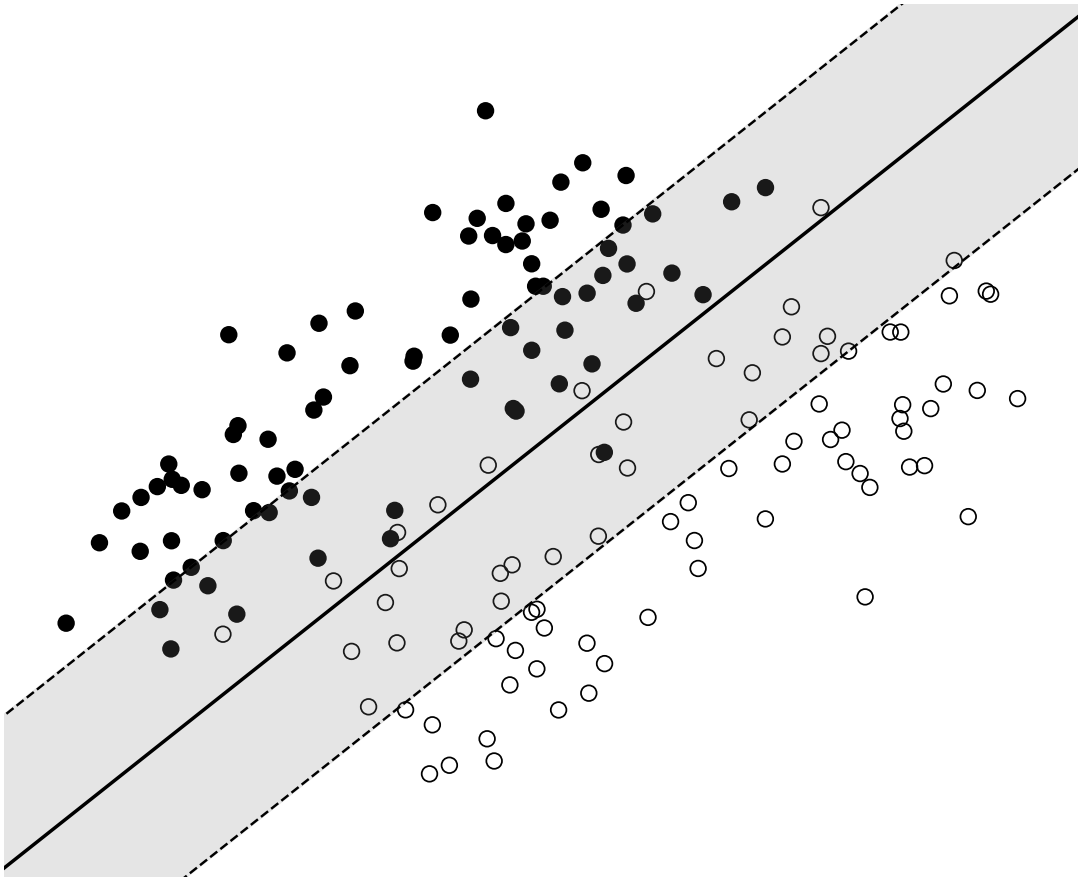
Standard form support vector classifier

trade-off between margin width and margin violation penalty:

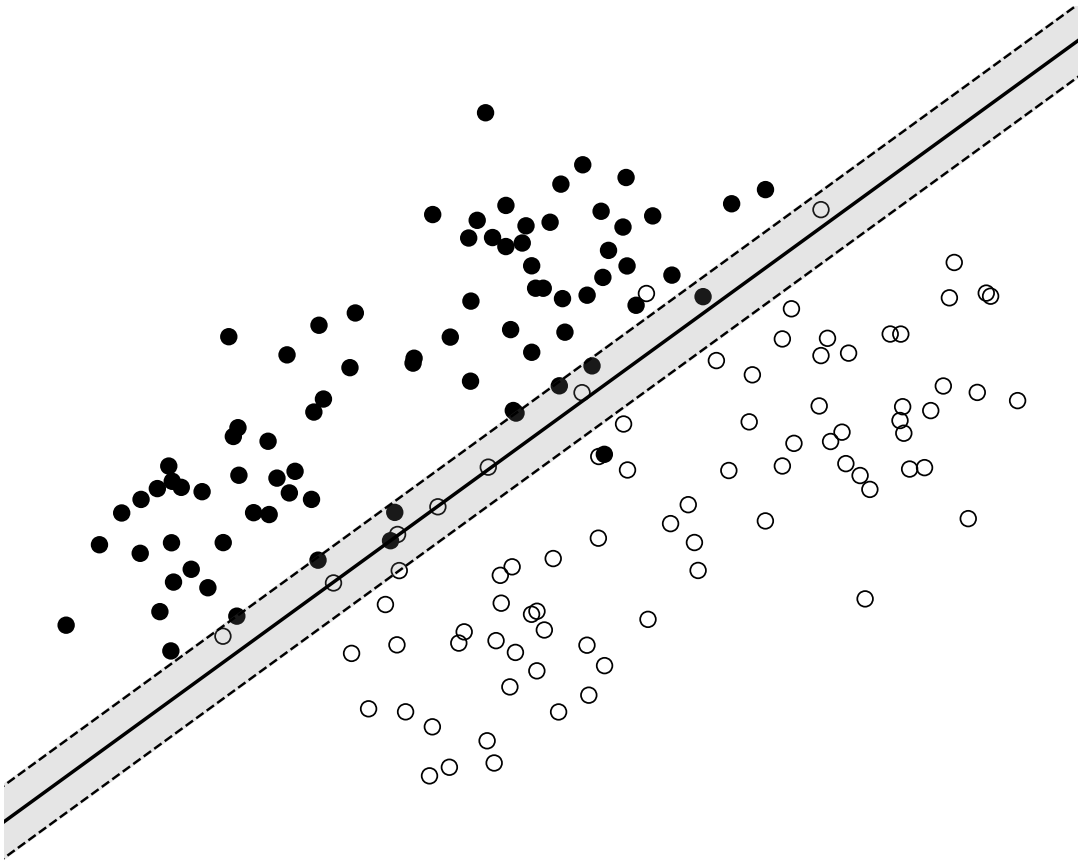
$$\begin{array}{ll} \text{minimize} & \|a\|_2 + \gamma(\mathbf{1}^T u + \mathbf{1}^T v) \\ \text{subject to} & a^T x_i - b \geq 1 - u_i, \quad i = 1, \ldots, M \\ & a^T y_i - b \leq -1 + v_i, \quad i = 1, \ldots, N \\ & u \succeq 0, \quad v \succeq 0 \end{array}$$

- variables $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, $u \in \mathbf{R}^M$, $v \in \mathbf{R}^N$
- $\gamma > 0$ is regularization weight
- separating slab width is $2/\|a\|_2$
- equals to feasibility problem as $\gamma \to \infty$ (no margin violations allowed)

for $\gamma = 0.01$:



for $\gamma = 1$:



9. Latent factor estimation

- mixture models and biconvex formulation
- cost functions, regularization, and constraints
- clustering with prior information
- principal component analysis
- solution and interpretations
- quadratic regularization
- robust low rank approximation
- structural regularization and constraints

Mixture models

inverse problem for general machine learning models:

$$\text{minimize} \quad \sum_{i=1}^m f_\theta(x_i, y_i)$$

- $(x_i, y_i) \in \mathbf{R}^n \times \mathbf{R}$ are data points
- variable $\theta \in \mathbf{R}^n$ are model parameters
- $f_\theta \colon \mathbf{R}^n \times \mathbf{R} \to \mathbf{R}$ is some cost function

mixture model assumes data generated by k different models,

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m \sum_{j=1}^k z_{ij} f_{\theta_j}(x_i, y_i) \\ \text{subject to} & z_i \in \{e_1, \ldots, e_k\}, \quad i = 1, \ldots, m \end{array}$$

- variables $\theta_1, \ldots, \theta_k \in \mathbf{R}^n$ and $z_1, \ldots, z_m \in \mathbf{R}^k$
- $z_i \in \mathbf{R}^k$ are *one-hot* labels indicating which of the k models generated the ith data point; sometimes called **latent factors**

rewrite the problem as

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m z_i^T r_i \\ \text{subject to} & r_{ij} = f_{\theta_j}(x_i, y_i), \quad i = 1, \ldots, m, \quad j = 1, \ldots, k \\ & z_i \in \{0,1\}^k, \quad \mathbf{card}\, z_i = 1, \quad i = 1, \ldots, m \end{array}$$

- $r_i \in \mathbf{R}^k$ are auxiliary, denote the vector of costs
- last constraint requires z_i to be one-hot

interpretation:

- first separate m data points into k groups, each corresponding to a different model, according to the latent factors z_i
- fit parameters θ_j of each model to the data in different groups

however, solving the inverse problem is very difficult (actually, NP-hard)

Biconvex formulation

relaxation of integer constraints on z_i to convex constraints:

$$0 \preceq z_i \preceq \mathbf{1}, \quad \mathbf{1}^T z_i = 1, \quad i = 1, \ldots, m$$

so we have **biconvex** inverse problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m z_i^T r_i \\ \text{subject to} & r_{ij} = f_{\theta_j}(x_i, y_i), \quad i = 1, \ldots, m, \quad j = 1, \ldots, k \\ & 0 \preceq z_i \preceq \mathbf{1}, \quad \mathbf{1}^T z_i = 1, \quad i = 1, \ldots, m \end{array}$$

- variables $\theta_1, \ldots, \theta_k$ and $z_1, \ldots, z_m$
- convex in θ_j for fixed z_i, and convex in z_i for fixed θ_j
- can be (approximately) solved via alternate convex search
- relaxation is often tight, *i.e.*, the solution z_i is often one-hot

interpretation:

- latent factors z_i lie in the probability simplex

$$\{z \in \mathbf{R}^k \mid 0 \preceq z \preceq \mathbf{1}, \quad \mathbf{1}^T z = 1\} = \mathbf{conv}\{e_1, \ldots, e_k\}$$

  (whose vertices are one-hot vectors)

- relax 'hard' labels to 'soft' labels, $i.e.$, probabilities of each data point belonging to each model

- objective is the *expected* cost of the data points under the soft labels

in ACS iterations, we alternate between

- fit θ_j given data points assigned to each model according to $\tilde{z}_i$

- find best label assignments z_i for each data point given $\tilde{\theta}_j$

Cost functions

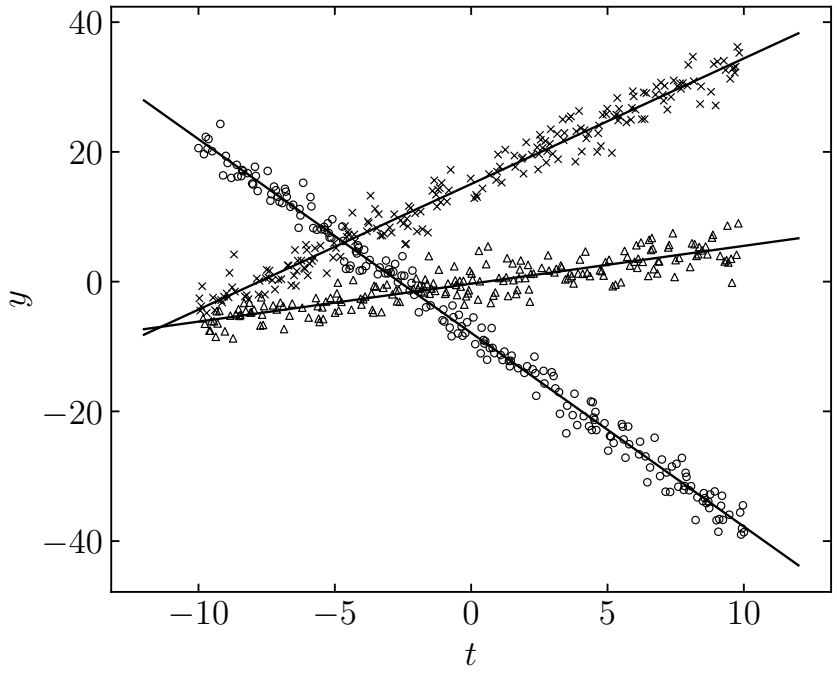
regression models:

$$f_{\theta_i}(x, y) = \phi(x^T\theta_i - y), \quad i = 1, \ldots, k$$

penalty function ϕ: $\mathbf{R} \to \mathbf{R}$ can be quadratic, ℓ_p-norm, Huber, etc.

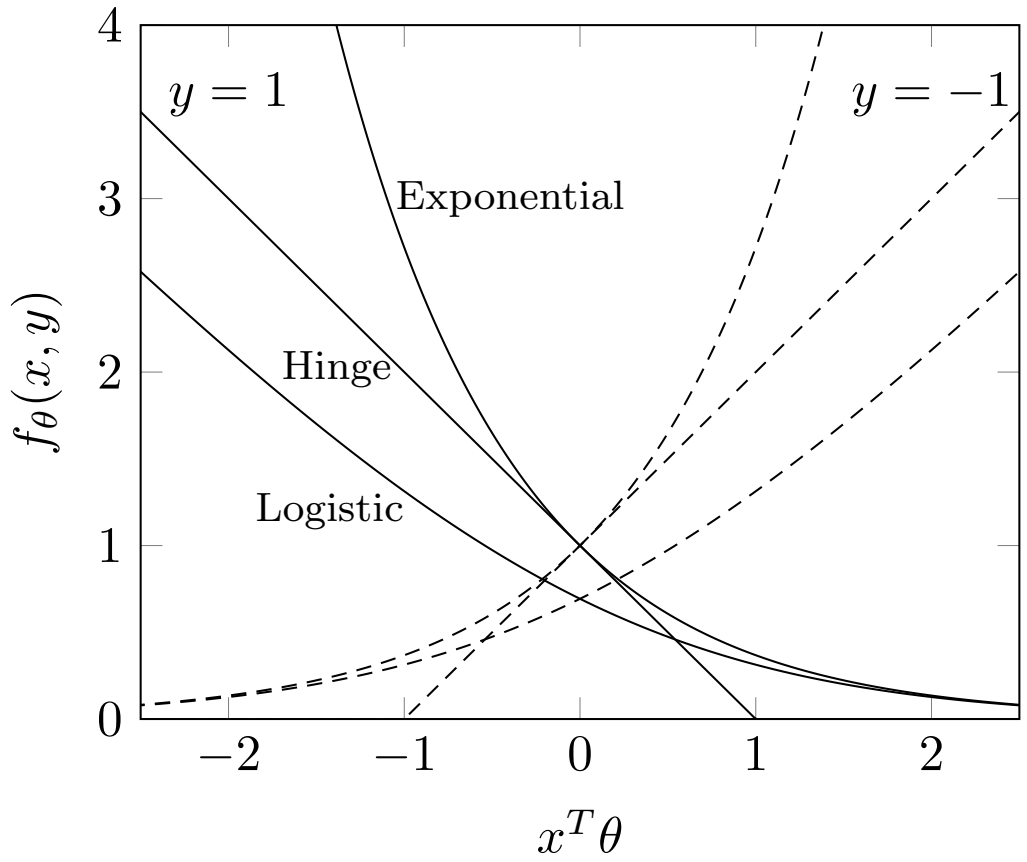
example: mixture of linear regressions

$$\begin{array}{ll} \text{min.} & \sum_{i=1}^m z_i^T r_i \\ \text{s.t.} & r_{ij} = \left(x_i^T\theta_j - y_i\right)^2 \\ & z_i \in \{0,1\}^k, \ \mathbf{card}\, z_i = 1 \\ & i = 1, \ldots, m, \ j = 1, \ldots, k \end{array}$$



classification models:

- logistic regression: $f_{\theta_i}(x, y) = \log(1 + \exp(-yx^T\theta_i))$
- support vector classifier: $f_{\theta_i}(x, y) = \max\{0,\ 1 - yx^T\theta_i\}$
- exponential loss: $f_{\theta_i}(x, y) = \exp(-yx^T\theta_i)$



- multiclass logistic regression: given $(X^{(i)}, y^{(i)})$, $i = 1, \ldots, m$, where

$$X^{(i)} = \left[\begin{array}{ccc} x_1^{(i)} & \cdots & x_q^{(i)} \end{array} \right]^T \in \mathbf{R}^{q \times n}, \qquad y^{(i)} \in \{e_1, \ldots, e_q\} \subseteq \mathbf{R}^q$$

  – $x_j^{(i)} \in \mathbf{R}^n$: feature vector contributing to the jth class

  cost function given by

$$f_{\theta_i}(X, y) = -\log(y^T u / \mathbf{1}^T u), \qquad u = \left[\begin{array}{c} \exp(x_1^T \theta_i) \\ \vdots \\ \exp(x_q^T \theta_i) \end{array} \right]$$

Regularization and constraints

probabilistic properties in z_i: $e.g.$, maximum entropy regularizer

$$\gamma \sum_{i=1}^{m} \sum_{j=1}^{k} z_{ij} \log z_{ij}$$

encourages stochastic z_i, $i.e.$, avoid hard model assignments

parameter alignment: positivity, monotonicity, etc. for different θ_j

time series models: total variation regularizer on $z(t)$ for $t = 1, \ldots, m$,

$$\gamma \sum_{t=1}^{m-1} \|z(t+1) - z(t)\|_1$$

encourages piecewise constant $z(t)$, $i.e.$, the model assignment does not change too often

example: hierarchical logistic regression for time series,

$$
\begin{array}{ll}
\text{minimize} & \sum_{t=1}^{m} z(t)^T r(t) + \gamma \sum_{t=1}^{m-1} \|z(t+1) - z(t)\|_1 \\
\text{subject to} & r_i(t) = \log(1 + \exp(-y(t)x(t)^T \theta_i)), \\
& \qquad\qquad t = 1, \ldots, m, \quad i = 1, 2 \\
& 0 \preceq z(t) \preceq \mathbf{1}, \quad \mathbf{1}^T z(t) = 1, \quad t = 1, \ldots, m \\
& \theta_1 \succeq 0, \quad \theta_{1,1} \geq \cdots \geq \theta_{1,n} \\
& \theta_2 \preceq 0, \quad \theta_{2,1} \leq \cdots \leq \theta_{2,n}
\end{array}
$$

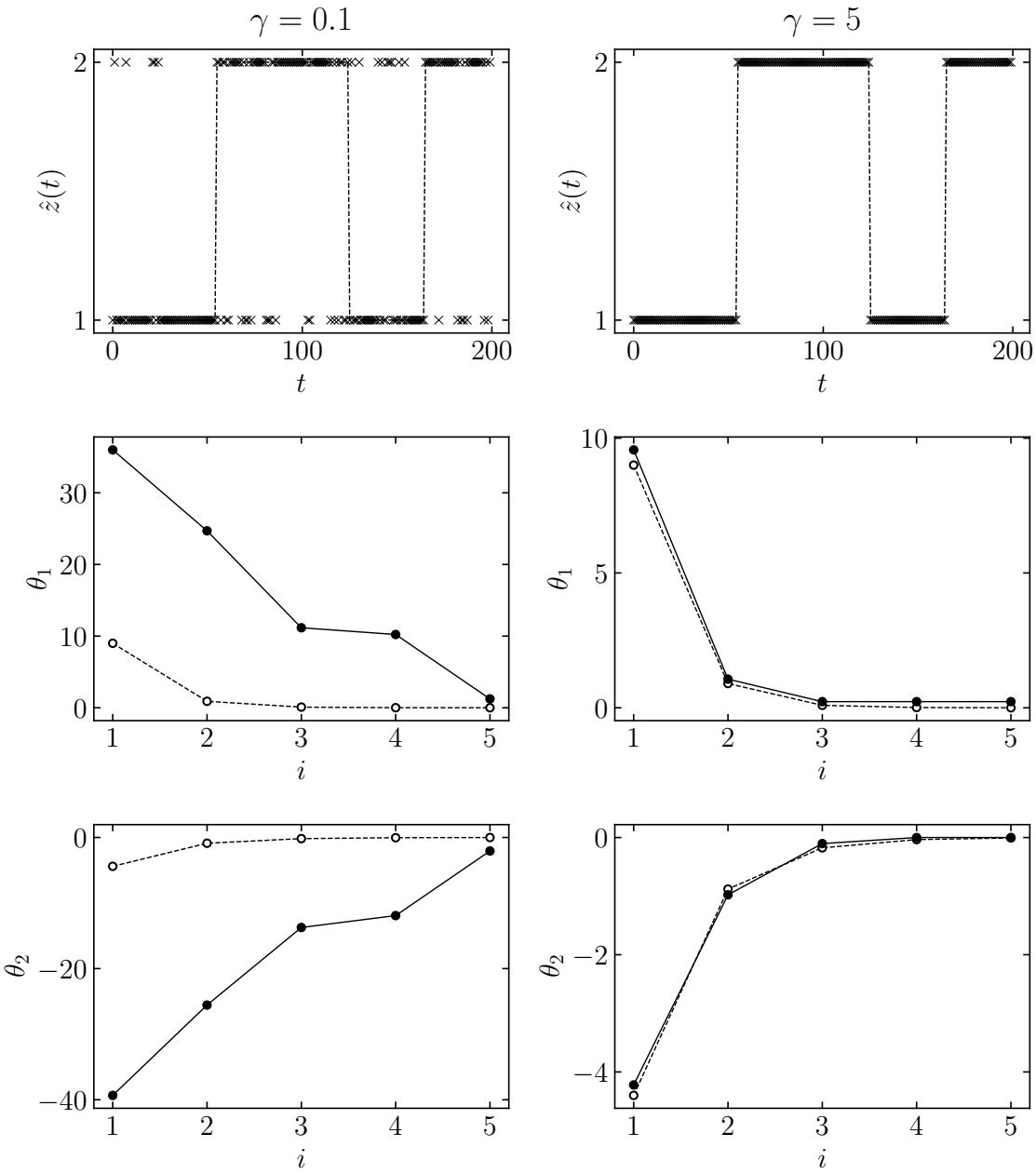
- mixture of two logistic regression models for time series data
- encourage piecewise constant latent factors $z(t)$
- θ_1 nonnegative and nonincreasing
- θ_2 nonpositive and nondecreasing

numerical results:

- one-hat label recovery:

$$\hat{z}(t) = \underset{i \in \{1,2\}}{\operatorname{argmax}}\, z_i(t)$$

- ground truth dashed



Clustering

$$\text{minimize} \quad \sum_{i=1}^{m} \min \left\{ \|x_i - \mu_j\|_2^2 \;\middle|\; j = 1, \ldots, k \right\}$$

- variables $\mu_1, \ldots, \mu_k \in \mathbf{R}^n$ denote cluster representatives
- data points $x_1, \ldots, x_m \in \mathbf{R}^n$ are given
- minimize the distance from each data point to its closest representative
- equivalent to

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{m} \|x_i - \mu_{c_i}\|_2^2 \\ \text{subject to} & c \in \{1, \ldots, k\}^m \end{array}$$

  with variables

  - $\mu_1, \ldots, \mu_k \in \mathbf{R}^n$
  - $c \in \mathbf{R}^m$ denoting cluster assignments

mixture model interpretation:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^m z_i^T r_i \\ \text{subject to} & r_{ij} = \|x_i - \mu_j\|_2^2, \quad i = 1, \ldots, m, \quad j = 1, \ldots, k \\ & z_i \in \{0,1\}^k, \quad \mathbf{card}\, z_i = 1, \quad i = 1, \ldots, m \end{array}$$

- variables $\mu_1, \ldots, \mu_k \in \mathbf{R}^n$ and $z_1, \ldots, z_m \in \mathbf{R}^k$

k-**means algorithm**: biconvex relaxation and ACS for clustering problem

---

k-means algorithm.

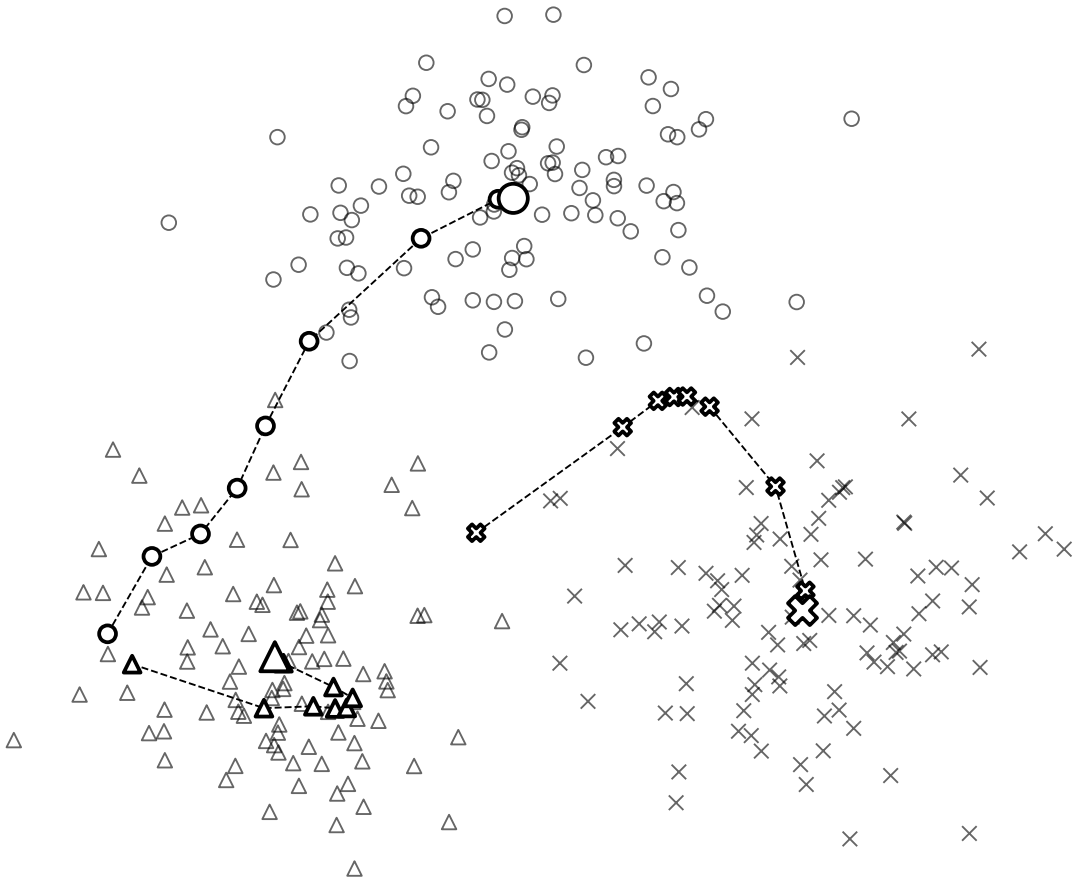
given $x_1, \ldots, x_m \in \mathbf{R}^n$; initial $\tilde{\mu}_1, \ldots, \tilde{\mu}_k \in \mathbf{R}^n$.
repeat

1. *Cluster assignment.* $\tilde{c}_i := \mathrm{argmin}_{j \in \{1,\ldots,k\}} \|x_i - \tilde{\mu}_j\|_2^2$ for $i = 1, \ldots, m$.
2. *Group formation.* $G_j := \{i \in \{1, \ldots, m\} \mid \tilde{c}_i = j\}$ for $j = 1, \ldots, k$.
3. *Representative update.* $\tilde{\mu}_j := (1/|G_j|) \sum_{i \in G_j} x_i$ for $j = 1, \ldots, k$.

until convergence.

---

numerical example:



- path shows results from each iteration
- final cluster representatives shown larger

Clustering with prior information

assignment and relation constraints:

- x_i must in one of the clusters in $S_i \subseteq \{1, \ldots, k\}$:

$$z_{ij} = 0, \quad j \in \{1, \ldots, k\} \setminus S_i$$

- *must-link* constraint: x_p, x_q must in the same cluster,

$$z_p = z_q \quad \Longleftrightarrow \quad z_{pj} = z_{qj}, \quad j = 1, \ldots, k$$

- *cannot-link* constraint: x_p, x_q must in different clusters,

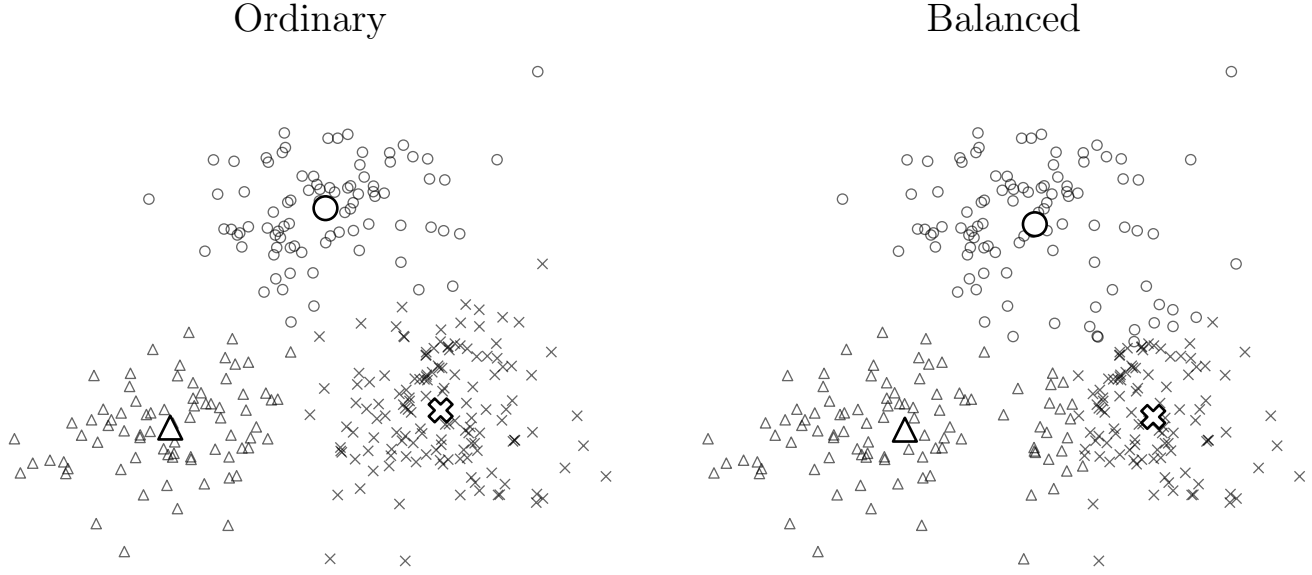
$$z_{pj} + z_{qj} \leq 1, \quad j = 1, \ldots, k$$

  requires only one of z_{pj} and z_{qj} can be nonzero

size constraints: define size measurement $s \in \mathbf{R}^k$ as

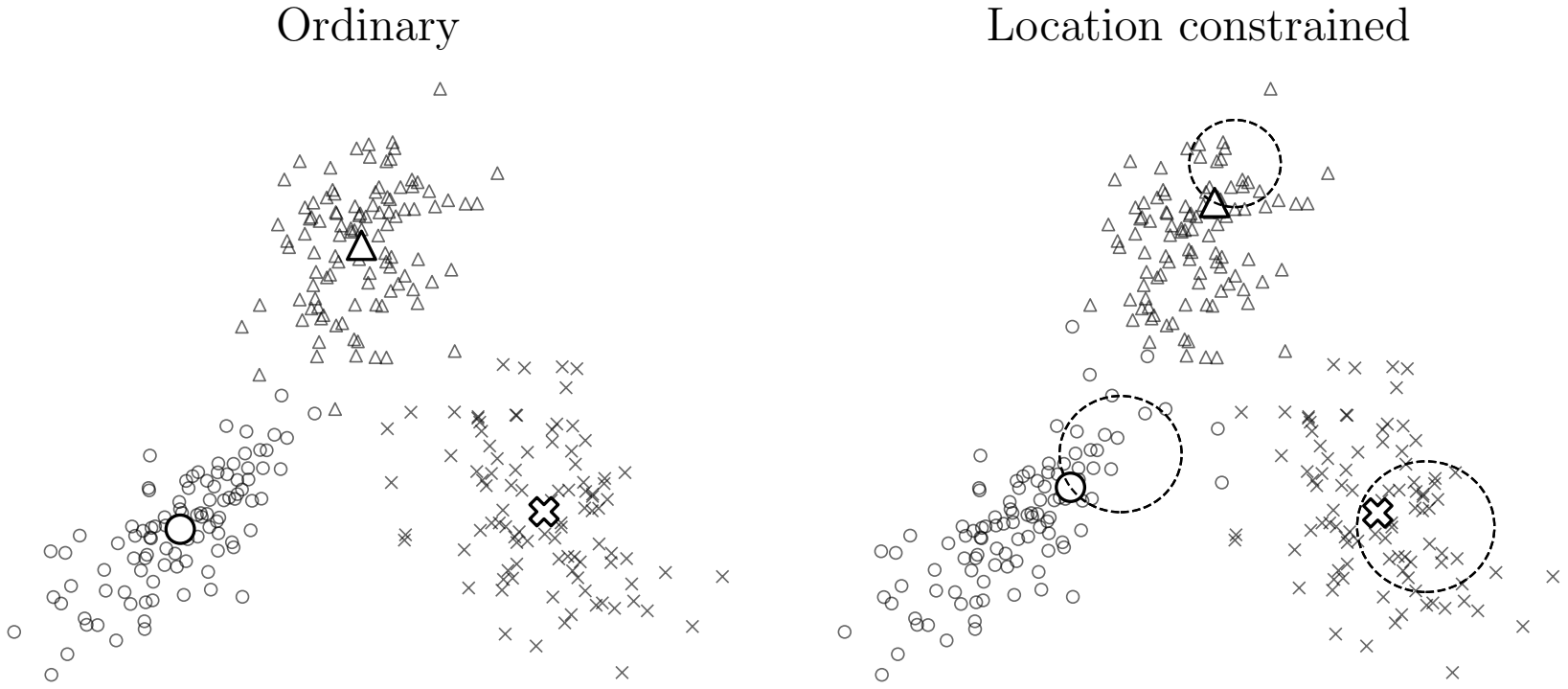
$$s_j = \sum_{i=1}^{m} z_{ij}, \quad j = 1, \ldots, k$$

- bounded size for each cluster: $L_j \leq s_j \leq U_j$, $j = 1, \ldots, k$
- balance constraint: $|s_j - m/k| \leq \epsilon$ for $j = 1, \ldots, k$ (with small $\epsilon > 0$) requires each cluster to have roughly the same number of points



geometric constraints on cluster representatives μ_i:

- polyhedron: $A_i\mu_i \preceq b_i$, $i = 1, \ldots, k$
- norm ball with center c_i and radius r_i: $\|\mu_i - c_i\|_2 \leq r_i$, $i = 1, \ldots, k$



Principal component analysis (PCA)

data matrix $A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \in \mathbf{R}^{m \times n}$:

- rows $a_i \in \mathbf{R}^n$: (feature vector of) ith data point
- jth column is jth feature across all m data points
- often assume $m \gg n$, *i.e.*, more data points than features

standardization:

$$\tilde{A} = (A - \mathbf{1}\mu^T)D^{-1}$$

- μ_i is mean of ith feature for $i = 1, \ldots, n$
- $D = \mathbf{diag}(\sigma_1, \ldots, \sigma_n)$ with σ_i the standard deviation of ith feature
- often assume A is already standardized, *i.e.*, $\mu = 0$ and $D = I$

PCA finds **low rank approximation** Z of A:

$$\begin{array}{ll} \text{minimize} & \|A - Z\|_F^2 \\ \text{subject to} & \mathbf{rank}\, Z \leq k \end{array}$$

- componentwise quadratic residual matrix penalty:

$$\|A - Z\|_F^2 = \sum_{i=1}^{m} \|a_i - z_i\|_2^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij} - z_{ij})^2$$

- equivalent to **factorized form**:

$$\text{minimize} \quad \|A - XY\|_F^2$$

  by change of variables $Z = XY$ with $X \in \mathbf{R}^{m \times k}$ and $Y \in \mathbf{R}^{k \times n}$

  – implicit rank constraint since $\mathbf{rank}(XY) \leq k$

Solution via singular value decomposition

PCA is nonconvex problem but has analytical solution via SVD:

$$A = U\Sigma V^T$$

- $U \in \mathbf{R}^{m\times r}$, $V \in \mathbf{R}^{n\times r}$ have orthonormal columns
- $\Sigma = \mathbf{diag}(\sigma_1, \ldots, \sigma_r)$ with singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$ of A
- PCA solution is then given by

$$Z^\star = U_k\Sigma_k V_k^T$$

  with U_k, Σ_k, V_k the first k columns of U, Σ, V

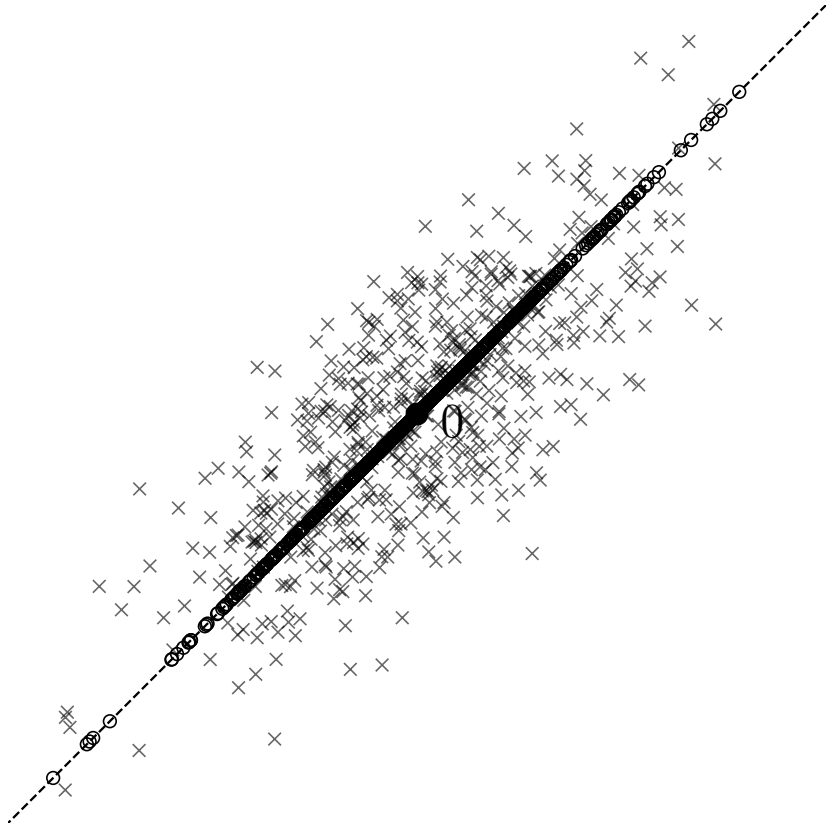
- for factorized form, *a* solution is given by

$$X^\star = U_k\Sigma_k^{1/2}, \qquad Y^\star = \Sigma_k^{1/2}V_k^T$$

  – for any invertible $W \in \mathbf{R}^{k\times k}$, $(X^\star W,\ W^{-1}Y^\star)$ is also a solution

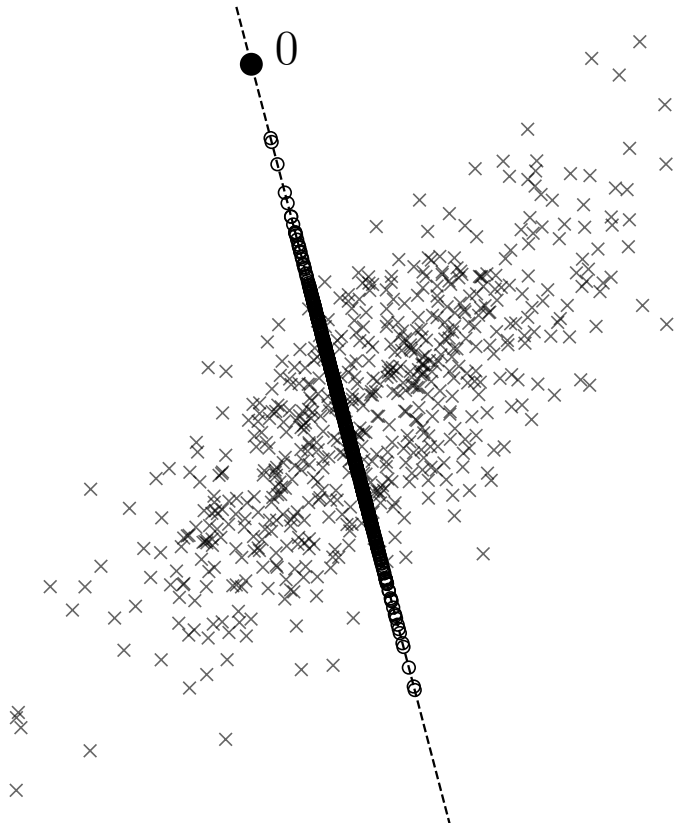
Interpretations

geometric interpretation: objective $\|A - Z\|_F^2 = \sum_{i=1}^m \|a_i - z_i\|_2^2$ finds z_i in a k-dimensional subspace of $\mathbf{R}^n$ that has the smallest sum of squared distances to the data points a_i



- A must be centered, $i.e.$, each column has zero mean
- due to projection only onto some subspace, not any affine set

example of PCA on a noncentered data matrix:



statistical interpretation:

- assuming a_i have zero mean and unit variance (so A standardized)
- find k orthogonal directions in $\mathbf{R}^n$
- such that projections of data points onto these directions have the top k largest variance

MLE interpretation: assuming components a_{ij} of A generated from

$$a_{ij} = x_i^T y_j + \epsilon_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$$

- coefficients $x_i, y_j \in \mathbf{R}^k$ and noise ϵ_{ij} are unknown
- let $X = \begin{bmatrix} x_1 & \cdots & x_m \end{bmatrix}^T \in \mathbf{R}^{m \times k}$, $Y = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix} \in \mathbf{R}^{k \times n}$
- assume IID $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$

so log-likelihood is $l(X, Y) = -\dfrac{mn}{2} \log(2\pi\sigma^2) - \dfrac{1}{2\sigma^2} \|A - XY\|_F^2$

latent factor estimation: compress n features into $k < n$ features

- rows x_i^T of X: latent factors of a_i, $i.e.$, representation in new features
- columns y_i of Y: linear maps of latent factors to original features

archetype representation:

- rows $\tilde{y}_i^T$ of Y: archetypes of the data points, $i.e.$, representative features
- rows x_i^T of X: coefficients of the archetypes for each data point

  $i.e.$, *archetype representation* of the data points in terms of $\tilde{y}_i$

Quadratically regularized PCA

$$\text{minimize} \quad \|A - XY\|_F^2 + \gamma(\|X\|_F^2 + \|Y\|_F^2)$$

- variables $X \in \mathbf{R}^{m\times k}$, $Y \in \mathbf{R}^{k\times n}$; $\gamma > 0$ is regularization parameter
- equivalent to the vector form

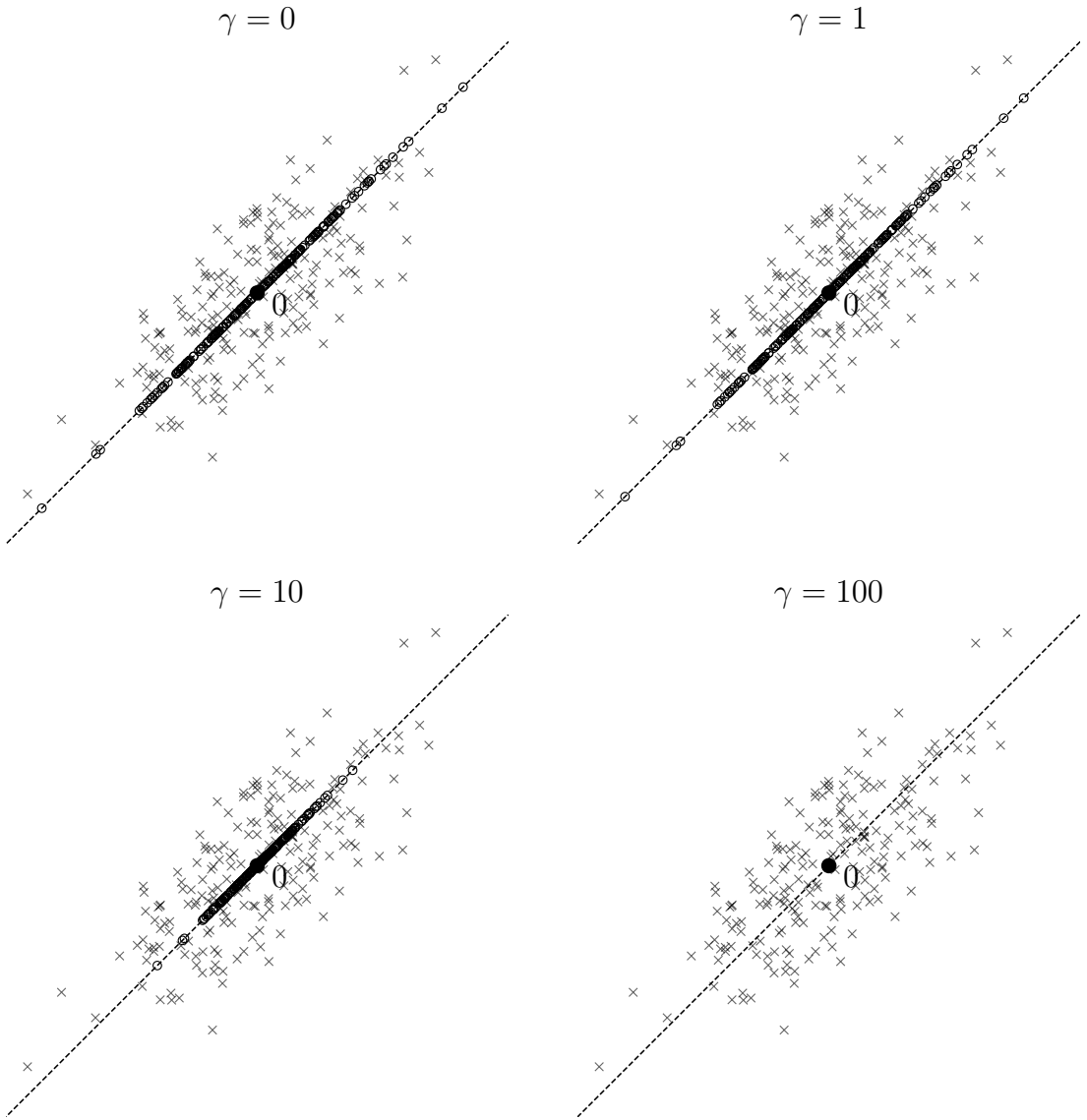
$$\text{minimize} \quad \sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij} - x_i^T y_j)^2 + \gamma\left(\sum_{i=1}^{m}\|x_i\|_2^2 + \sum_{j=1}^{n}\|y_j\|_2^2\right)$$

- analytical solution via SVD: $X^\star = U_k\tilde{\Sigma}_k^{1/2}$ and $Y^\star = \tilde{\Sigma}_k^{1/2}V_k^T$, where $A = U\Sigma V^T$; U_k, V_k are first k columns of U, V, and

$$\tilde{\Sigma}_k = \begin{bmatrix} (\sigma_1 - \gamma)_+ & & 0 \\ & \ddots & \\ 0 & & (\sigma_k - \gamma)_+ \end{bmatrix} \in \mathbf{R}^{k\times k}$$

- still nonunique solutions but bounded

geometric interpretation: project onto a k-dimensional subspace, then shrink coefficients along singular vectors using γ



MAP estimation interpretation:

- assume linear model $a_{ij} = x_i^T y_j + \epsilon_{ij}$ with IID $\epsilon_{ij} \sim \mathcal{N}(0, 1)$
- assume componentwise IID priors $x_{ij}, y_{ij} \sim \mathcal{N}(0, 1/\gamma)$
  (so regularization weight interpreted as inverse variance $\gamma = 1/\sigma^2$)

Robust low rank approximation

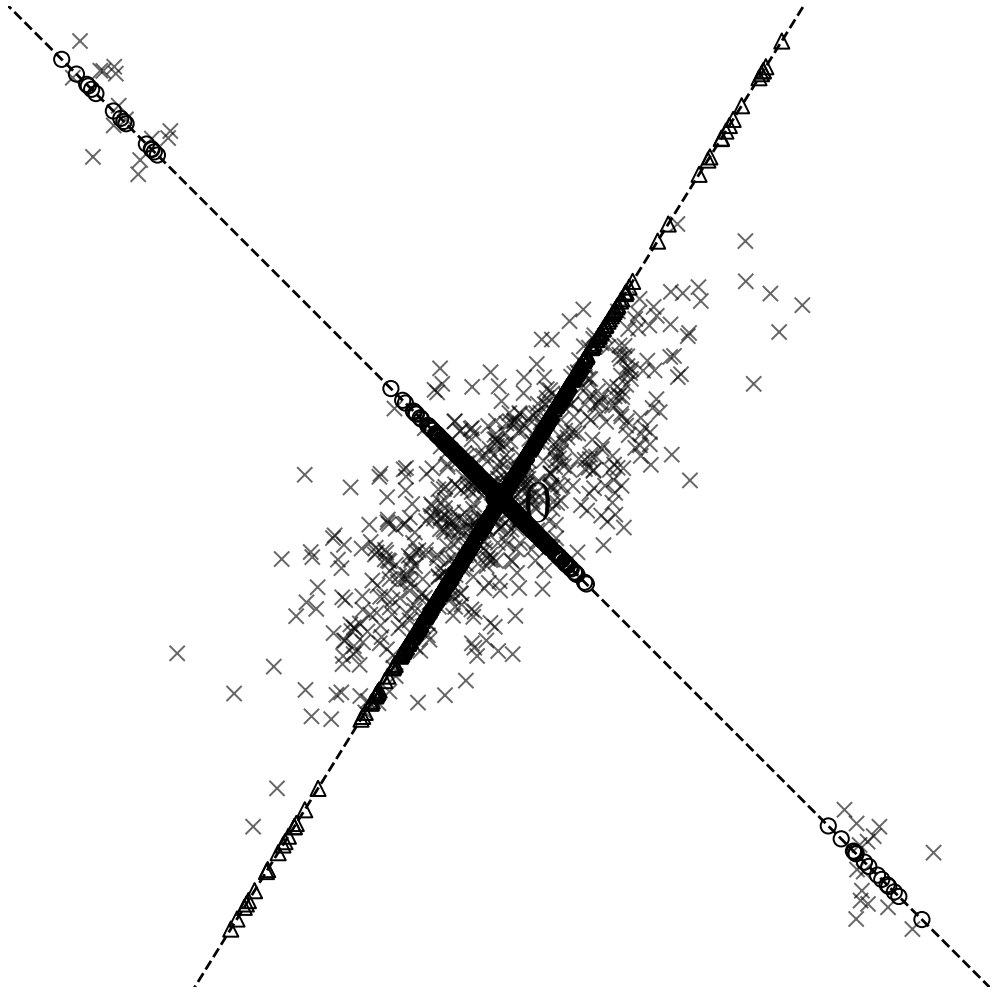
$$\begin{array}{ll} \text{minimize} & \|A - Z\|_{\text{sav}} = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij} - z_{ij}| \\ \text{subject to} & \mathbf{rank}\, Z \leq k \end{array}$$

- componentwise absolute value penalty on residual matrix
- less sensitive to outliers than ordinary PCA
- no analytical solution
- equivalent to factorized form

$$\text{minimize} \quad \|A - XY\|_{\text{sav}} = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij} - x_i^T y_j|$$

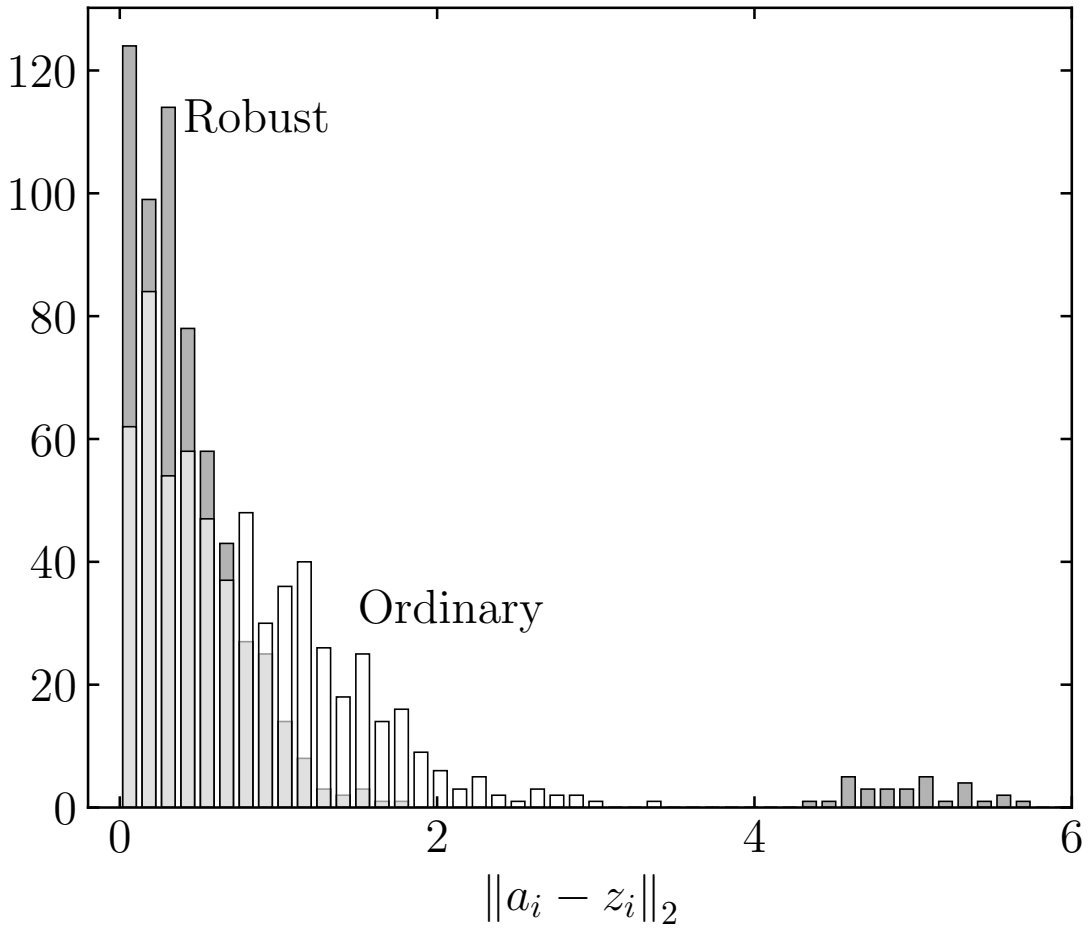
so can be solved via ACS

numerical example: data points (with outliers) in crosses



ordinary PCA in circles; robust approximation in triangles

histogram of distances between data points and approximation



robust PCA:

$$\text{minimize} \quad \|A - XY\|_{\text{sav}} + \gamma(\|X\|_F^2 + \|Y\|_F^2)$$

- equivalent to

$$\begin{array}{ll} \text{minimize} & \|A - Z\|_{\text{sav}} + 2\gamma\|Z\|_* \\ \text{subject to} & \mathbf{rank}\, Z \leq k \end{array}$$

- removing rank constraint gives convex problem

$$\begin{array}{ll} \text{minimize} & \|S\|_{\text{sav}} + 2\gamma\|Z\|_* \\ \text{subject to} & S + Z = A \end{array}$$

  - S denotes the sparse component of A ($e.g.$, outliers)
  - Z denotes the low rank component of A

Structural regularization and constraints

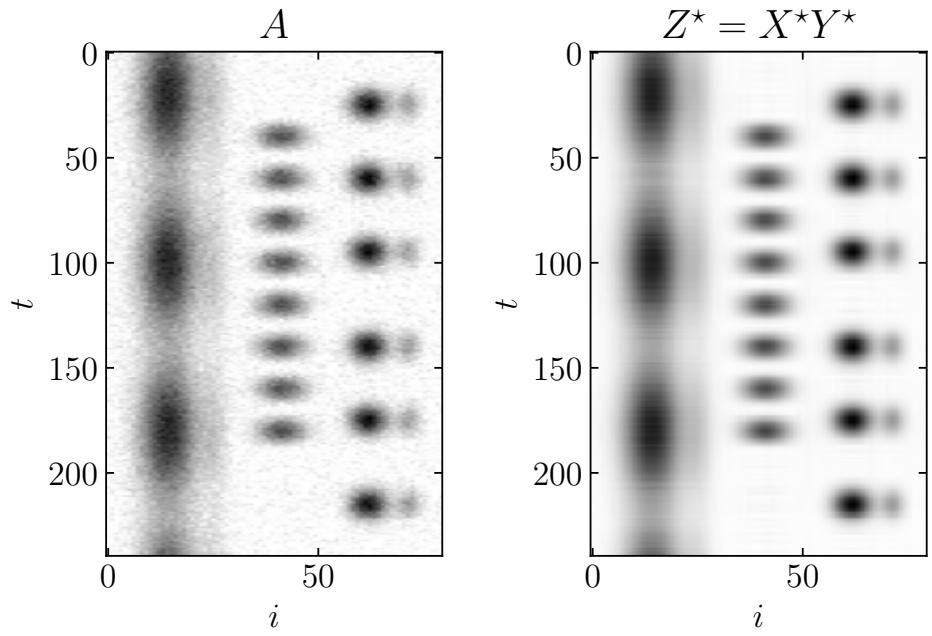
nonnegative matrix factorization (NMF):

$$\begin{array}{ll} \text{minimize} & \|A - XY\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n \left(a_{ij} - x_i^T y_j\right)^2 \\ \text{subject to} & x_i \succeq 0, \quad i = 1, \ldots, m \\ & y_i \succeq 0, \quad i = 1, \ldots, n \end{array}$$

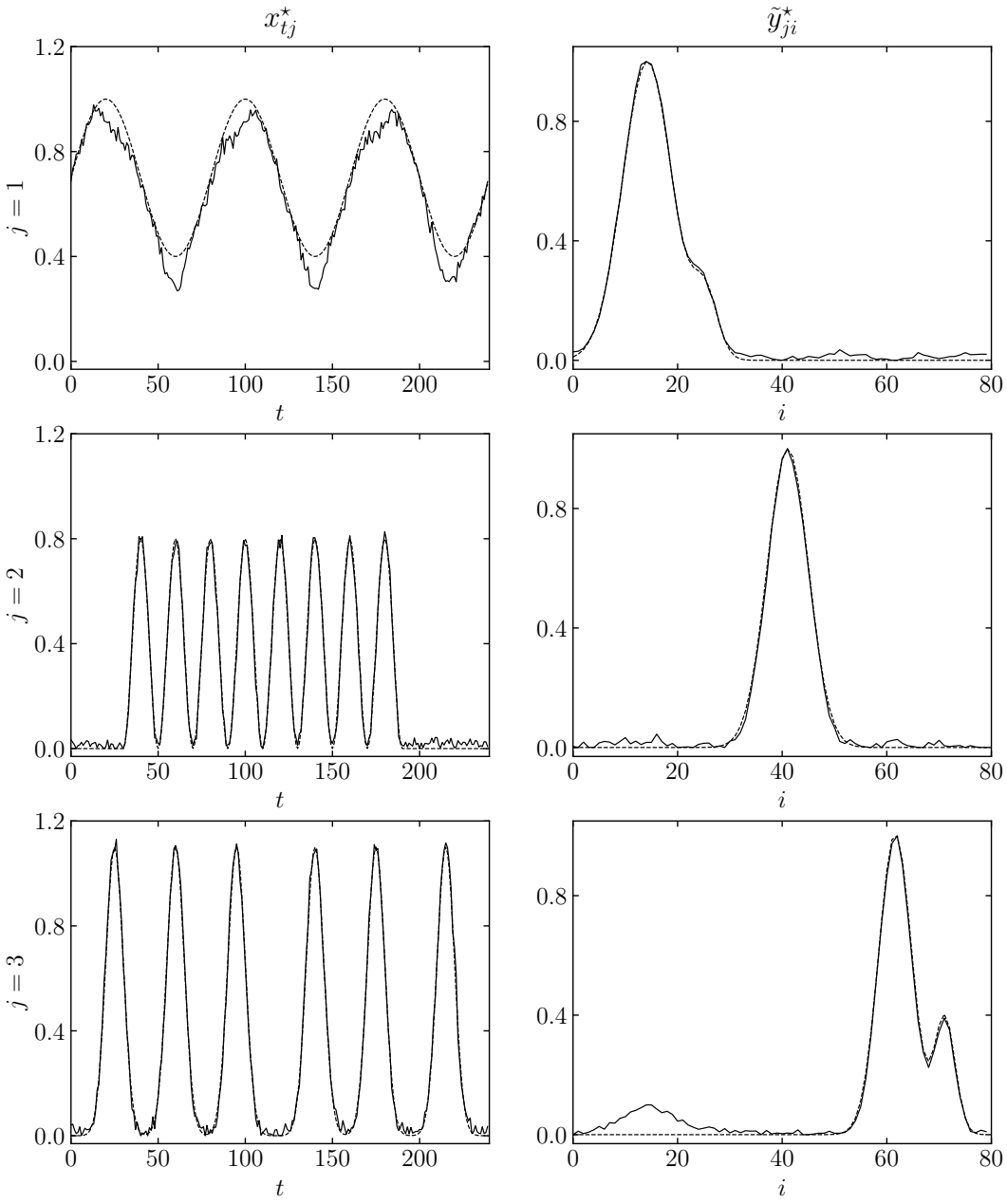
- no analytical solution, but can be solved via ACS

example:

- (corrupted) data $A \in \mathbf{R}_+^{m \times n}$
- m: number of time steps
- n: number of features
- NMF result $Z^\star = X^\star Y^\star$ with $X^\star \in \mathbf{R}_+^{m \times 3}$ and $Y^\star \in \mathbf{R}_+^{3 \times n}$



- right: archetypes
- left: coefficients for each time step of archetypes



sparse PCA:

$$\text{minimize} \quad \|A - XY\|_F^2 + \gamma \sum_{i=1}^m \|x_i\|_1$$

- latent factors x_i encouraged to be sparse
- or each point is represented by a few archetypes

example: *dictionary learning* problem

$$\begin{array}{ll} \text{minimize} & \|A - XY\|_F^2 + \gamma \|X\|_{\text{sav}} \\ \text{subject to} & \|Y\|_F^2 \leq \beta \end{array}$$

- Y called *dictionary*, $i.e.$, bounded set of archetypes

feature selection:

$$\text{minimize} \quad \|A - XY\|_F^2 + \gamma \sum_{i=1}^n \|y_i\|_2$$

- encourage columnwise sparsity in Y, $i.e.$, only a few of original n features are used in the archetypes

10. Conclusions

- main ideas of the course
- further topics

Modeling

inverse problem formulation of machine learning models:

- as mathematical optimization problems
- say what you want and how to evaluate a candidate
- incorporate prior knowledge and requirements

tractability

- roughly speaking, requires convexity
- surprisingly many inverse problems can be formulated as convex
- heuristics (local solutions) for nonconvex problems via (sequential) convex optimization, often works quite well

Consequences of convexity

theoretical consequences

- local optima are global
- systematic way of deriving lower bounds on optimal value
- certificates of optimality, unboundedness, or infeasibility

practical consequences

- (most) convex problems can be solved globally and efficiently
- high quality implementations of numerical algorithms
- many hight level symbolic modeling languages enable fast prototyping

Further topics

- convex optimization
  - subgradient calculus, convex analysis
  - algorithms for convex optimization
  - methods for very large scale problems
  - distributed optimization
- numerical linear algebra
  - preconditioning
  - iterative methods
- domain specific languages for optimization
  - symbolic modeling, verification, and canonicalization
  - differentiating through optimization problems
  - embedded code generation