Convex Optimization: Applications

Lecturer: Pradeep Ravikumar
Co-instructor: Aarti Singh

Convex Optimization 10-725/36-725

Based on material from Boyd, Vandenberghe

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 $x^\star = \mathrm{argmin}_x \|Ax - b\|$:

- **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

Norm Approximation: ell_2 norm

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

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

- 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 $\mathbf{rank}\, A = n$)

Norm Approximation: ell_infty norm

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

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

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

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

Norm Approximation: ell_1 norm

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

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

- 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)

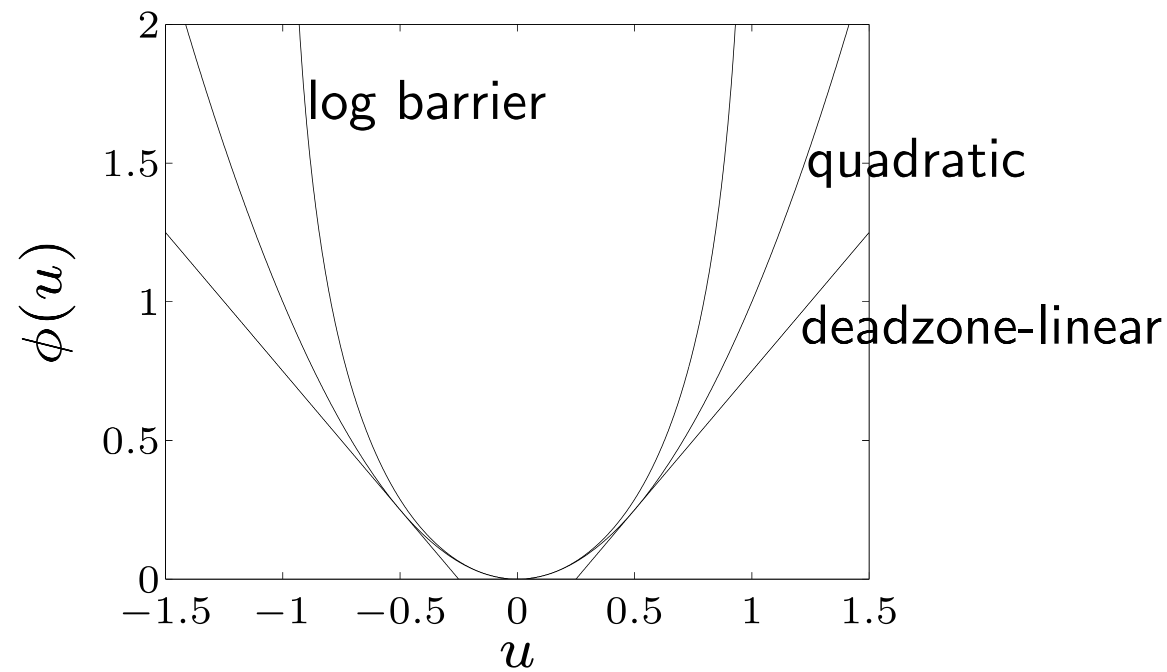
examples

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

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

- log-barrier with limit a:

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

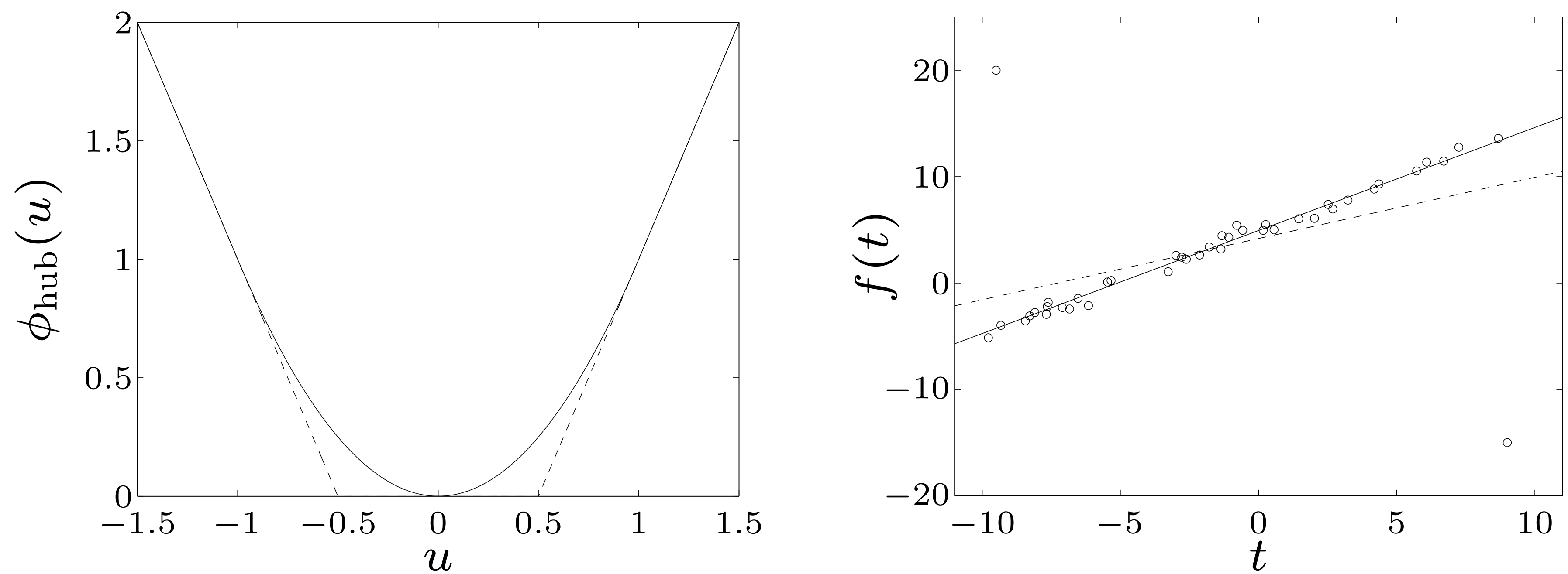


Penalty Function Approximation

Huber penalty function (with parameter M)

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

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

Least Norm Problems: ell_1 norm

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

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

- 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 $x^\star$

Least Norm Problems: least penalty extension

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

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

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

Regularized Approximation

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

- 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

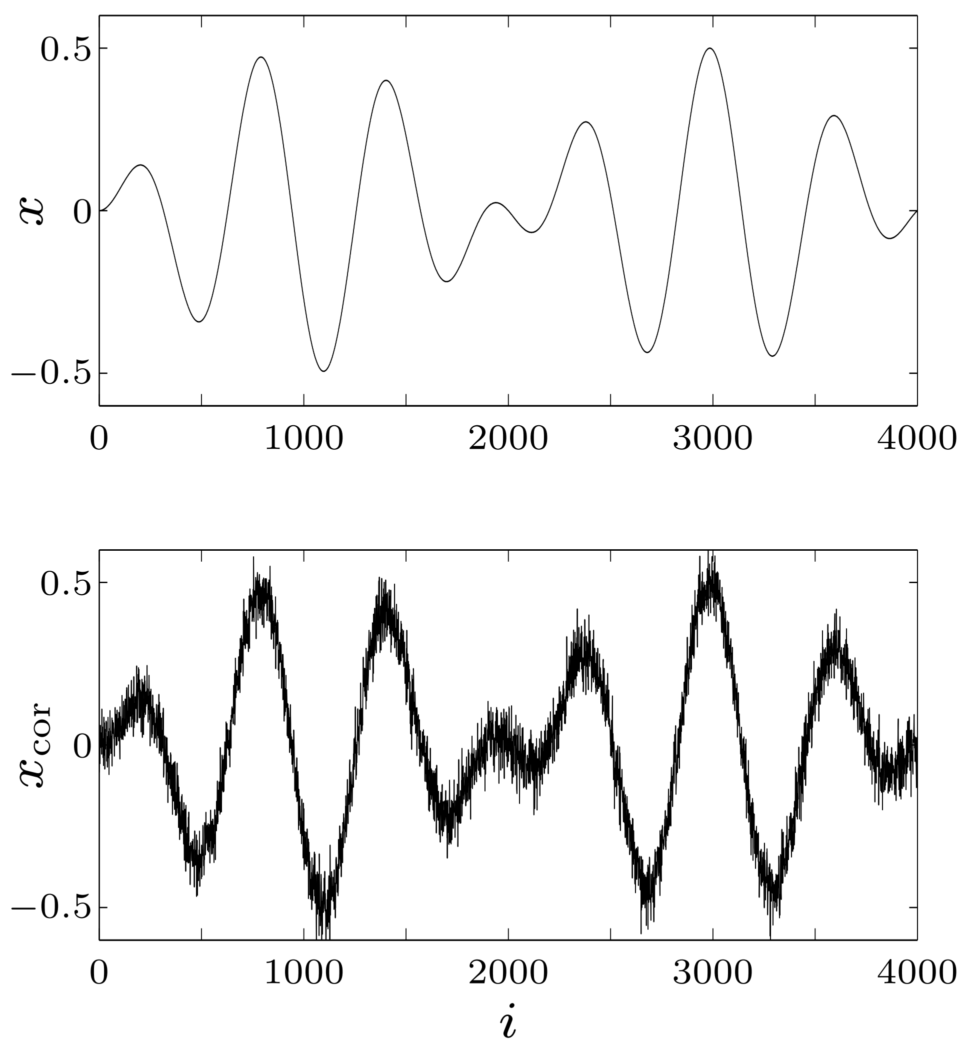
$$\text{minimize (w.r.t. } \mathbf{R}_+^2) \quad (\|\hat{x} - x_{\text{cor}}\|_2, \phi(\hat{x}))$$

- $x \in \mathbf{R}^n$ 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

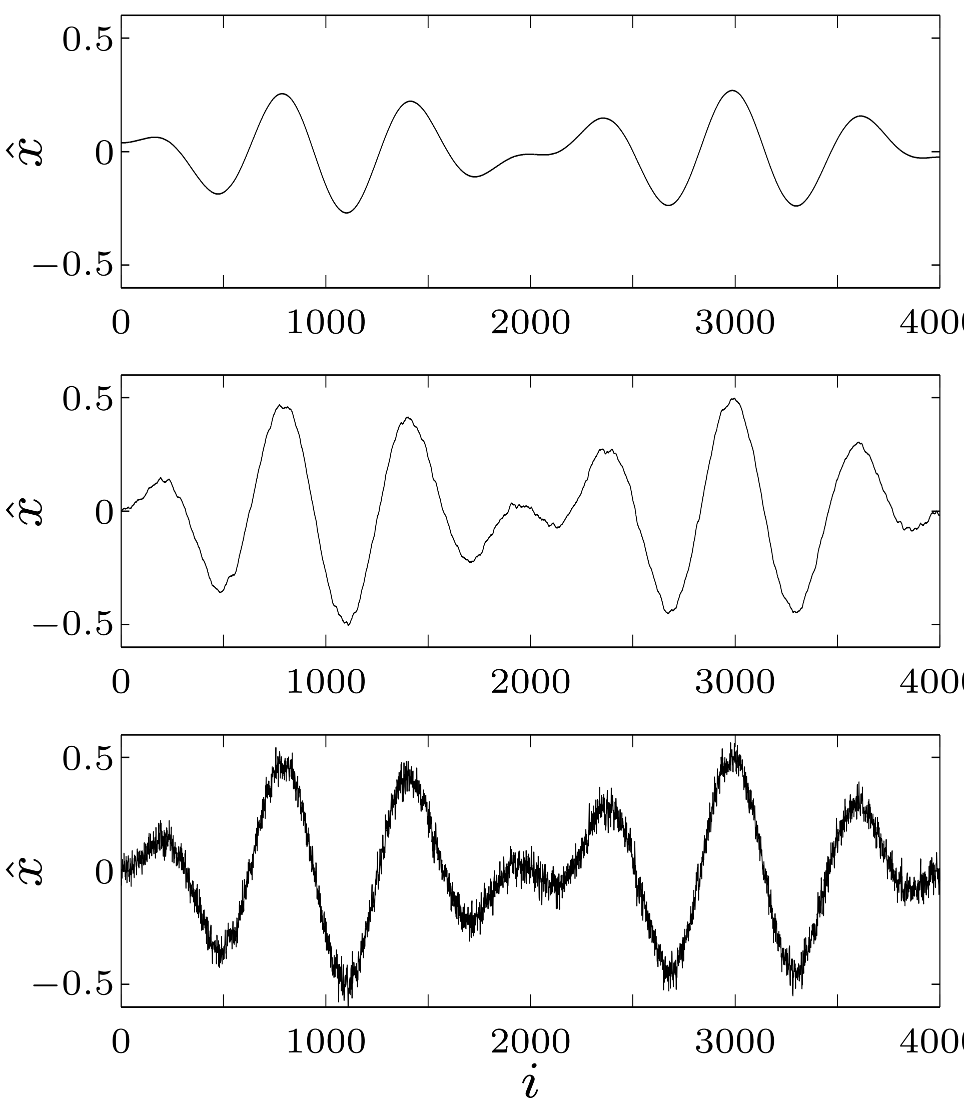
examples: quadratic smoothing, total variation smoothing:

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

Signal Reconstruction: Quadratic Smoothing

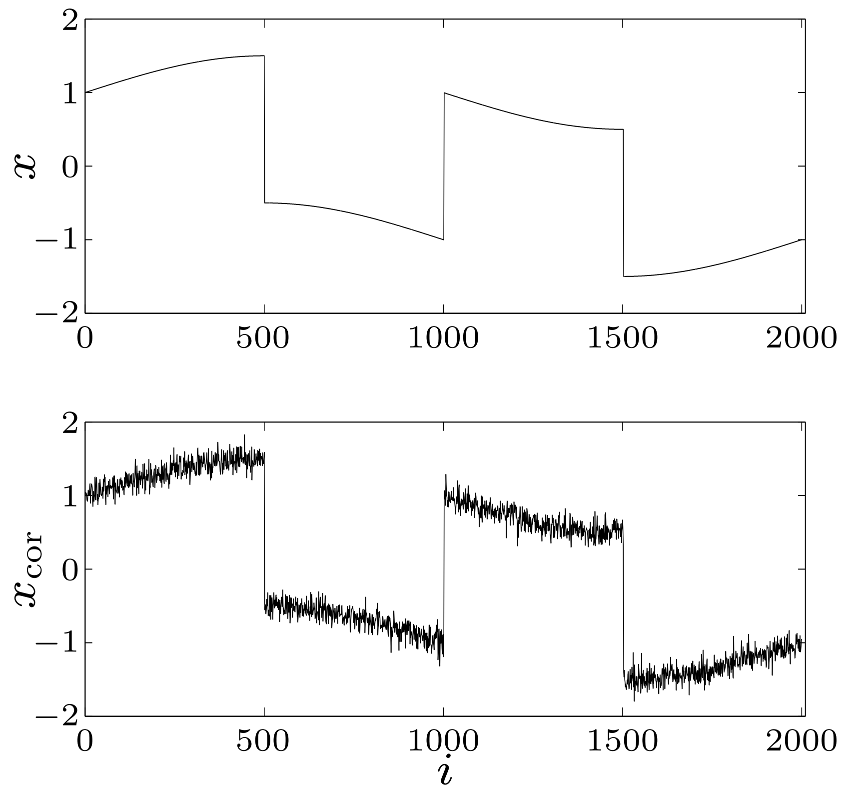


original signal x and noisy signal x_{cor}

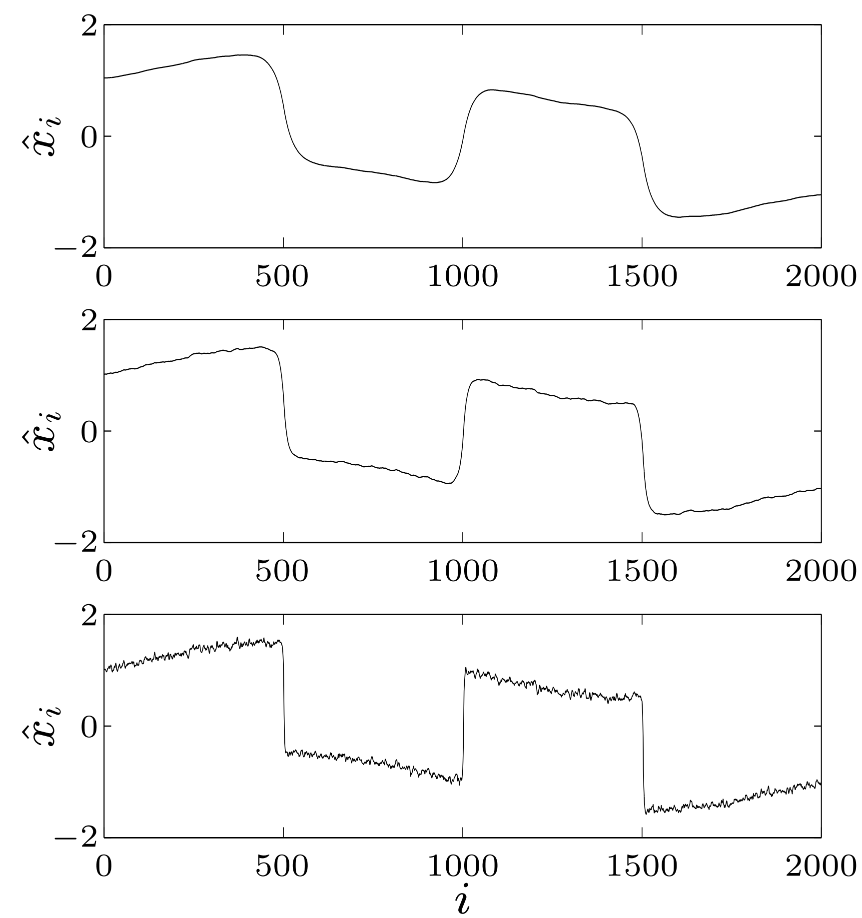


three solutions on trade-off curve $\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

Signal Reconstruction: Total Variation Smoothing



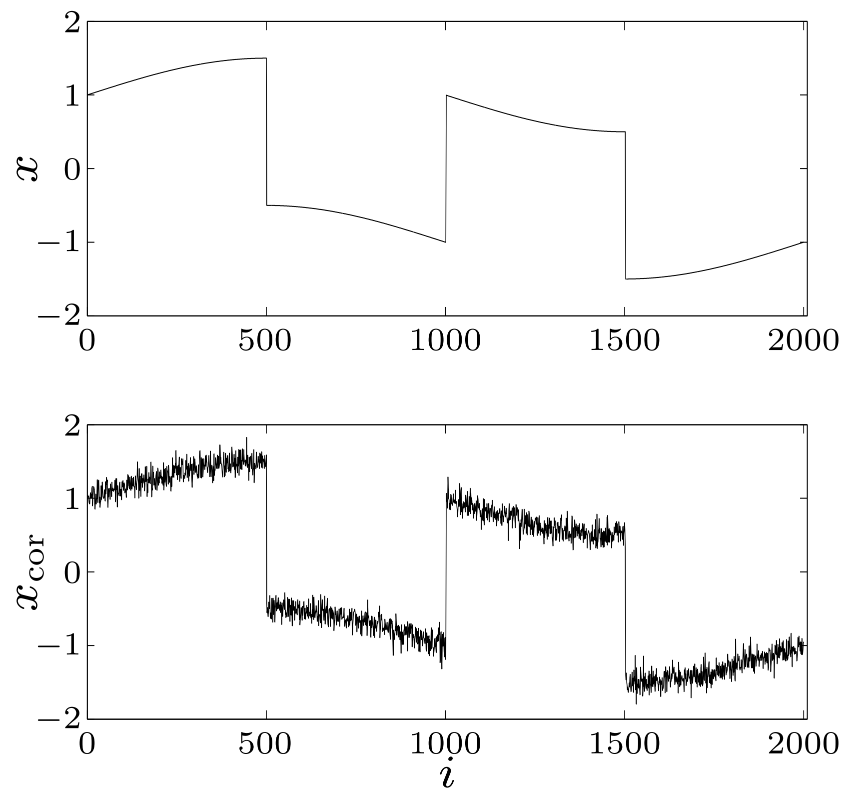
original signal x and noisy signal x_{cor}



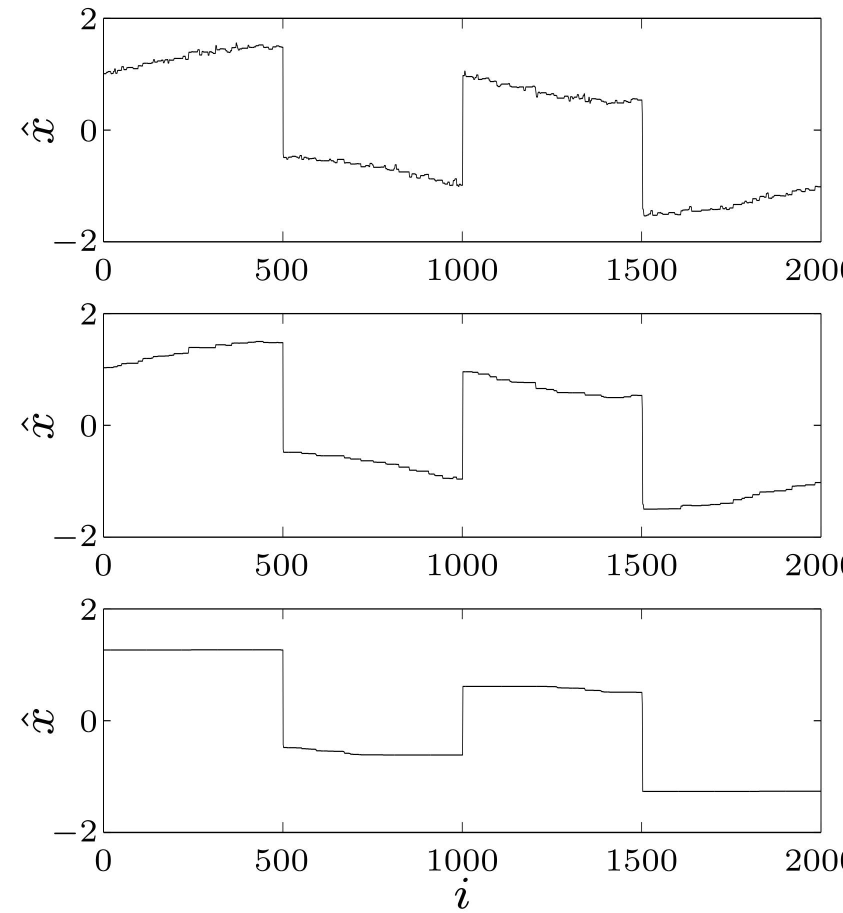
three solutions on trade-off curve $\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

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

Signal Reconstruction: Total Variation Smoothing



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

Robust Approximation

minimize $\|Ax - b\|$ with uncertain A

two approaches:

- **stochastic**: assume A is random, minimize $\mathbf{E}\,\|Ax - b\|$
- **worst-case:** set $\mathcal{A}$ of possible values of A, minimize $\sup_{A \in \mathcal{A}} \|Ax - b\|$

tractable only in special cases (certain norms $\|\cdot\|$, distributions, sets $\mathcal{A}$)

Robust Approximation

minimize $\|Ax - b\|$ with uncertain A

two approaches:

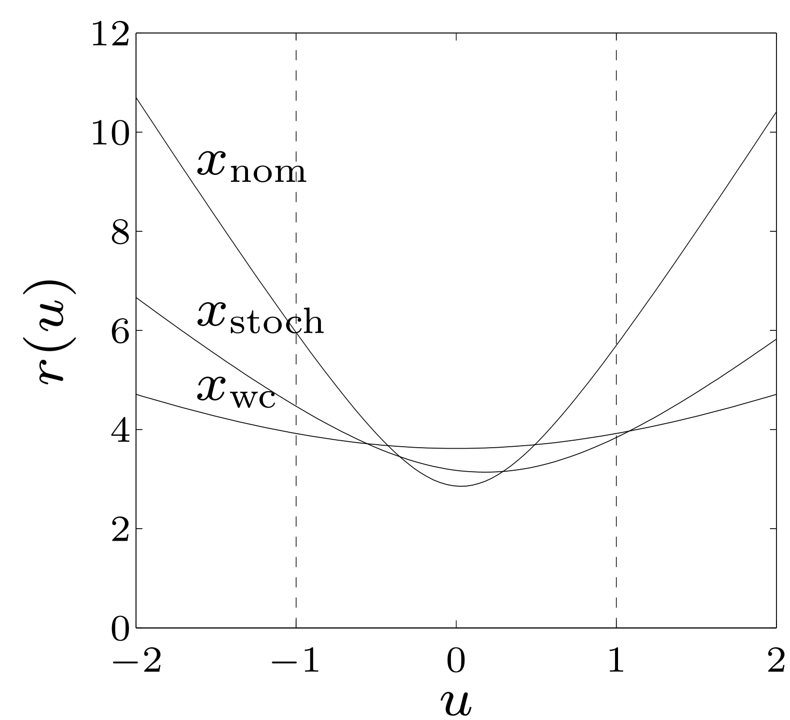
- **stochastic**: assume A is random, minimize $\mathbf{E}\,\|Ax - b\|$
- **worst-case:** set $\mathcal{A}$ of possible values of A, minimize $\sup_{A \in \mathcal{A}} \|Ax - b\|$

tractable only in special cases (certain norms $\|\cdot\|$, distributions, sets $\mathcal{A}$)

example: $A(u) = A_0 + uA_1$

- $x_{\rm nom}$ minimizes $\|A_0 x - b\|_2^2$
- $x_{\rm stoch}$ minimizes $\mathbf{E}\,\|A(u)x - b\|_2^2$ with u uniform on $[-1, 1]$
- $x_{\rm wc}$ minimizes $\sup_{-1 \le u \le 1} \|A(u)x - b\|_2^2$

figure shows $r(u) = \|A(u)x - b\|_2$



Robust Approximation: Stochastic

stochastic robust LS with $A = \bar{A} + U$, U random, $\mathbf{E}\, U = 0$, $\mathbf{E}\, U^T U = P$

$$\text{minimize} \quad \mathbf{E}\, \|(\bar{A} + U)x - b\|_2^2$$

- explicit expression for objective:

$$\begin{aligned}
\mathbf{E}\, \|Ax - b\|_2^2 &= \mathbf{E}\, \|\bar{A}x - b + Ux\|_2^2 \\
&= \|\bar{A}x - b\|_2^2 + \mathbf{E}\, x^T U^T U x \\
&= \|\bar{A}x - b\|_2^2 + x^T P x
\end{aligned}$$

- hence, robust LS problem is equivalent to LS problem

$$\text{minimize} \quad \|\bar{A}x - b\|_2^2 + \|P^{1/2}x\|_2^2$$

- for $P = \delta I$, get Tikhonov regularized problem

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

Robust Approximation: Worst-Case

worst-case robust LS with $\mathcal{A} = \{\bar{A} + u_1 A_1 + \cdots + u_p A_p \mid \|u\|_2 \leq 1\}$

$$\text{minimize} \quad \sup_{A \in \mathcal{A}} \|Ax - b\|_2^2 = \sup_{\|u\|_2 \leq 1} \|P(x)u + q(x)\|_2^2$$

where $P(x) = \left[\begin{array}{cccc} A_1 x & A_2 x & \cdots & A_p x \end{array}\right]$, $q(x) = \bar{A}x - b$

Robust Approximation: Worst-Case

worst-case robust LS with $\mathcal{A} = \{\bar{A} + u_1 A_1 + \cdots + u_p A_p \mid \|u\|_2 \le 1\}$

$$\text{minimize} \quad \sup_{A \in \mathcal{A}} \|Ax - b\|_2^2 = \sup_{\|u\|_2 \le 1} \|P(x)u + q(x)\|_2^2$$

where $P(x) = \begin{bmatrix} A_1 x & A_2 x & \cdots & A_p x \end{bmatrix}$, $q(x) = \bar{A}x - b$

- It can be shown strong duality holds between the following problems

$$\begin{array}{ll} \text{maximize} & \|Pu + q\|_2^2 \\ \text{subject to} & \|u\|_2^2 \le 1 \end{array} \qquad \begin{array}{ll} \text{minimize} & t + \lambda \\ \text{subject to} & \begin{bmatrix} I & P & q \\ P^T & \lambda I & 0 \\ q^T & 0 & t \end{bmatrix} \succeq 0 \end{array}$$

- hence, robust LS problem is equivalent to SDP

$$\begin{array}{ll} \text{minimize} & t + \lambda \\ \text{subject to} & \begin{bmatrix} I & P(x) & q(x) \\ P(x)^T & \lambda I & 0 \\ q(x)^T & 0 & t \end{bmatrix} \succeq 0 \end{array}$$

Statistical Estimation

- distribution estimation problem: estimate probability density $p(y)$ of a random variable from observed values
- parametric distribution estimation: choose from a family of densities $p_x(y)$, indexed by a parameter x

maximum likelihood estimation

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

- y is observed value
- $l(x) = \log p_x(y)$ is called log-likelihood function
- can add constraints $x \in C$ explicitly, or define $p_x(y) = 0$ for $x \notin C$
- a convex optimization problem if $\log p_x(y)$ is concave in x for fixed y

Statistical Estimation: Linear Measurements with Noise

linear measurement model

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

- $x \in \mathbf{R}^n$ is vector of unknown parameters
- v_i is IID measurement noise, with density $p(z)$
- y_i is measurement: $y \in \mathbf{R}^m$ has density $p_x(y) = \prod_{i=1}^m p(y_i - a_i^T x)$

maximum likelihood estimate: any solution x of

$$\text{maximize} \quad l(x) = \sum_{i=1}^m \log p(y_i - a_i^T x)$$

(y is observed value)

Statistical Estimation: Linear Measurements with Noise

- Gaussian noise $\mathcal{N}(0, \sigma^2)$: $p(z) = (2\pi\sigma^2)^{-1/2} e^{-z^2/(2\sigma^2)}$,

$$l(x) = -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(a_i^T x - y_i)^2$$

  ML estimate is LS solution

- Laplacian noise: $p(z) = (1/(2a))e^{-|z|/a}$,

$$l(x) = -m\log(2a) - \frac{1}{a}\sum_{i=1}^{m}|a_i^T x - y_i|$$

  ML estimate is ℓ_1-norm solution

- uniform noise on $[-a, a]$:

$$l(x) = \begin{cases} -m\log(2a) & |a_i^T x - y_i| \leq a, \quad i = 1, \ldots, m \\ -\infty & \text{otherwise} \end{cases}$$

  ML estimate is any x with $|a_i^T x - y_i| \leq a$

Statistical Estimation: Logistic Regression

random variable $y \in \{0,1\}$ with distribution

$$p = \mathbf{prob}(y = 1) = \frac{\exp(a^T u + b)}{1 + \exp(a^T u + b)}$$

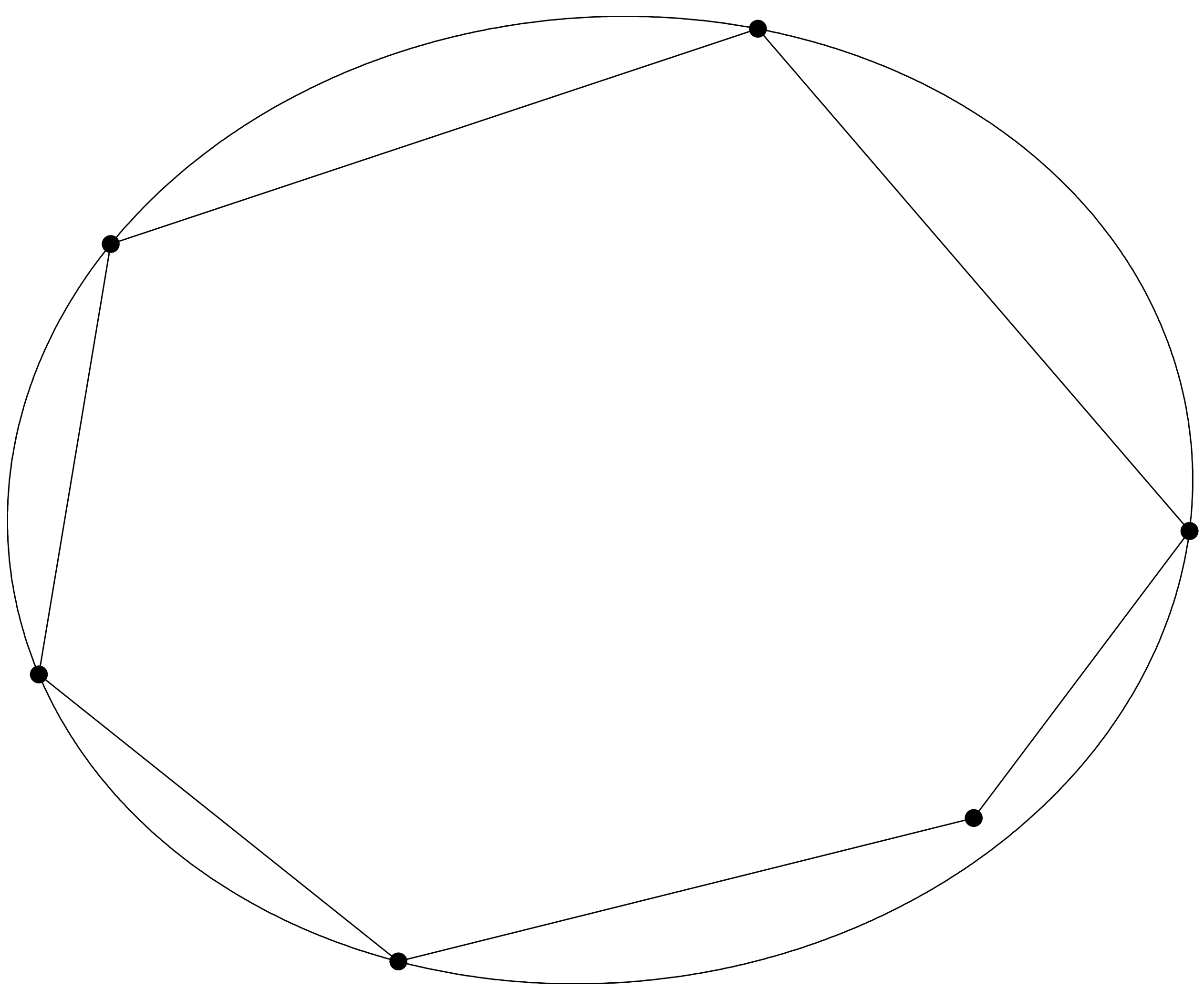
- a, b are parameters; $u \in \mathbf{R}^n$ are (observable) explanatory variables
- estimation problem: estimate a, b from m observations (u_i, y_i)

log-likelihood function (for $y_1 = \cdots = y_k = 1$, $y_{k+1} = \cdots = y_m = 0$):

$$\begin{aligned} l(a,b) &= \log\left(\prod_{i=1}^{k} \frac{\exp(a^T u_i + b)}{1 + \exp(a^T u_i + b)} \prod_{i=k+1}^{m} \frac{1}{1 + \exp(a^T u_i + b)}\right) \\ &= \sum_{i=1}^{k}(a^T u_i + b) - \sum_{i=1}^{m} \log(1 + \exp(a^T u_i + b)) \end{aligned}$$

concave in a, b

Minimum Volume Ellipsoid Around a Set



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)

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,

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

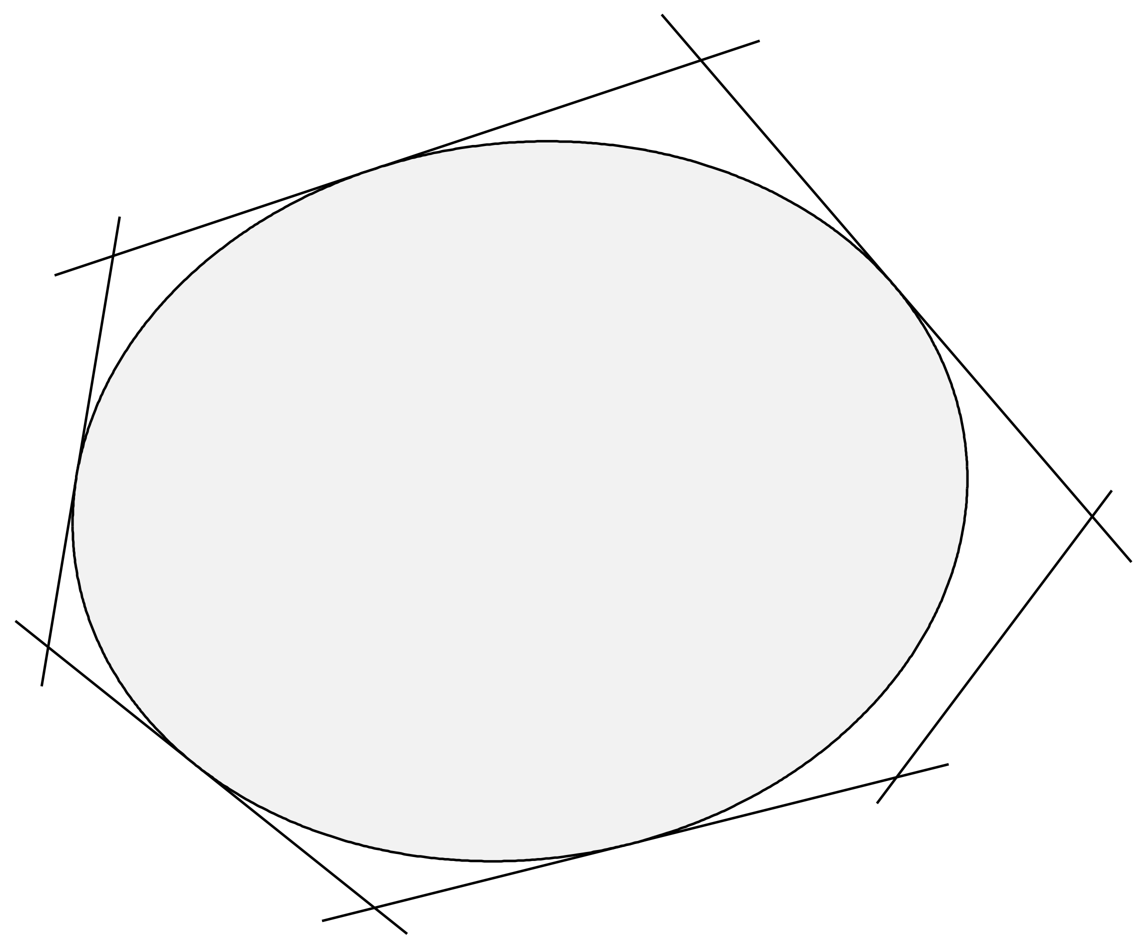
convex, but evaluating the constraint can be hard (for general C)

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

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

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

Maximum Volume Inscribed Ellipsoid



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)

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