Approximation and fitting

Norm Approximation

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

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

$b \in \mathcal{R}(A)$ $\longleftrightarrow$ optimal value is zero

Approximation interpretation

$a_1, \ldots, a_n \in \mathbf{R}^m$ are the columns of A

approximate the vector b by a linear combination of the columns of A

regressors

By expressing Ax as

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

regression of b

The approximation problem is also called the regression problem.

Norm Approximation

Estimation interpretation

Linear measurement model $$y = Ax + v$$

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

unknown, but presumed to be small

Smaller values of v are more plausible than larger values

v has the value $y - A\hat{x}$

$$\hat{x} = \mathrm{argmin}_z ||Az - y||.$$

Norm Approximation

Geometric l interpretation

subspace $\mathcal{A} = \mathcal{R}(A) \subseteq \mathbf{R}^m$

point $b \in \mathbf{R}^m$

projection of
the point b onto the subspace $\mathcal{A}$, in the norm $||\cdot||$

$$\begin{array}{ll} \text{minimize} & ||u - b|| \\ \text{subject to} & u \in \mathcal{A}. \end{array} \iff \text{minimize} \quad ||Ax - b||$$

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

$Ax^\star$ is point in $\mathcal{R}(A)$ closest to b

Norm Approximation

Design interpretation

The n variables $x_1, \ldots, x_n$ are *design variables* (input), Ax is result (output)

The vector b is a vector of *target* or *desired results*.

Goal: choose a vector of design variables that achieves, as closely as possible, the desired results

$$Ax \approx b$$

Norm Approximation

Least-squares approximation

$$\text{minimize} \quad \|Ax - b\|_2^2 = r_1^2 + r_2^2 + \cdots + r_m^2,$$

$$r = Ax - b$$

$$f(x) = x^T A^T A x - 2b^T A x + b^T b.$$

$$\nabla f(x) = 2A^T A x - 2A^T b = 0,$$

$$\boxed{A^T A x = A^T b}$$

$$x^\star = (A^T A)^{-1} A^T b$$

Chebyshev or minimax approximation

$$\text{minimize} \quad \|Ax - b\|_\infty = \max\{|r_1|, \ldots, |r_m|\}$$

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

$x \in \mathbf{R}^n$ and $t \in \mathbf{R}$.

Norm Approximation

Sum of absolute residuals approximation

$$\text{minimize} \quad \|Ax - b\|_1 = |r_1| + \cdots + |r_m|$$

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

$x \in \mathbf{R}^n$ and $t \in \mathbf{R}^m$

Penalty function approximation

minimize $\quad \|Ax - b\|$ $\qquad$ for $1 \leq p < \infty$ $\qquad$ $(|r_1|^p + \cdots + |r_m|^p)^{1/p}$

generalization of the ℓ_p-norm approximation problem

penalty function approximation problem

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

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

penalty function assesses a cost or penalty for each component of residual

Penalty function approximation

Some common penalty functions

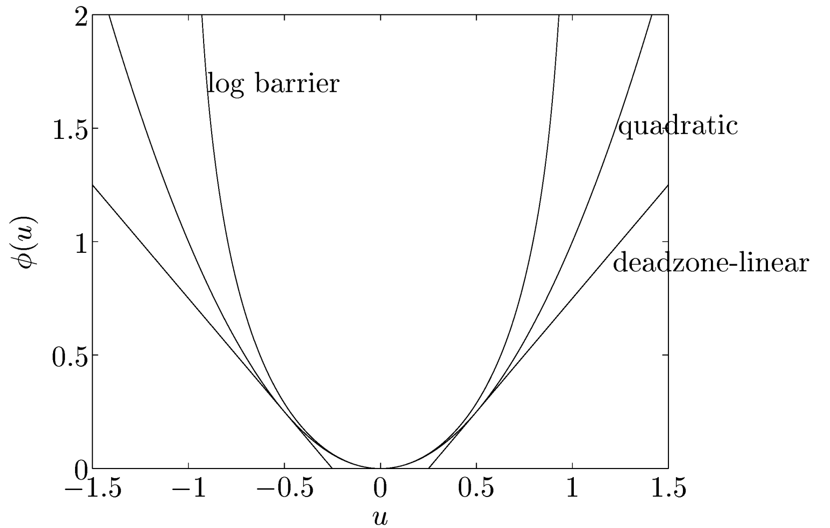
- $\phi(u) = |u|^p$, where $p \geq 1$ $\longrightarrow$ ℓ_p-norm approximation problem

- deadzone-linear with width a:

$$\phi(u) = \begin{cases} 0 & |u| \leq a \\ |u| - a & |u| > a. \end{cases}$$

- log-barrier with limit a:

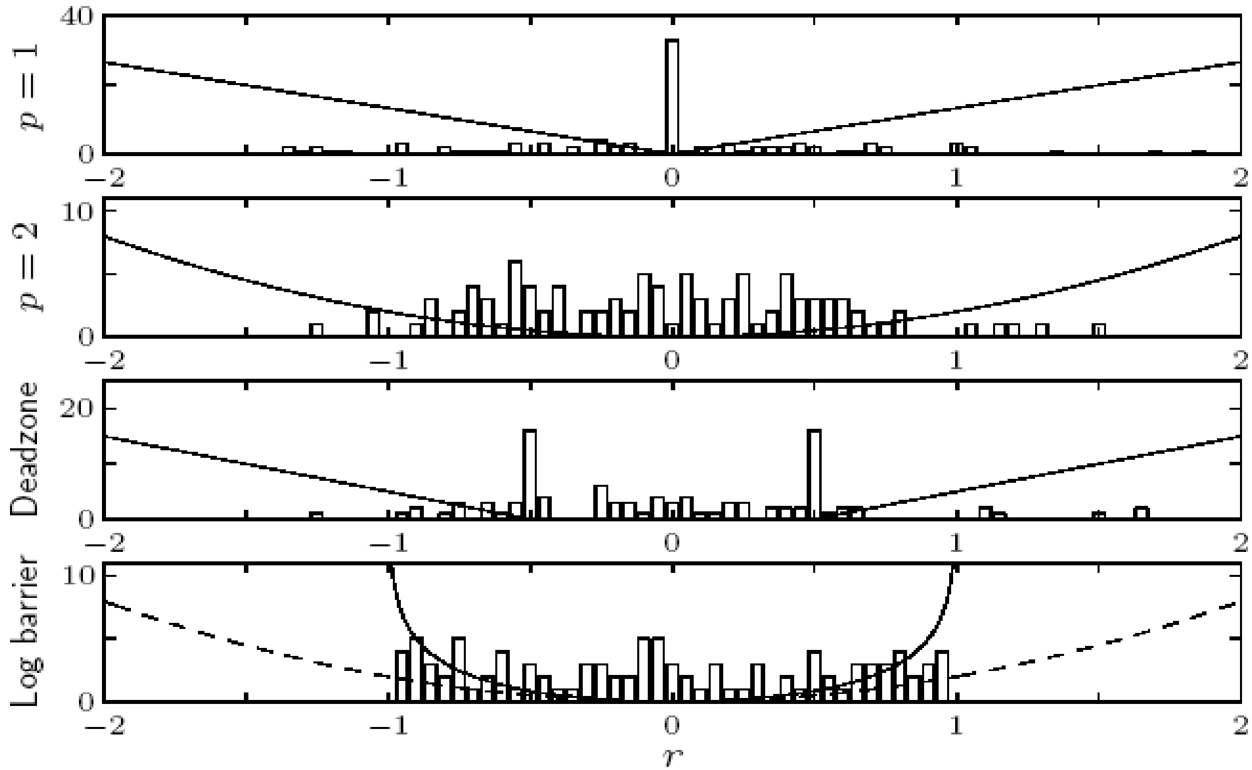
$$\phi(u) = \begin{cases} -a^2 \log(1 - (u/a)^2) & |u| < a \\ \infty & |u| \geq a. \end{cases}$$



Penalty function approximation

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

Least-norm problems

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

underdetermined

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

When $m = n$, the only feasible point is $x = A^{-1}b$;

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

Geometric interpretation

$x^\star$ is point in affine set $\{x \mid Ax = b\}$ with minimum distance to 0

Least-norm problems

Estimation interpretation

✓ x is a vector of parameters to be estimated.

$$Ax = b$$

✓ We have $m < n$ perfect (noise free) linear measurements

✓ Any parameter vector x that satisfies Ax = b is consistent with our measurements.

✓ without taking further measurements, use prior information

✓ x is more likely to be small than large.

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

✓ Smallest (most plausible) estimate consistent with measurements

Least-norm problems

Control or Design interpretation

x are design variables (inputs);

b are required results (outputs)

$Ax = b$ represent m *specifications* or *requirements* on the design

$n - m$ degrees of freedom

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

Least-norm problems

Least-squares solution of linear equations

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

can be solved via optimality conditions

dual variable $\nu \in \mathbf{R}^m$

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

$$x^\star = -(1/2)A^T\nu^\star \qquad -(1/2)AA^T\nu^\star = b$$

$$\nu^\star = -2(AA^T)^{-1}b, \qquad x^\star = A^T(AA^T)^{-1}b.$$

Least-norm problems

Sparse solutions via least ℓ_1-norm

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

tends to produce sparse solution $x^\star$

can be solved as an LP

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

Least-penalty problems

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

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

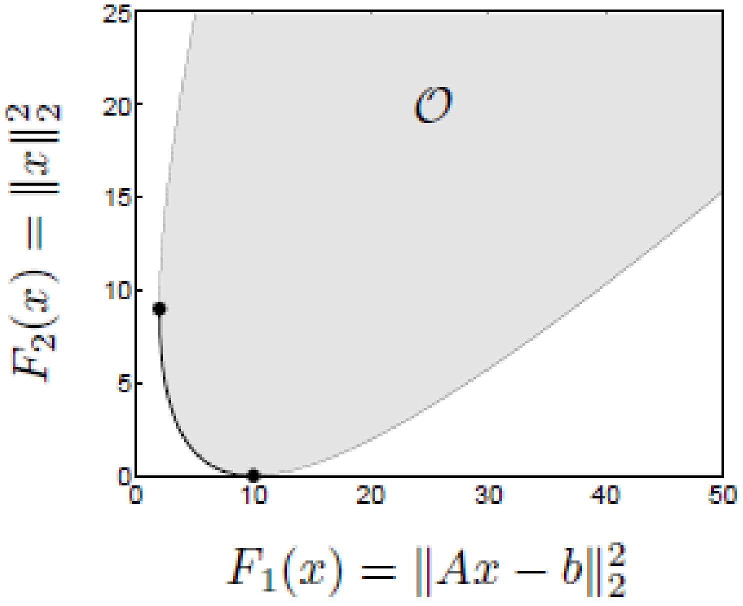
$A \in \mathbf{R}^{m \times n}$, norms on $\mathbf{R}^m$ and $\mathbf{R}^n$ can be different

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

Multi-objective optimization

to find Pareto optimal points: choose $\lambda \succ 0$ and solve scalar problem



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

example for $A \in \mathbf{R}^{100\times 10}$; heavy line is formed by Pareto optimal points

Regularization -Scalarized problem

Regularization is a common scalarization method used to solve the bi-criterion 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

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

$$= x^T(A^TA + \delta I)x - 2b^TAx + b^Tb.$$

$$x = (A^TA + \delta I)^{-1}A^Tb.$$

Optimal input design

linear dynamical system with impulse response h:

$$y(t)=\sum_{\tau=0}^{t} h(\tau)u(t-\tau), \quad t=0,1,\ldots,N.$$

input design problem: multicriterion problem with 3 objectives

1. tracking error with desired output y_{des}: $J_{\text{track}}=\sum_{t=0}^{N}(y(t)-y_{\text{des}}(t))^2$
2. input magnitude: $J_{\text{mag}}=\sum_{t=0}^{N}u(t)^2$
3. input variation: $J_{\text{der}}=\sum_{t=0}^{N-1}(u(t+1)-u(t))^2$

track desired output using a small and slowly varying input signal

$$\text{minimize} \quad J_{\text{track}}+\delta J_{\text{der}}+\eta J_{\text{mag}}$$

Optimal input design

$$J_{\text{track}} + \delta J_{\text{der}} + \eta J_{\text{mag}},$$

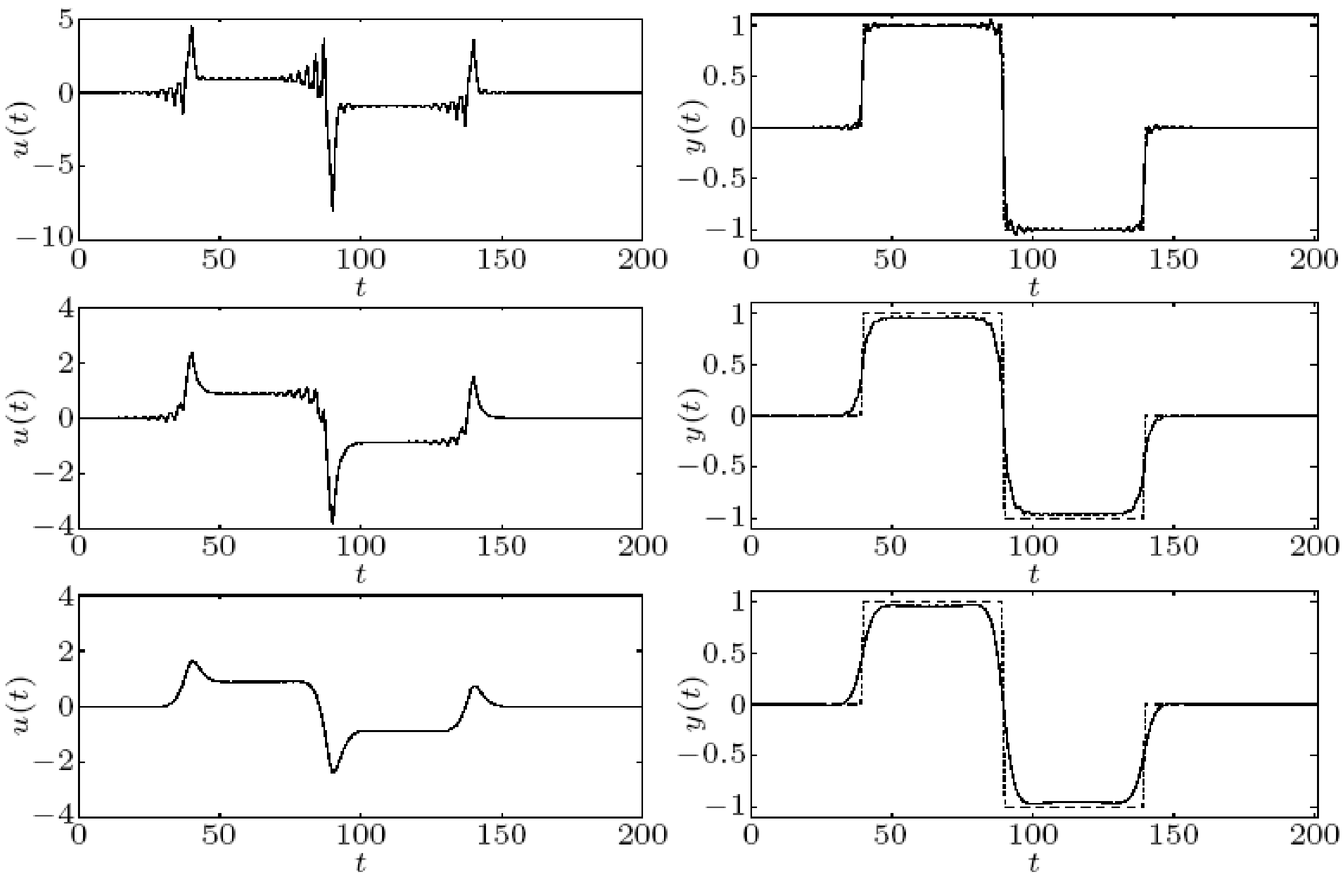
$$h(t) = \frac{1}{9}(0.9)^t(1 - 0.4\cos(2t))$$

$$N = 200$$

$\delta = 0,\ \eta = 0.005$

$\delta = 0,\ \eta = 0.05$

$\delta = 0.3,\ \eta = 0.05$



Reconstruction, smoothing, and de-noising

$$x_{\text{cor}} = x + v.$$

✓Noise is unknown, small, and, rapidly varying.
✓Signal does not vary too rapidly.

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

- variable $\hat{x}$ (reconstructed signal) is estimate of x
- $\phi : \mathbf{R}^n \to \mathbf{R}$ is regularization function or smoothing objective

measure the roughness, or lack of smoothness, of the estimate $\hat{x}$

1. Quadratic smoothing

$$\phi_{\text{quad}}(x) = \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 = \|Dx\|_2^2,$$

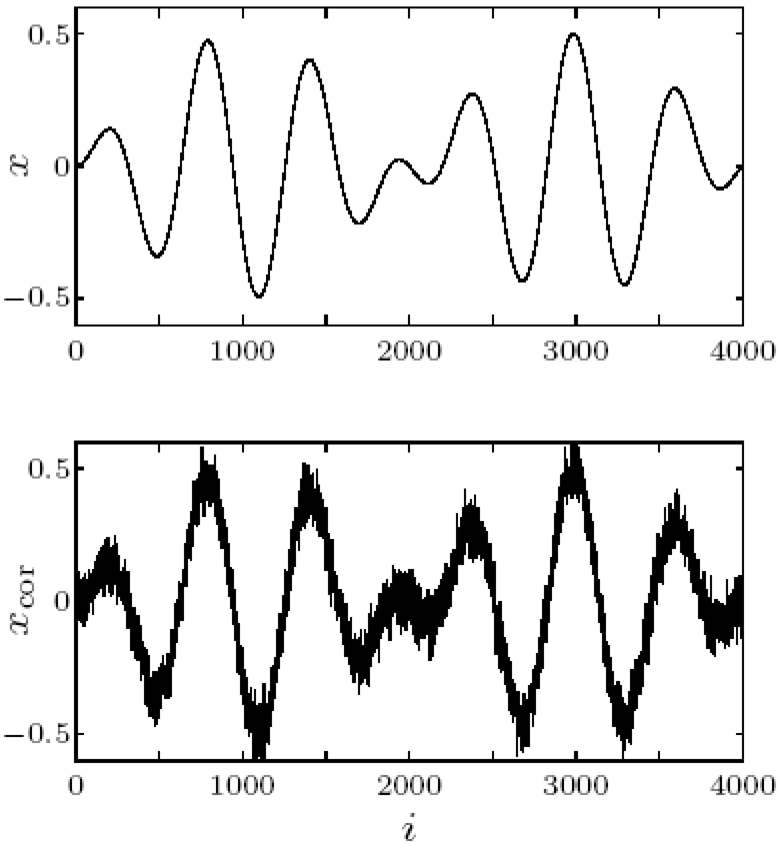
$D \in \mathbf{R}^{(n-1)\times n}$

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

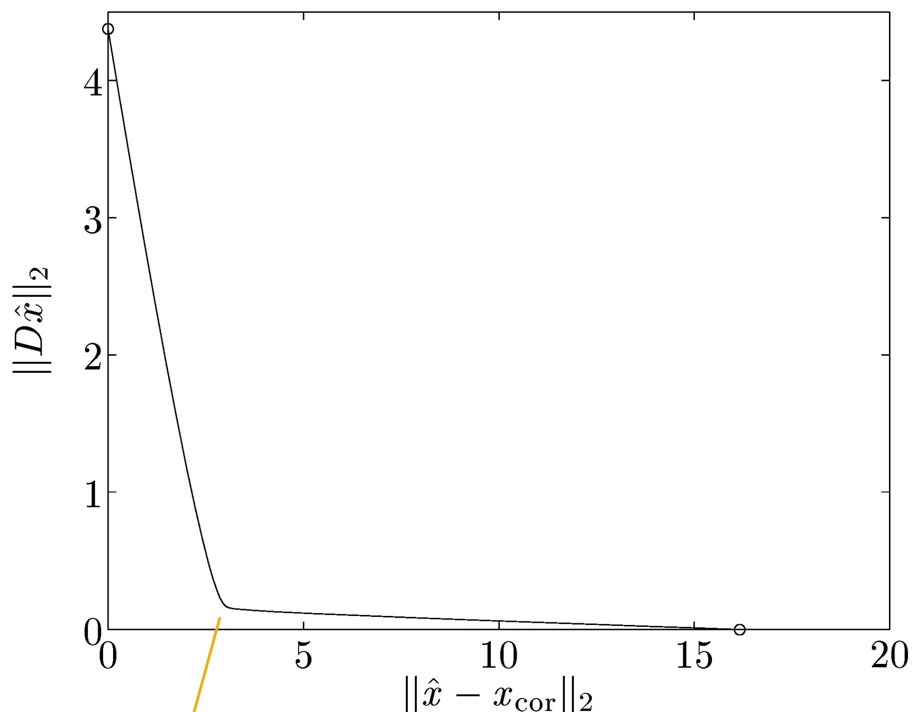
$$\min \quad \|\hat{x} - x_{\text{cor}}\|_2^2 + \delta \|D\hat{x}\|_2^2,$$

$$\hat{x} = (I + \delta D^T D)^{-1} x_{\text{cor}},$$

1. Quadratic smoothing

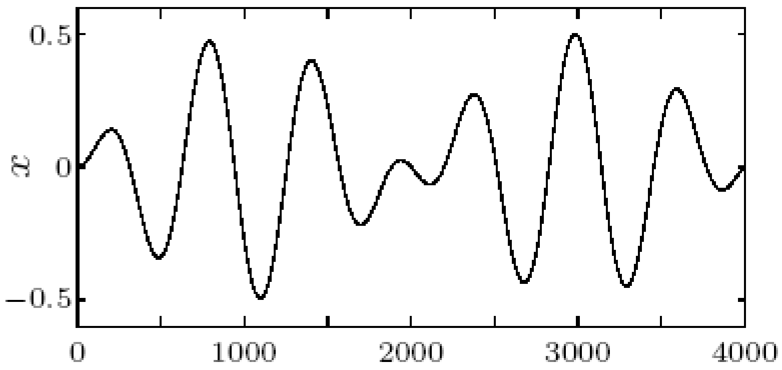


original signal x and noisy signal $x_{\rm cor}$



a clear knee near $\|\hat{x} - x_{\rm cor}\| \approx 3$.

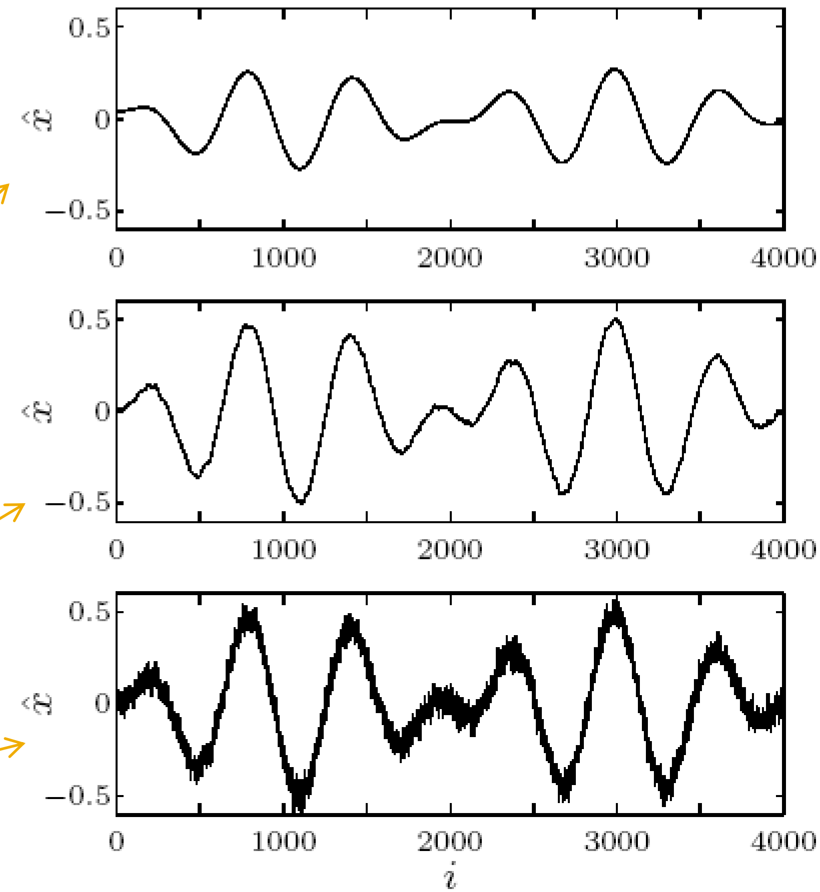
1. Quadratic smoothing



$$\|\hat{x} - x_{\mathrm{cor}}\|_2 = 8$$

$$\|\hat{x} - x_{\mathrm{cor}}\|_2 = 3$$

$$\|\hat{x} - x_{\mathrm{cor}}\|_2 = 1$$



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

2. Total variation smoothing

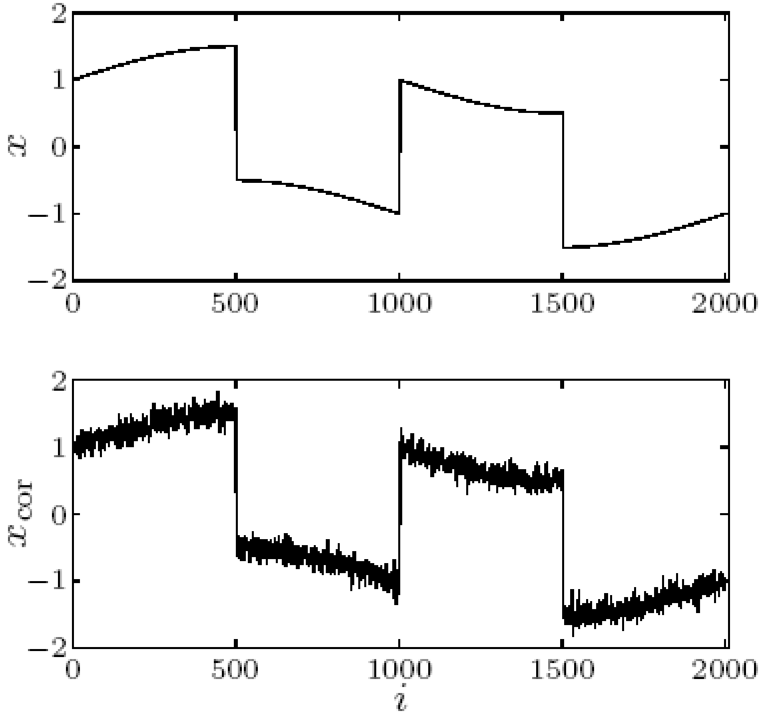
2. Total variation smoothing

✓ quadratic smoothing works well when the original signal is very smooth, and the noise is rapidly varying.

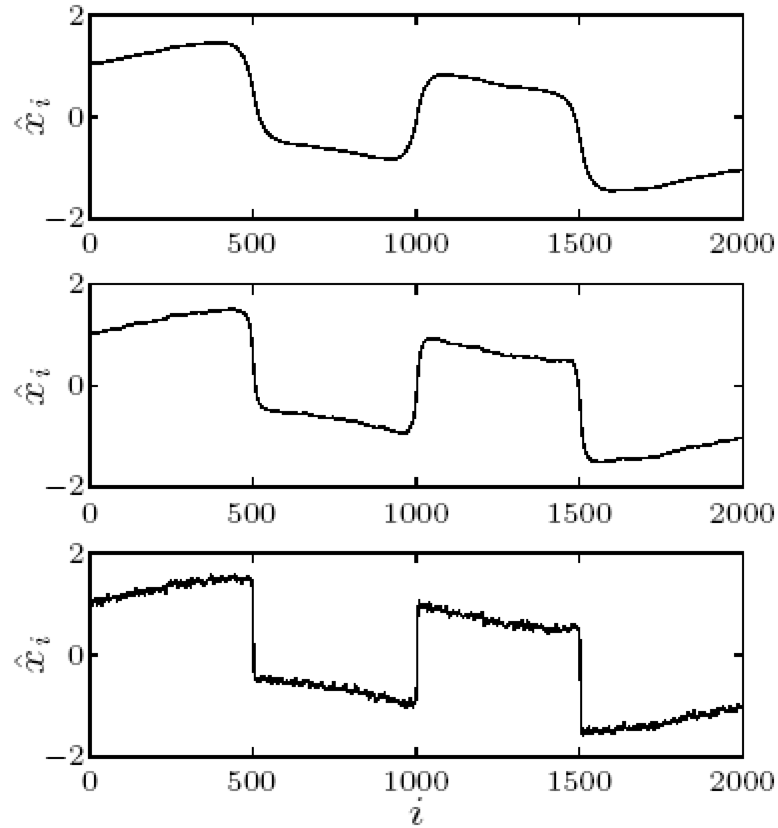
✓any rapid variations in the original signal will be attenuated or removed by quadratic smoothing

$$\phi_{\mathrm{tv}}(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$$

Compare quad and tv

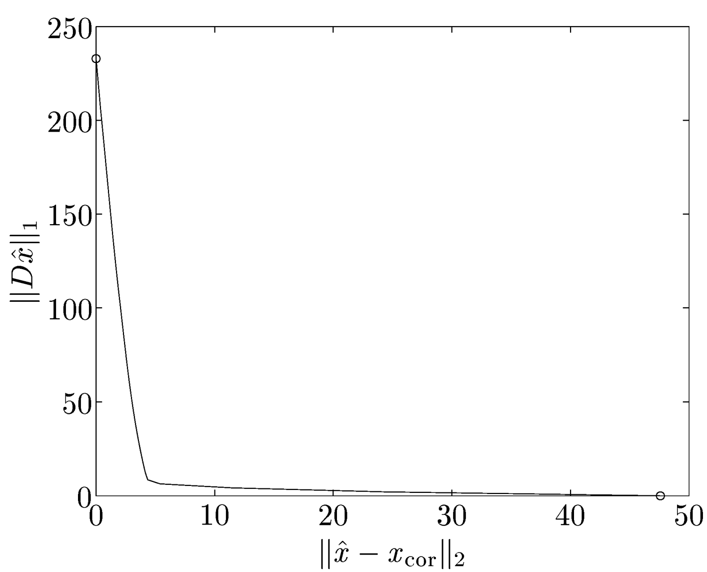


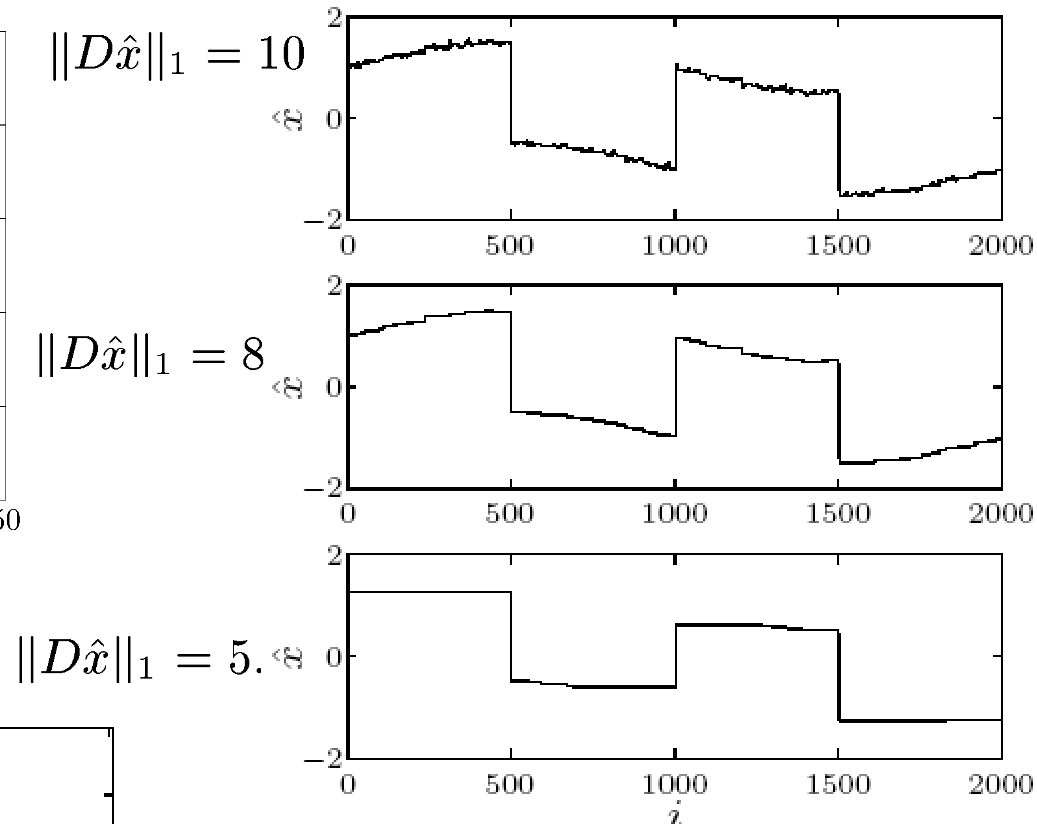
original signal x and noisy signal x_{cor}

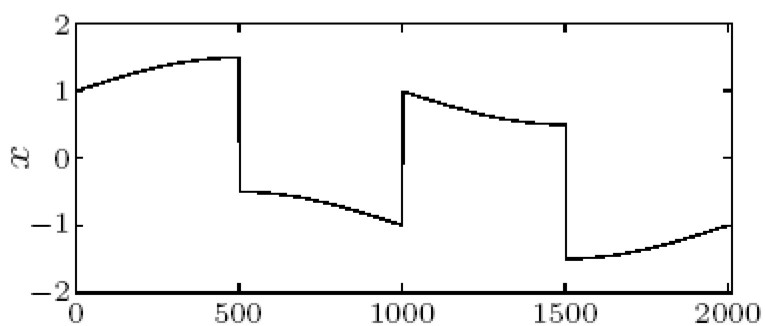


three solutions on trade-off curve $\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

Compare quad and tv







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

Robust approximation

some uncertainty or possible variation in the data matrix A.

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

1. Stochastic robust approximation

assume that A is a random variable taking values in $\mathbf{R}^{m\times n}$, with mean $\bar{A}$

$$A = \bar{A} + U,$$

U is a random matrix with zero mean

$$\text{minimize} \quad \mathbf{E}\,\|Ax - b\|.$$

Stochastic robust approximation

Special cases

$$\mathbf{prob}(A = A_i) = p_i, \quad i = 1, \dots, k,$$

$$\text{minimize} \quad p_1\|A_1x - b\| + \cdots + p_k\|A_kx - b\|,$$

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

$$\begin{aligned} \mathbf{E}\,\|Ax - b\|_2^2 &= \mathbf{E}(\bar{A}x - b + Ux)^T(\bar{A}x - b + Ux) \\ &= (\bar{A}x - b)^T(\bar{A}x - b) + \mathbf{E}\,x^TU^TUx \\ &= \|\bar{A}x - b\|_2^2 + x^TPx, \end{aligned}$$

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

$$x = (\bar{A}^T\bar{A} + P)^{-1}\bar{A}^Tb.$$

Robust approximation

2. Worst-case robust approximation

describe the uncertainty by a set of possible values for A: $A \in \mathcal{A} \subseteq \mathbf{R}^{m \times n}$,

$$\text{minimize} \quad e_{\mathrm{wc}}(x) = \sup\{\|Ax - b\| \mid A \in \mathcal{A}\},$$

Comparison

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

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

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