Convex Optimization

Geometrical and Approximation Problems

Approximation

- norm approximation
- least-norm problems
- regularized approximation
- robust approximation

Norm approximation

$$\text{minimize} \quad \|Ax - b\|$$

($A \in \mathbf{R}^{m \times n}$ with $m \geq n$, $\|\cdot\|$ is a norm on $\mathbf{R}^m$)

interpretations of solution $x^\star = \operatorname{argmin}_x \|Ax - b\|$:

- **geometric**: $Ax^\star$ is point in $\mathcal{R}(A)$ closest to b
- **estimation**: linear measurement model

$$y = Ax + v$$

  y are measurements, x is unknown, v is measurement error

  given $y = b$, best guess of x is $x^\star$

- **optimal design**: x are design variables (input), Ax is result (output)

  $x^\star$ is design that best approximates desired result b

examples

- least-squares approximation ($\|\cdot\|_2$): solution satisfies normal equations

$$A^T A x = A^T b$$

  ($x^\star = (A^T A)^{-1} A^T b$ if $\mathbf{Rank}\, A = n$)

- Chebyshev approximation ($\|\cdot\|_\infty$): can be solved as an LP

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

- sum of absolute residuals approximation ($\|\cdot\|_1$): can be solved as an LP

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

Penalty function approximation

$$\begin{array}{ll} \text{minimize} & \phi(r_1) + \cdots + \phi(r_m) \\ \text{subject to} & r = Ax - b \end{array}$$

($A \in \mathbf{R}^{m \times n}$, $\phi : \mathbf{R} \to \mathbf{R}$ is a convex penalty function)

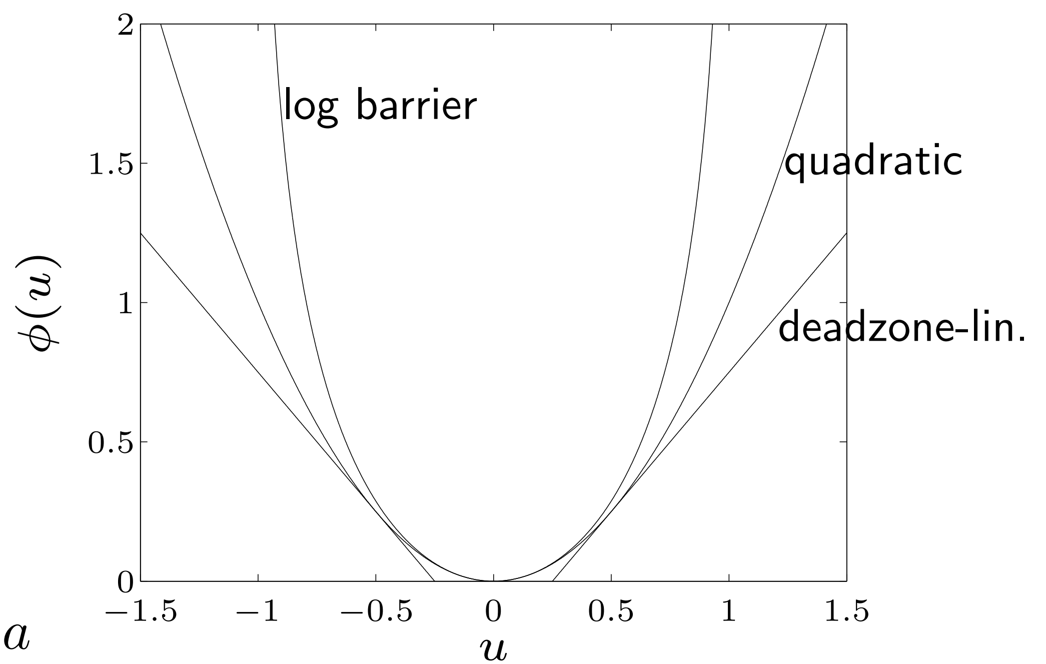
examples

- quadratic: $\phi(u) = u^2$
- deadzone-linear with width a:

$$\phi(u) = \max\{0, |u| - a\}$$

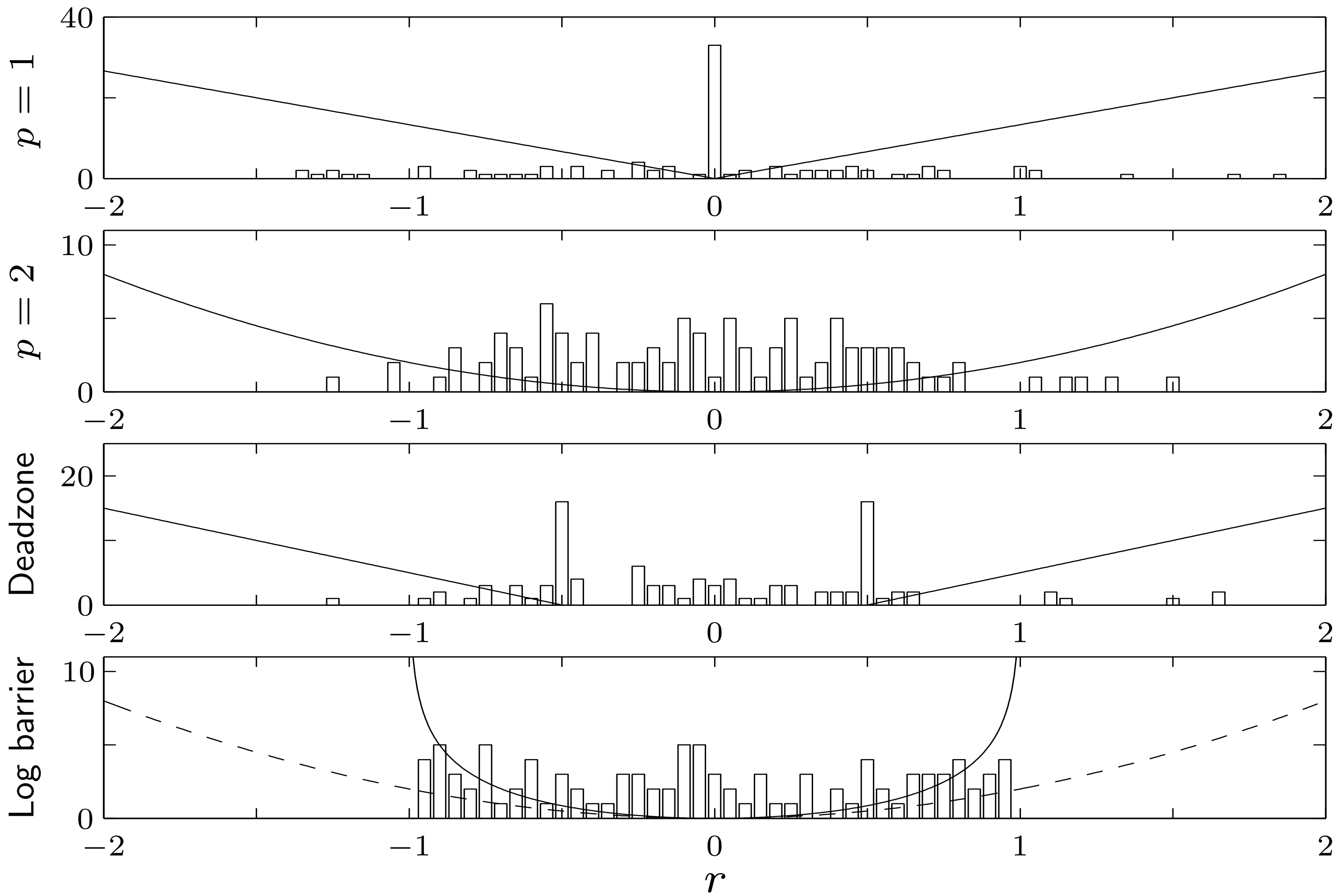
- log-barrier with limit a:

$$\phi(u) = \begin{cases} -a^2 \log(1 - (u/a)^2) & |u| < a \\ \infty & \text{otherwise} \end{cases}$$



example ($m = 100$, $n = 30$): histogram of residuals for penalties

$$\phi(u) = |u|, \quad \phi(u) = u^2, \quad \phi(u) = \max\{0, |u| - a\}, \quad \phi(u) = -\log(1 - u^2)$$

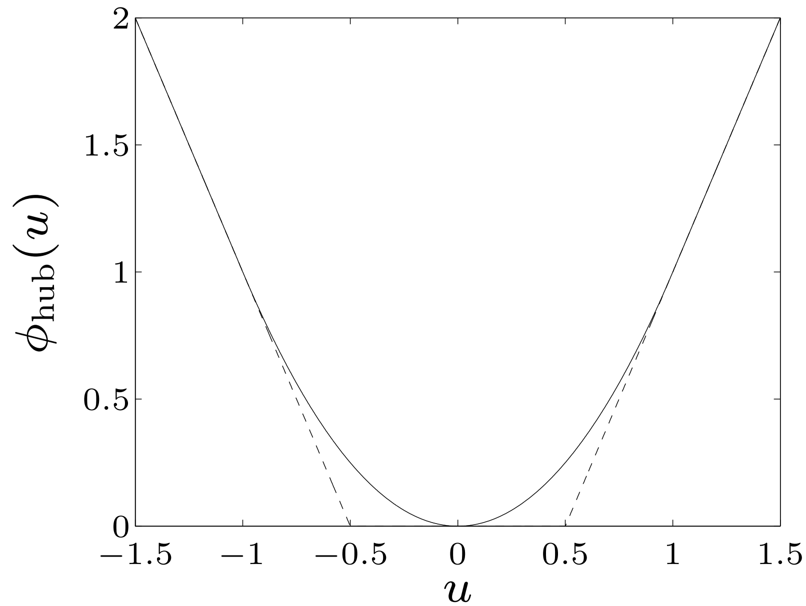


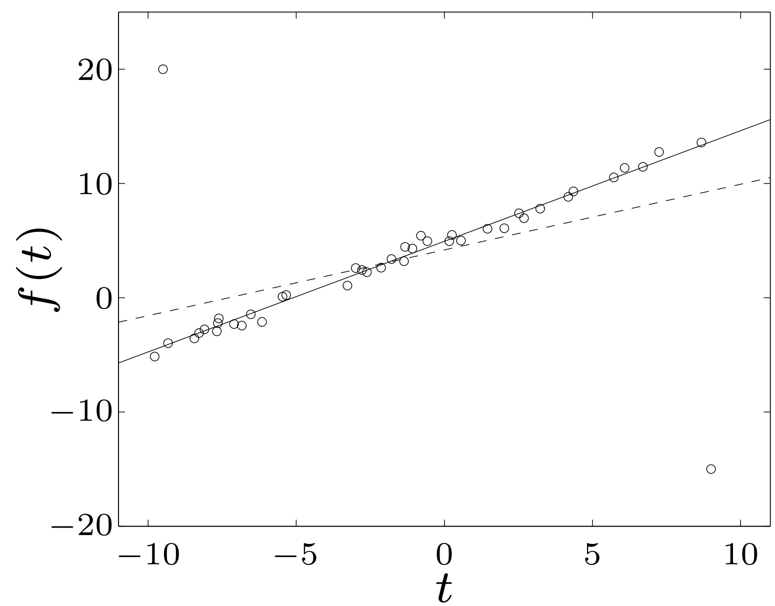
shape of penalty function has large effect on distribution of residuals

Huber penalty function (with parameter M)

$$\phi_{\text{hub}}(u) = \begin{cases} u^2 & |u| \leq M \\ M(2|u| - M) & |u| > M \end{cases}$$

linear growth for large u makes approximation less sensitive to outliers





- left: Huber penalty for $M = 1$
- right: affine function $f(t) = \alpha + \beta t$ fitted to 42 points t_i, y_i (circles) using quadratic (dashed) and Huber (solid) penalty

Least-norm problems

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

($A \in \mathbf{R}^{m \times n}$ with $m \leq n$, $\| \cdot \|$ is a norm on $\mathbf{R}^n$)

interpretations of solution $x^\star = \mathrm{argmin}_{Ax=b} \|x\|$:

- **geometric:** $x^\star$ is point in affine set $\{x \mid Ax = b\}$ with minimum distance to 0
- **estimation:** $b = Ax$ are (perfect) measurements of x; $x^\star$ is smallest ('most plausible') estimate consistent with measurements
- **design:** x are design variables (inputs); b are required results (outputs)

  $x^\star$ is smallest ('most efficient') design that satisfies requirements

examples

- least-squares solution of linear equations ($\|\cdot\|_2$):

  can be solved via optimality conditions

$$2x + A^T\nu = 0, \qquad Ax = b$$

- minimum sum of absolute values ($\|\cdot\|_1$): can be solved as an LP

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

  tends to produce sparse solution $x^\star$

extension: least-penalty problem

$$\begin{array}{ll} \text{minimize} & \phi(x_1) + \cdots + \phi(x_n) \\ \text{subject to} & Ax = b \end{array}$$

$\phi : \mathbf{R} \to \mathbf{R}$ is convex penalty function

Regularized approximation

$$\text{minimize (w.r.t. } \mathbf{R}^2_+) \quad (\|Ax - b\|, \|x\|)$$

$A \in \mathbf{R}^{m \times n}$, norms on $\mathbf{R}^m$ and $\mathbf{R}^n$ can be different

interpretation: find good approximation $Ax \approx b$ with small x

- **estimation:** linear measurement model $y = Ax + v$, with prior knowledge that $\|x\|$ is small
- **optimal design**: small x is cheaper or more efficient, or the linear model $y = Ax$ is only valid for small x
- **robust approximation:** good approximation $Ax \approx b$ with small x is less sensitive to errors in A than good approximation with large x

Scalarized problem

$$\text{minimize} \quad \|Ax - b\| + \gamma\|x\|$$

- solution for $\gamma > 0$ traces out optimal trade-off curve
- other common method: minimize $\|Ax - b\|^2 + \delta\|x\|^2$ with $\delta > 0$

Tikhonov regularization

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

can be solved as a least-squares problem

$$\text{minimize} \quad \left\| \begin{bmatrix} A \\ \sqrt{\delta} I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$$

solution $x^\star = (A^T A + \delta I)^{-1} A^T b$

Signal reconstruction

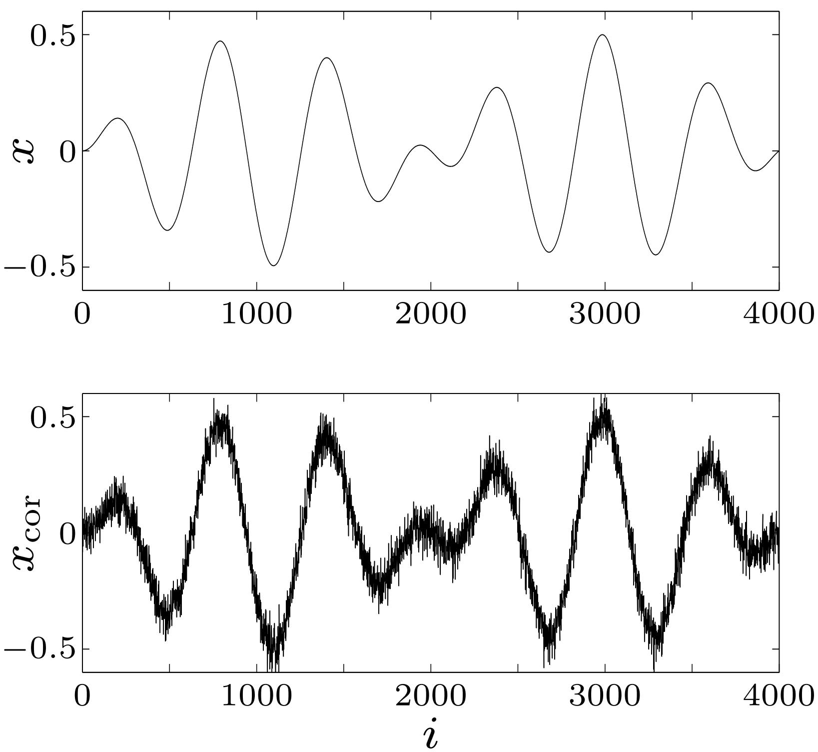
$$\text{minimize (w.r.t. } \mathbf{R}^2_+) \quad (\|\hat{x} - x_{\text{cor}}\|_2, \phi(\hat{x}))$$

- $x \in \mathbf{R}^n$ is unknown signal
- $x_{\text{cor}} = x + v$ is (known) corrupted version of x, with additive noise v
- variable $\hat{x}$ (reconstructed signal) is estimate of x
- $\phi : \mathbf{R}^n \to \mathbf{R}$ is regularization function or smoothing objective

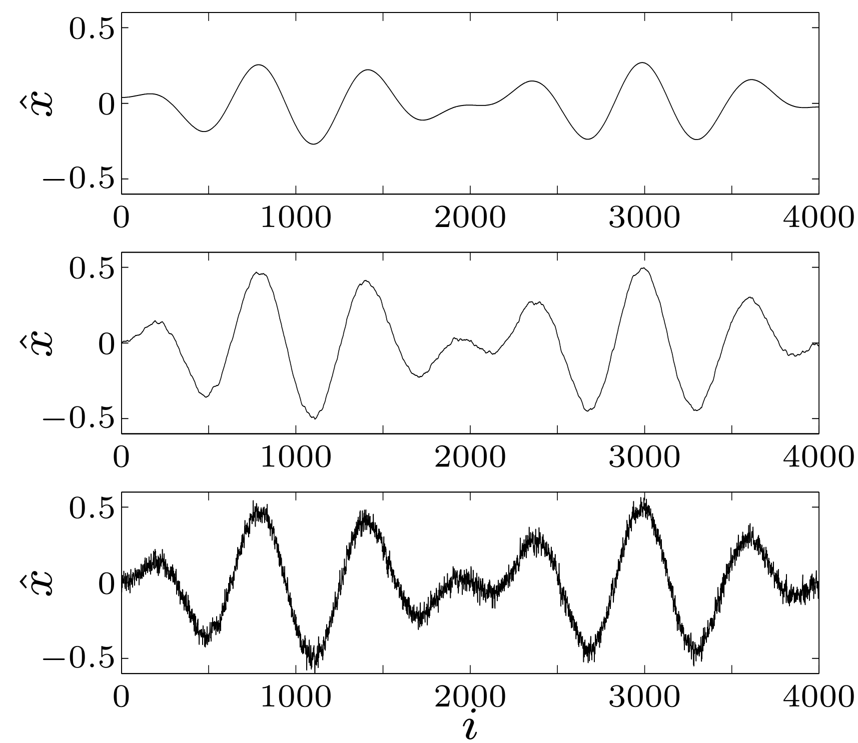
examples: quadratic smoothing, total variation smoothing:

$$\phi_{\text{quad}}(\hat{x}) = \sum_{i=1}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2, \qquad \phi_{\text{tv}}(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$$

quadratic smoothing example

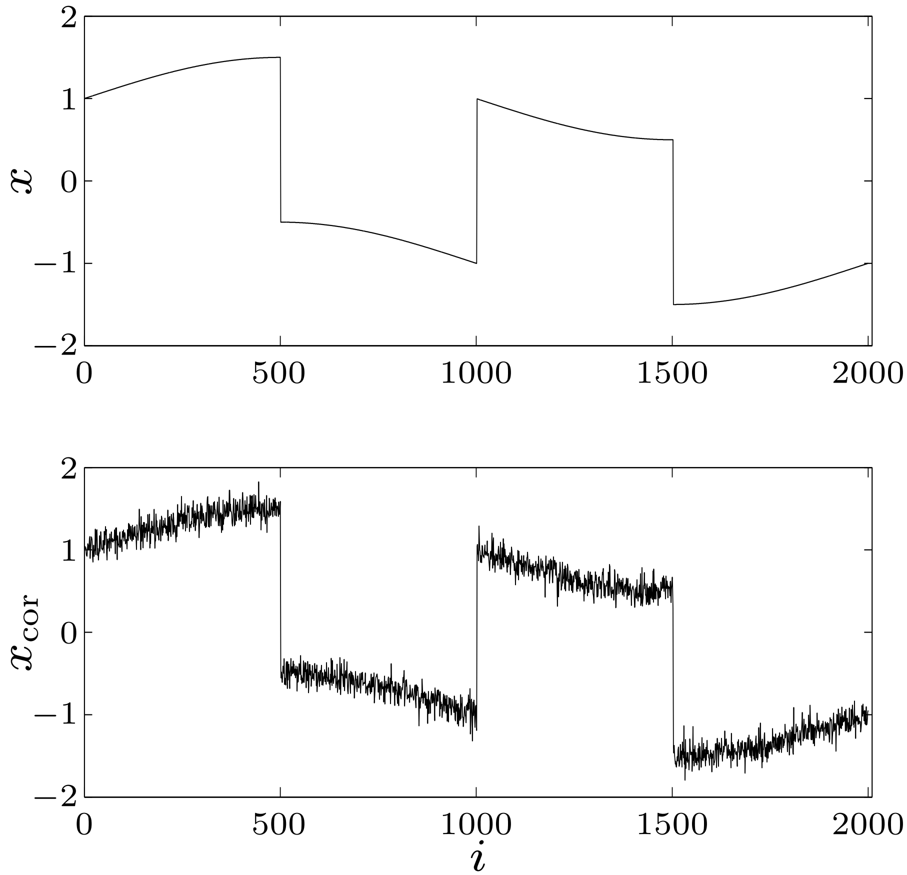


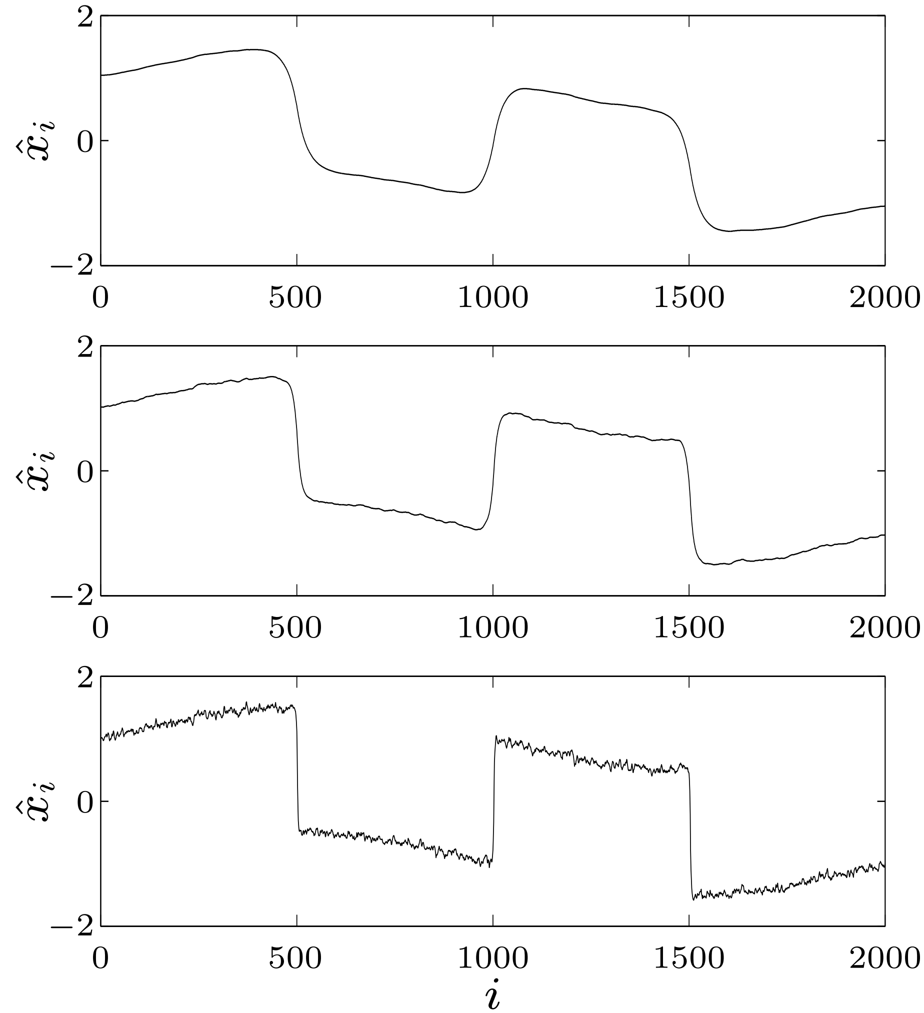
original signal x and noisy signal x_{cor}



three solutions on trade-off curve $\|\hat{x} - x_{\text{cor}}\|_2$ versus $\phi_{\text{quad}}(\hat{x})$

total variation reconstruction example

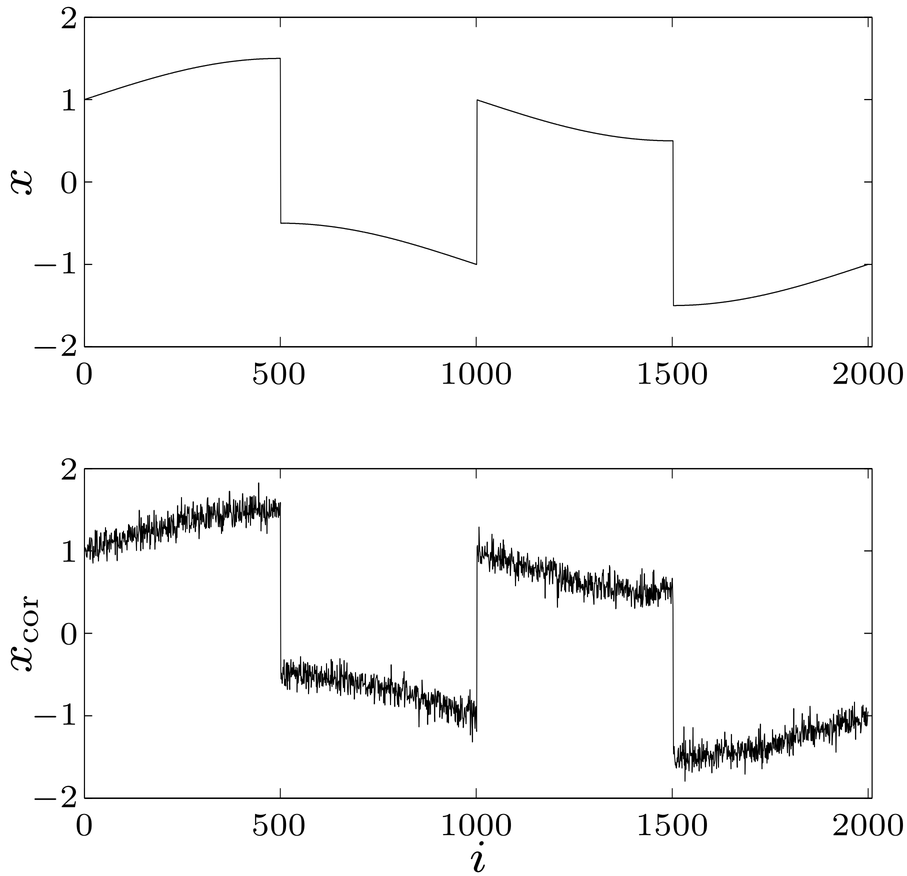




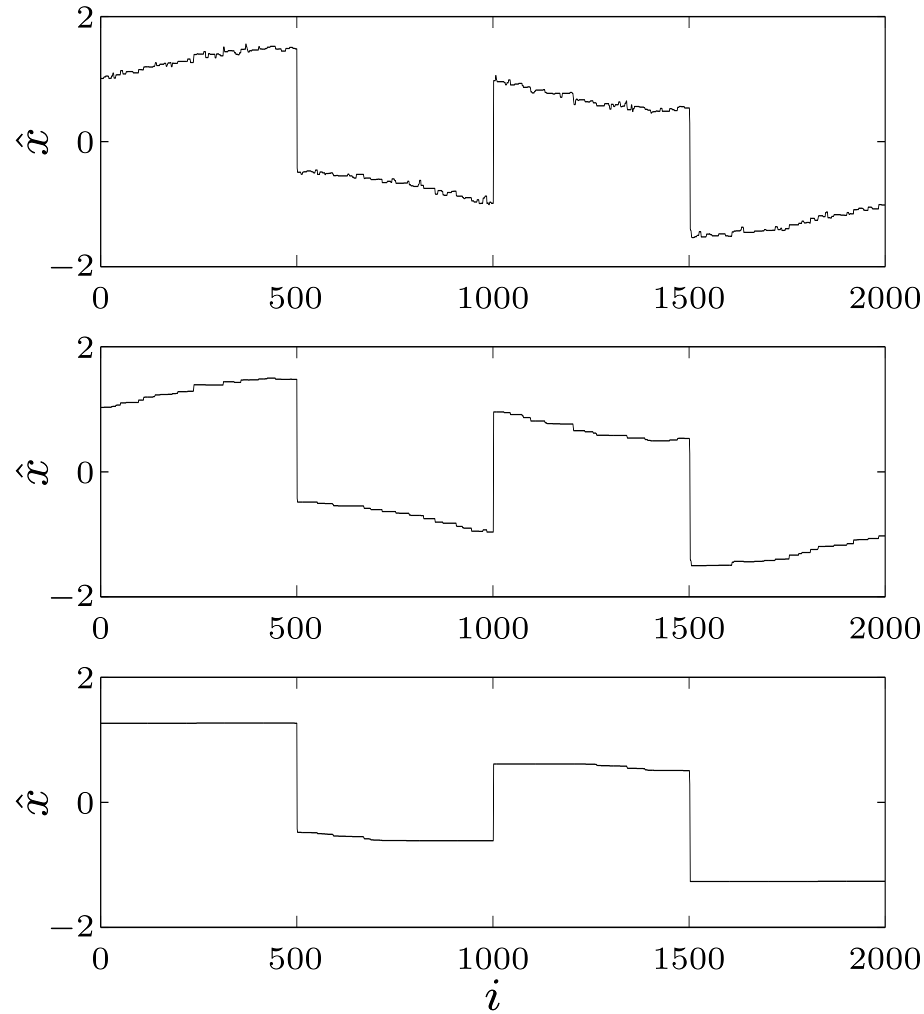
original signal x and noisy signal x_{cor}

three solutions on trade-off curve $\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

quadratic smoothing smooths out noise **and** sharp transitions in signal



original signal x and noisy signal x_{cor}



three solutions on trade-off curve
$\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{tv}}(\hat{x})$

total variation smoothing preserves sharp transitions in signal

Geometrical Problems

Geometrical Problems

- extremal volume ellipsoids
- centering
- placement and facility location.

Minimum volume ellipsoid around a set

Löwner-John ellipsoid of a set C: minimum volume ellipsoid $\mathcal{E}$ s.t. $C \subseteq \mathcal{E}$

- parametrize $\mathcal{E}$ as $\mathcal{E} = \{v \mid \|Av + b\|_2 \leq 1\}$; w.l.o.g. assume $A \in \mathbf{S}^n_{++}$
- $\mathbf{vol}\,\mathcal{E}$ is proportional to $\det A^{-1}$; to compute minimum volume ellipsoid,

$$\begin{array}{ll} \text{minimize (over } A, b) & \log\det A^{-1} \\ \text{subject to} & \sup_{v \in C} \|Av + b\|_2 \leq 1 \end{array}$$

convex, but evaluating the constraint can be hard (for general C)

finite set $C = \{x_1, \dots, x_m\}$:

$$\begin{array}{ll} \text{minimize (over } A, b) & \log\det A^{-1} \\ \text{subject to} & \|Ax_i + b\|_2 \leq 1, \quad i = 1, \dots, m \end{array}$$

also gives Löwner-John ellipsoid for polyhedron $\mathbf{Co}\{x_1, \dots, x_m\}$

Maximum volume inscribed ellipsoid

maximum volume ellipsoid $\mathcal{E}$ inside a convex set $C \subseteq \mathbf{R}^n$

- parametrize $\mathcal{E}$ as $\mathcal{E} = \{Bu + d \mid \|u\|_2 \leq 1\}$; w.l.o.g. assume $B \in \mathbf{S}^n_{++}$
- $\mathbf{vol}\,\mathcal{E}$ is proportional to $\det B$; can compute $\mathcal{E}$ by solving

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & \sup_{\|u\|_2 \leq 1} I_C(Bu + d) \leq 0 \end{array}$$

(where $I_C(x) = 0$ for $x \in C$ and $I_C(x) = \infty$ for $x \notin C$)

convex, but evaluating the constraint can be hard (for general C)

polyhedron $\{x \mid a_i^T x \leq b_i,\ i = 1, \ldots, m\}$:

$$\begin{array}{ll} \text{maximize} & \log\det B \\ \text{subject to} & \|Ba_i\|_2 + a_i^T d \leq b_i, \quad i = 1, \ldots, m \end{array}$$

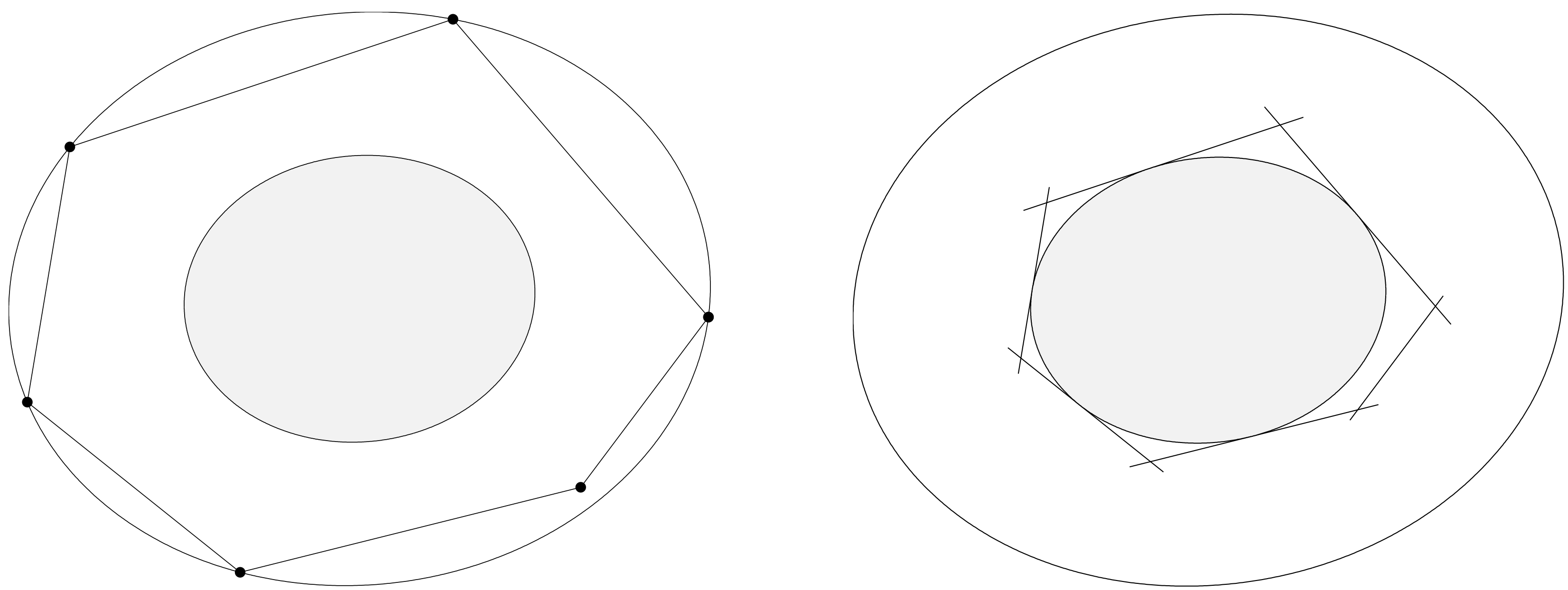
(constraint follows from $\sup_{\|u\|_2 \leq 1} a_i^T(Bu + d) = \|Ba_i\|_2 + a_i^T d$)

Efficiency of ellipsoidal approximations

$C \subseteq \mathbf{R}^n$ convex, bounded, with nonempty interior

- Löwner-John ellipsoid, shrunk by a factor n, lies inside C
- maximum volume inscribed ellipsoid, expanded by a factor n, covers C

example (for two polyhedra in $\mathbf{R}^2$)



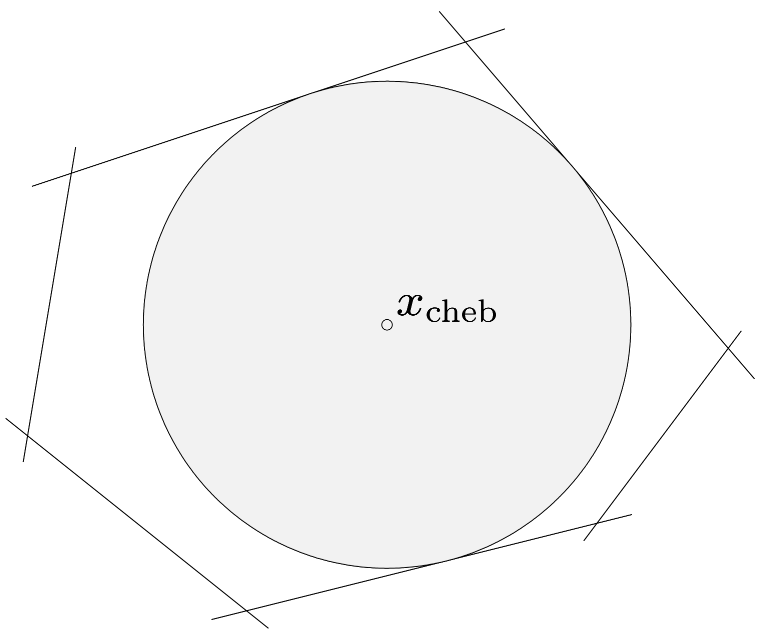
factor n can be improved to $\sqrt{n}$ if C is symmetric

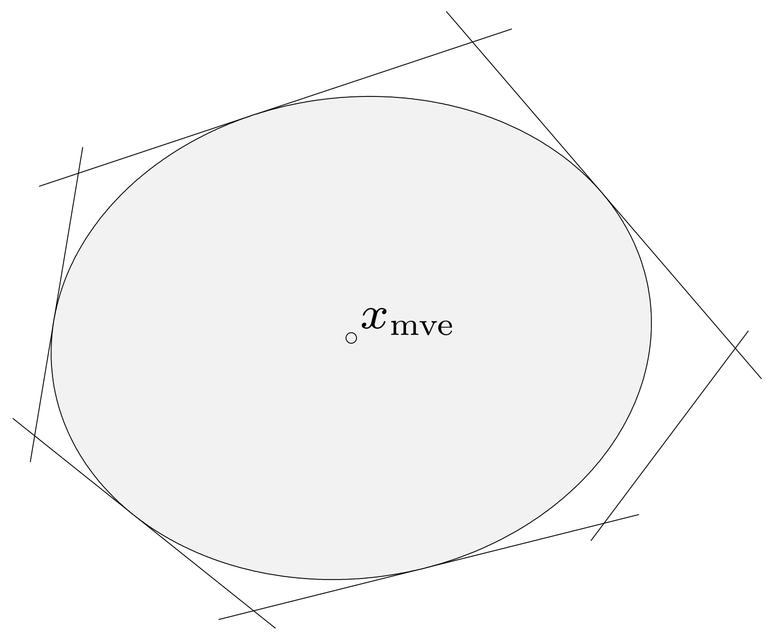
Centering

some possible definitions of 'center' of a convex set C:

- center of largest inscribed ball ('Chebyshev center')

  for polyhedron, can be computed via linear programming (page **??**)
- center of maximum volume inscribed ellipsoid (page 19)





MVE center is invariant under affine coordinate transformations

Analytic center of a set inequalities

the analytic center of set of convex inequalities and linear equations

$$f_i(x) \le 0, \quad i = 1, \dots, m, \qquad Fx = g$$

is defined as the optimal point of

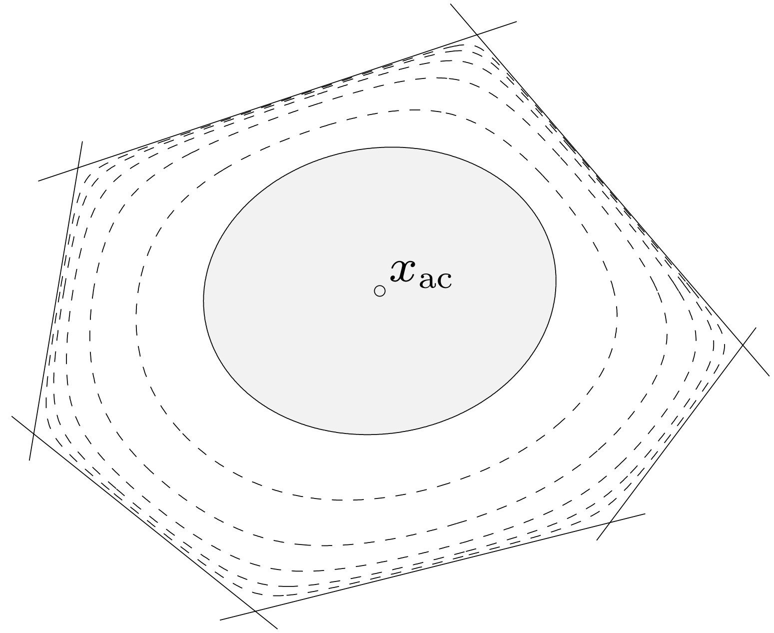
$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^m \log(-f_i(x)) \\ \text{subject to} & Fx = g \end{array}$$

- more easily computed than MVE or Chebyshev center (see later)
- not just a property of the feasible set: two sets of inequalities can describe the same set, but have different analytic centers

analytic center of linear inequalities $a_i^T x \le b_i$, $i = 1, \ldots, m$

x_{ac} is minimizer of

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



inner and outer ellipsoids from analytic center:

$$\mathcal{E}_{\text{inner}} \subseteq \{x \mid a_i^T x \le b_i,\ i = 1, \ldots, m\} \subseteq \mathcal{E}_{\text{outer}}$$

where

$$\begin{aligned} \mathcal{E}_{\text{inner}} &= \{x \mid (x - x_{\text{ac}})^T \nabla^2 \phi(x_{\text{ac}})(x - x_{\text{ac}} \le 1\} \\ \mathcal{E}_{\text{outer}} &= \{x \mid (x - x_{\text{ac}})^T \nabla^2 \phi(x_{\text{ac}})(x - x_{\text{ac}}) \le m(m-1)\} \end{aligned}$$

Placement and facility location

- N points with coordinates $x_i \in \mathbf{R}^2$ (or $\mathbf{R}^3$)
- some positions x_i are given; the other x_i's are variables
- for each pair of points, a cost function $f_{ij}(x_i, x_j)$

placement problem

$$\text{minimize} \quad \sum_{i \neq j} f_{ij}(x_i, x_j)$$

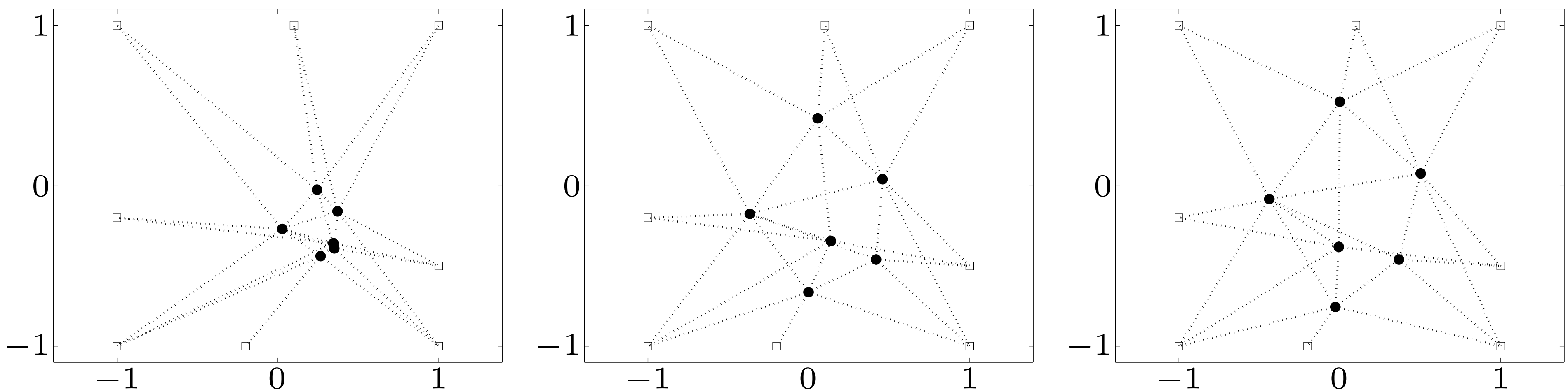
variables are positions of free points

interpretations

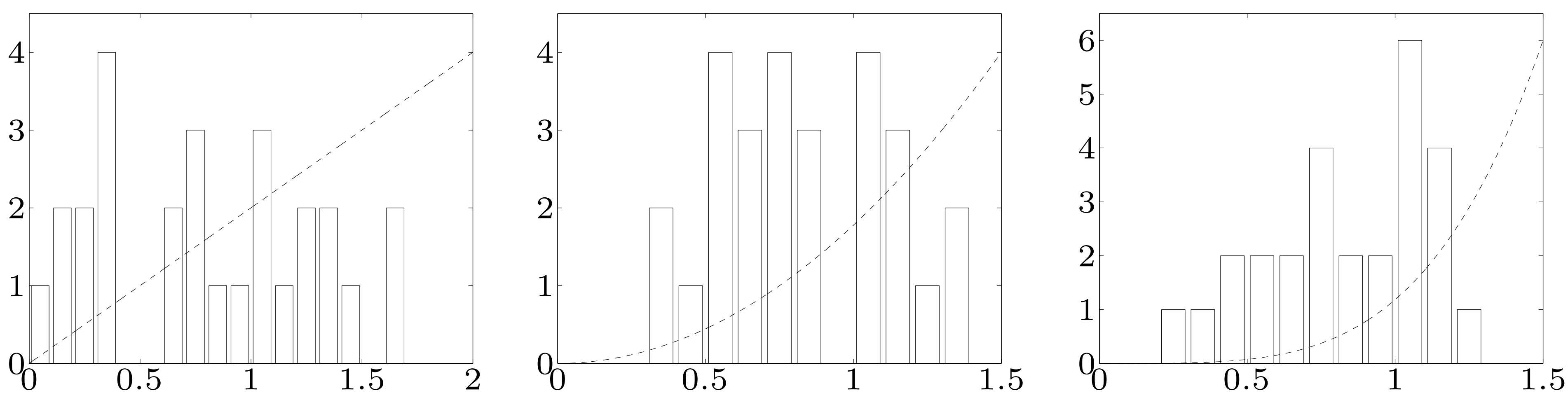
- points represent plants or warehouses; f_{ij} is transportation cost between facilities i and j
- points represent cells on an IC; f_{ij} represents wirelength

example: minimize $\sum_{(i,j)\in\mathcal{A}} h(\|x_i - x_j\|_2)$, with 6 free points, 27 links

optimal placement for $h(z) = z$, $h(z) = z^2$, $h(z) = z^4$



histograms of connection lengths $\|x_i - x_j\|_2$



Distance matrices

Distance matrices . . .

- The problem of reconstructing an N-point Euclidean metric, given **partial** information on pairwise distances between points $v_i,\ i = 1, \ldots, N$ can also be cast as an SDP, known as and **Euclidean Distance Matrix Completion** problem.

$$\begin{array}{ll} \text{find} & D \\ \text{subject to} & \mathbf{1}v^T + v\mathbf{1}^T - D \succeq 0 \\ & D_{ij} = \|v_i - v_j\|_2^2, \quad (i,j) \in S \\ & v \geq 0 \end{array}$$

  in the variables $D \in \mathbf{S}_n$ and $v \in \mathbf{R}^n$, on a subset $S \subset [1, N]^2$.

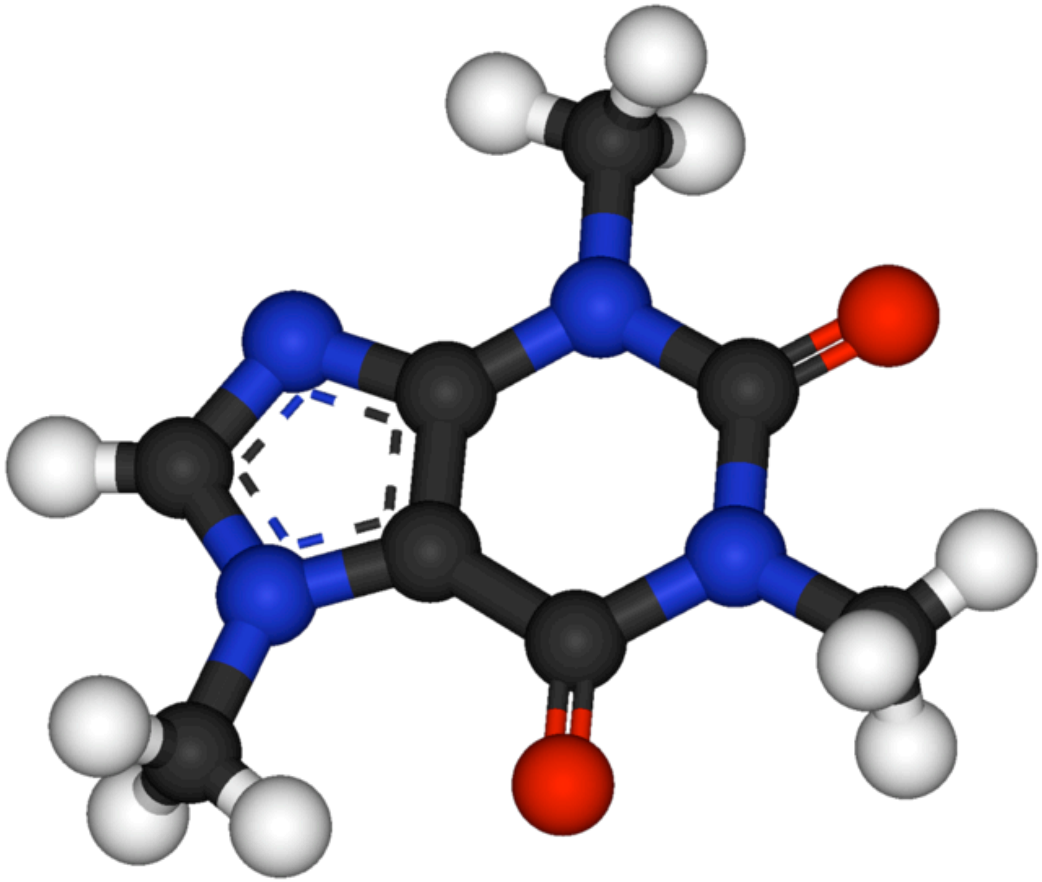
- We can add further constraints to this problem given additional structural info on the configuration.

- Applications in sensor networks, molecular conformation reconstruction etc. . .

Distance matrices . . .

[Dattorro, 2005] 3D map of the USA reconstructed from pairwise distances on 5000 points. Distances reconstructed from Latitude/Longitude data.

Distance matrices . . .

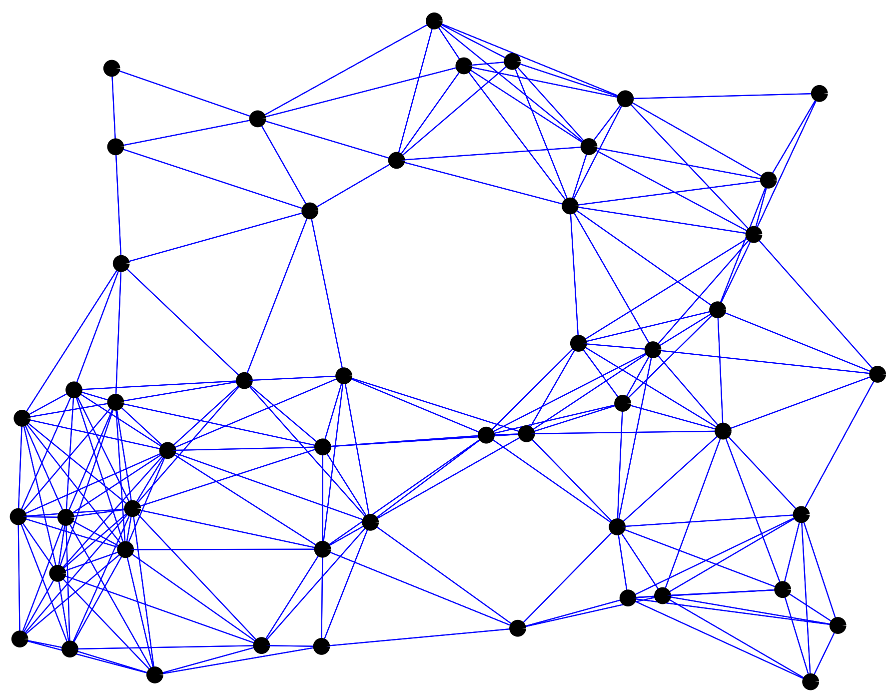


3D **Caffeine.** Reconstruct molecules from MRI data...

Mixing rates for Markov chains & maximum variance unfolding

Mixing rates for Markov chains & unfolding

- Let $G = (V, E)$ be an **undirected graph** with n vertices and m edges.
- We define a **Markov chain** on this graph, and let $w_{ij} \geq 0$ be the transition rate for edge $(i, j) \in V$.



Mixing rates for Markov chains & unfolding

- Let $\pi(t)$ be the state distribution at time t, its evolution is governed by the heat equation
$$d\pi(t) = -L\pi(t)dt$$
with
$$L_{ij} = \begin{cases} -w_{ij} & \text{if } i \neq j,\ (i,j) \in V \\ 0 & \text{if } (i,j) \notin V \\ \sum_{(i,k)\in V} w_{ik} & \text{if } i = j \end{cases}$$
the **graph Laplacian** matrix, which means
$$\pi(t) = e^{-Lt}\pi(0).$$

Mixing rates for Markov chains & unfolding

[Sun, Boyd, Xiao, and Diaconis, 2006]

- Maximizing the mixing rate of the Markov chain means solving

$$\begin{array}{ll} \text{maximize} & t \\ \text{subject to} & L(w) \succeq t(\mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^T) \\ & \sum_{(i,j)\in V} d_{ij}^2 w_{ij} \leq 1 \\ & w \geq 0 \end{array}$$

  in the variable $w \in \mathbf{R}^m$, with (normalization) parameters $d_{ij}^2 \geq 0$.
- Since $L(w)$ is an affine function of the variable $w \in \mathbf{R}^m$, this is a **semidefinite program** in $w \in \mathbf{R}^m$.

Mixing rates for Markov chains & unfolding

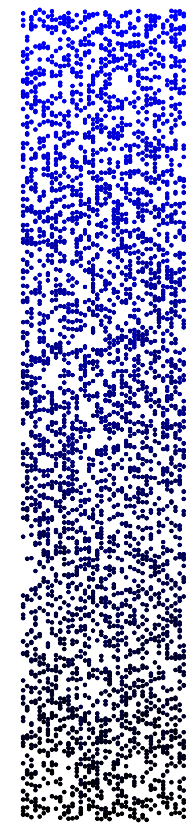
[Weinberger and Saul, 2006, Sun et al., 2006]

- The **dual** means solving

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(X(\mathbf{I}-(1/n)\mathbf{1}\mathbf{1}^T)) \\ \text{subject to} & X_{ii}-2X_{ij}+X_{jj} \leq d_{ij}^2, \quad (i,j) \in V \\ & X \succeq 0, \end{array}$$

  in the variable $X \in \mathbf{S}_n$.
- This is a **maximum variance unfolding problem.**

Mixing rates for Markov chains & unfolding



From [Sun et al., 2006]: we are given pairwise 3D distances for k-nearest neighbors in the point set on the right. We plot the maximum variance point set satisfying these pairwise distance bounds on the right.

*

References

J. Dattorro. *Convex optimization & Euclidean distance geometry*. Meboo Publishing USA, 2005.

J. Sun, S. Boyd, L. Xiao, and P. Diaconis. The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM Review*, 48(4):681–699, 2006.

K.Q. Weinberger and L.K. Saul. Unsupervised Learning of Image Manifolds by Semidefinite Programming. *International Journal of Computer Vision*, 70(1):77–90, 2006.