# Convex Optimization selections from Chapter 6

Audrey Hesse

# Approximation & Fitting

- Norm Approximation
  - Penalty functions
- Least-norm problems
- Regularized Approximation
- Signal Reconstruction

# Norm Approximation

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

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

- **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 $b$

- A solution to norm approximation is sometimes an approximate solution

$$Ax \approx b$$

- Residual: vector $r = Ax - b$
- Residuals: the individual components of the residual associated with x

## examples

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

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

Minimize the sum of the squares of the residuals $\| Ax - b \|_2^2 = r_1^2 + r_2^2 + \ldots + r_m^2$

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

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

Minimize: $\| Ax - b \|_\infty = \max\{| r_1 |, \ldots, | r_m |\}$

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

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

Minimize: $\| Ax - b \|_1 = | r_1 | + | r_2 | + \ldots + | r_m |$

# 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)

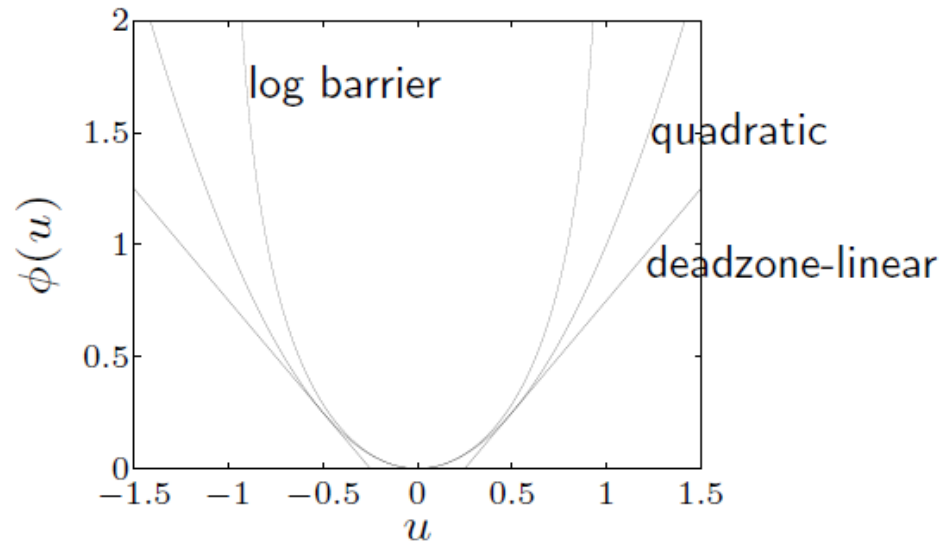
**examples**

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

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

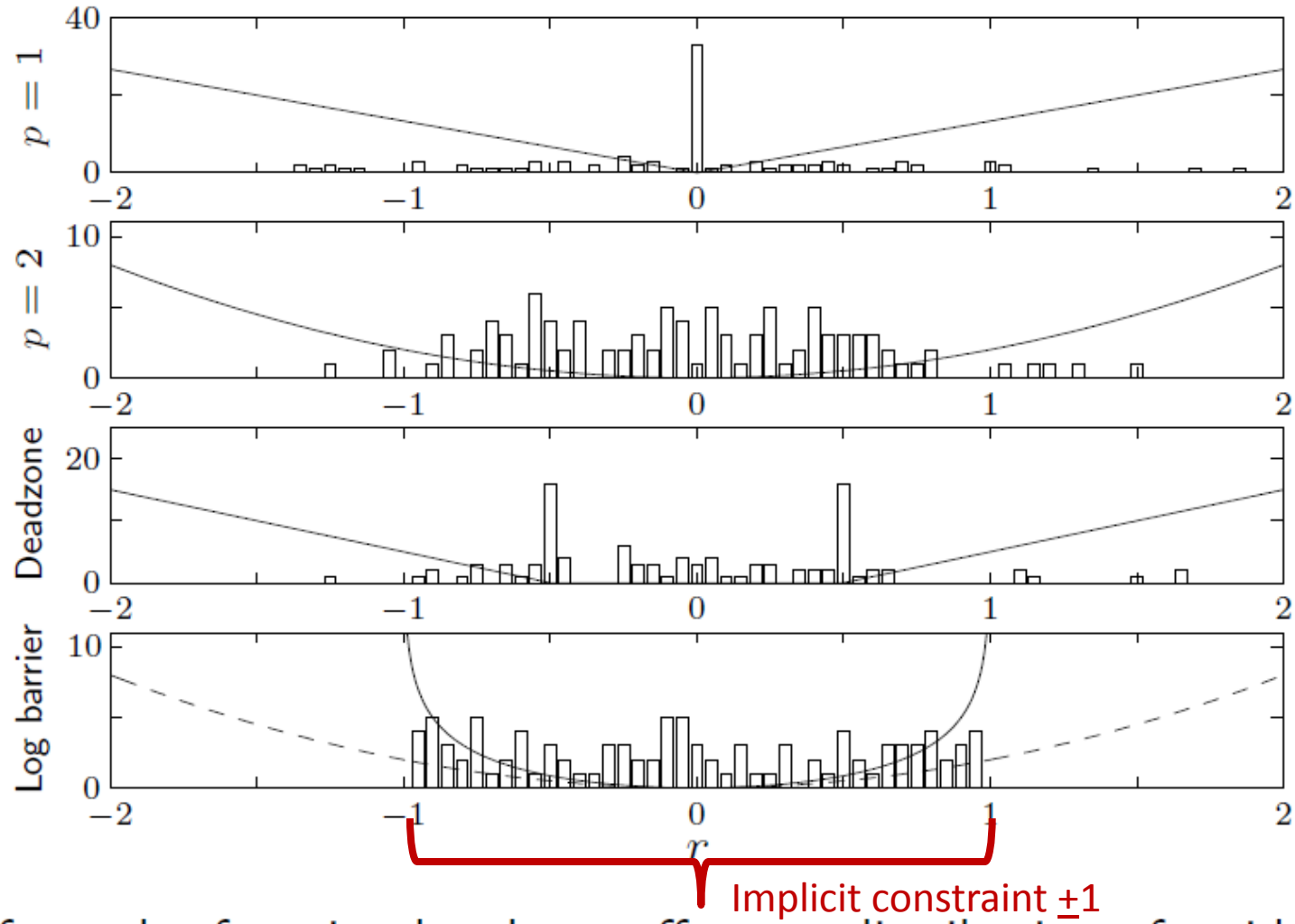
- log-barrier with limit $a$:

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



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

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

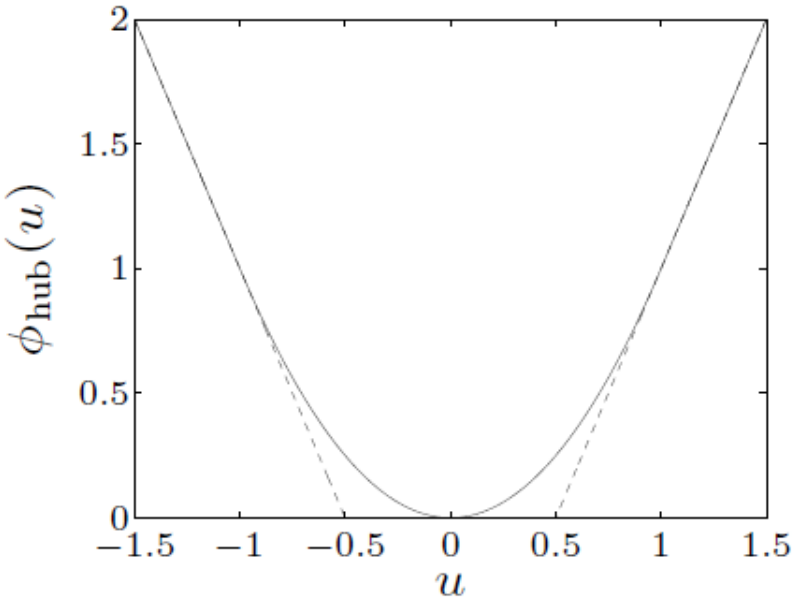


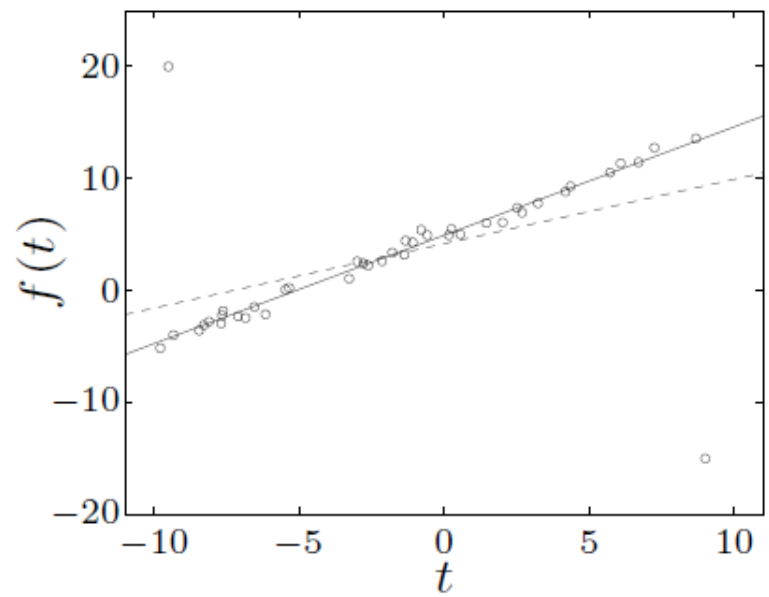

shape of penalty function has large effect on distribution of residuals

**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 $x^\star = \operatorname{argmin}_{Ax=b} \|x\|$:

- **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

$$\text{minimize} \ \| x \|_2^2$$
$$\text{subject to}: Ax = b$$
$$f = x^2 + \nu^T (Ax - b)$$

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

# 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 $y_{\text{des}}$: $J_{\text{track}} = \sum_{t=0}^{N}(y(t) - y_{\text{des}}(t))^2$
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

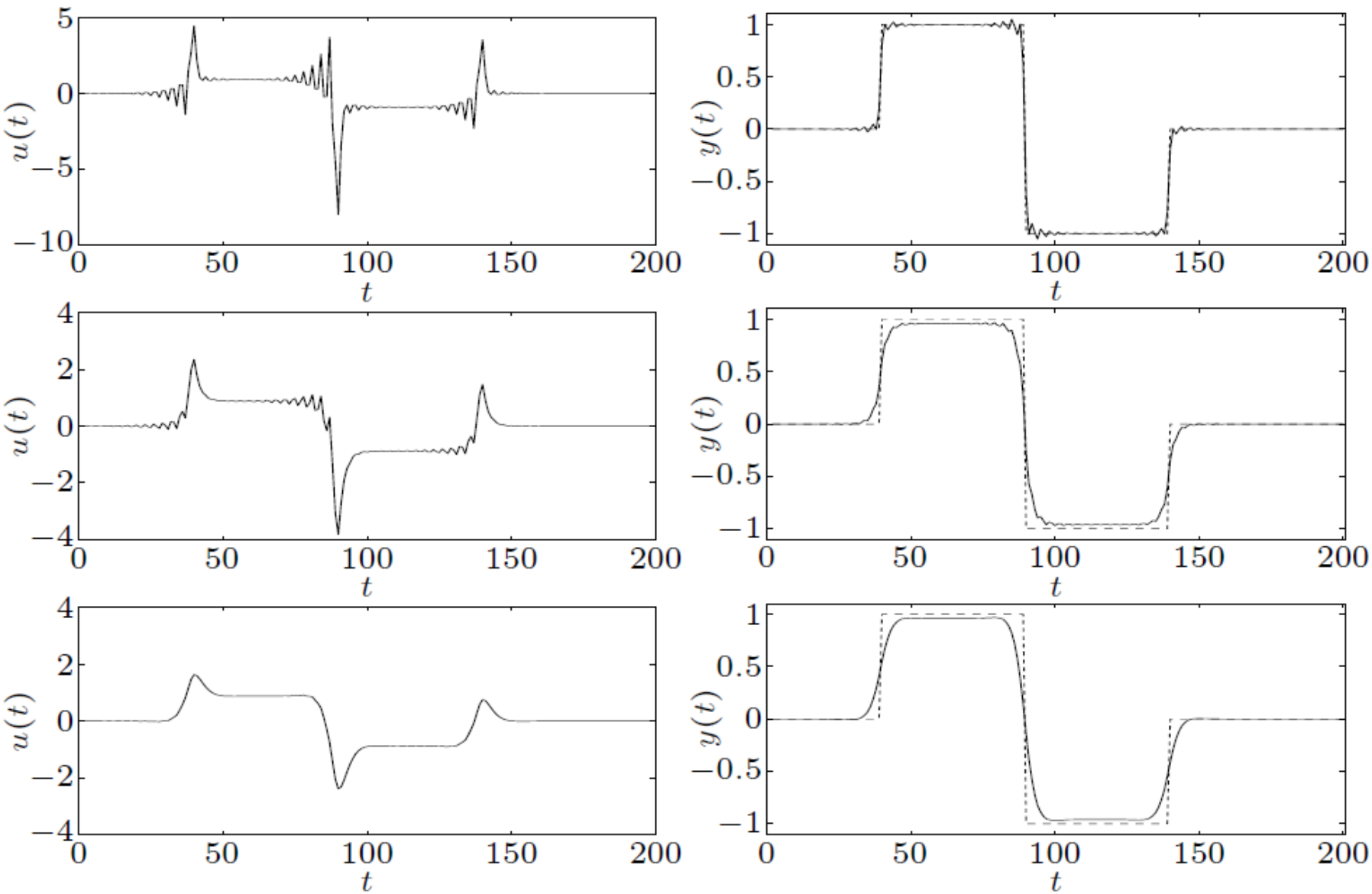
**regularized least-squares formulation**

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

for fixed $\delta, \eta$, a least-squares problem in $u(0), \ldots, u(N)$

**example**: 3 solutions on optimal trade-off surface

(top) $\delta = 0$, small $\eta$; (middle) $\delta = 0$, larger $\eta$; (bottom) large $\delta$



# Signal reconstruction

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

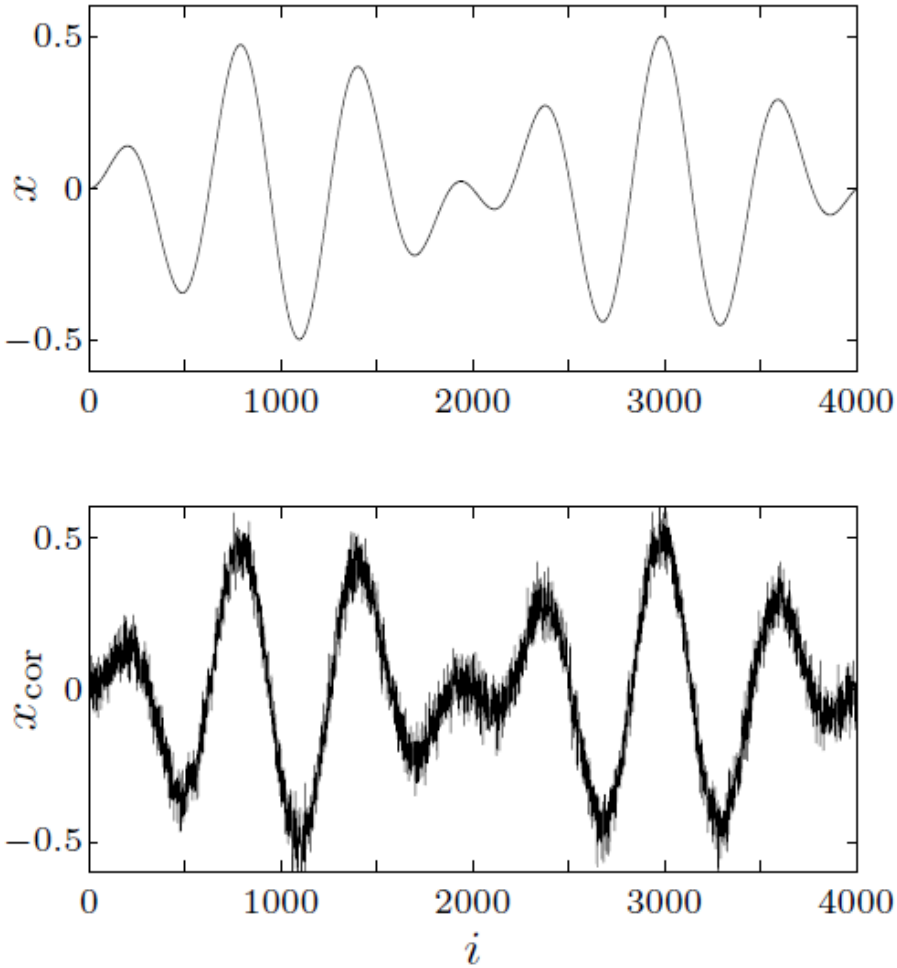
- $x \in \mathbf{R}^n$ is unknown signal
- $x_{\text{cor}} = x + v$ is (known) corrupted version of $x$, with additive noise $v$
- variable $\hat{x}$ (reconstructed signal) is estimate of $x$
- $\phi : \mathbf{R}^n \to \mathbf{R}$ is regularization function or smoothing objective

**examples:** quadratic smoothing, total variation smoothing:

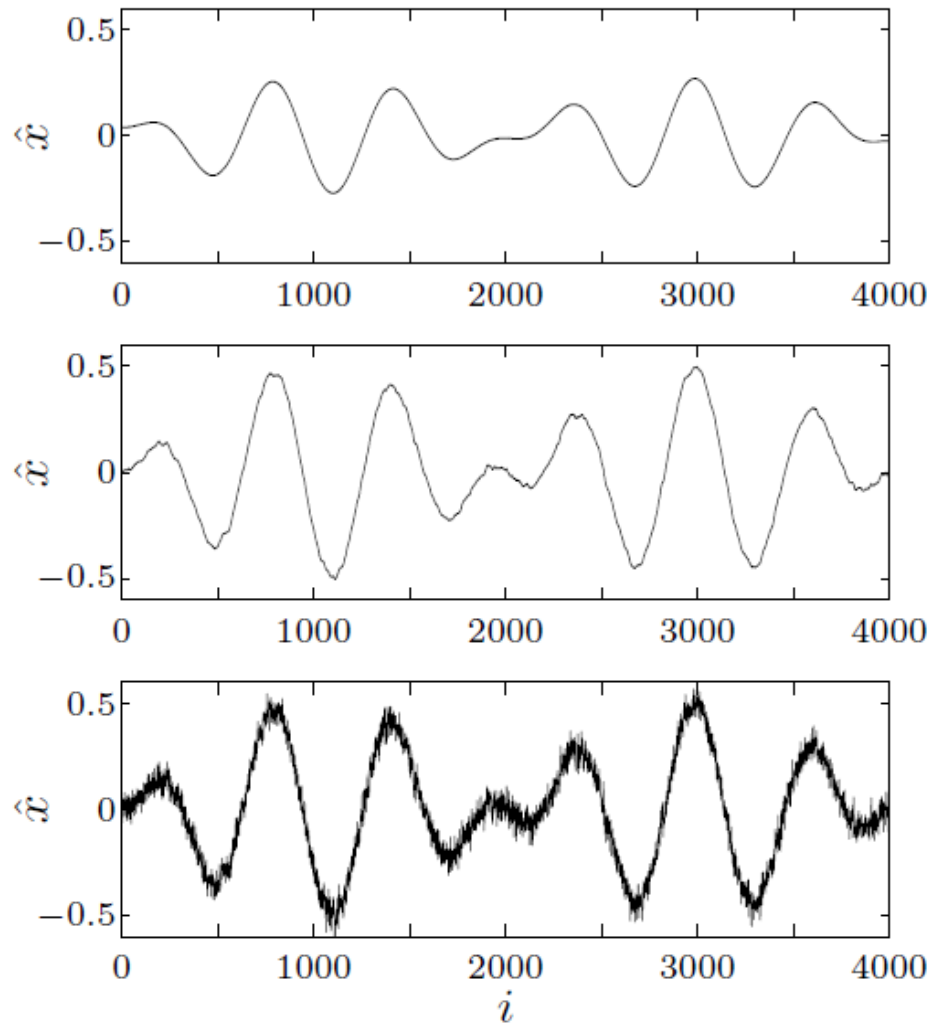
$$\phi_{\text{quad}}(\hat{x}) = \sum_{i=1}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2, \qquad \phi_{\text{tv}}(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$$

# quadratic smoothing example

$x \in R^{4000}$

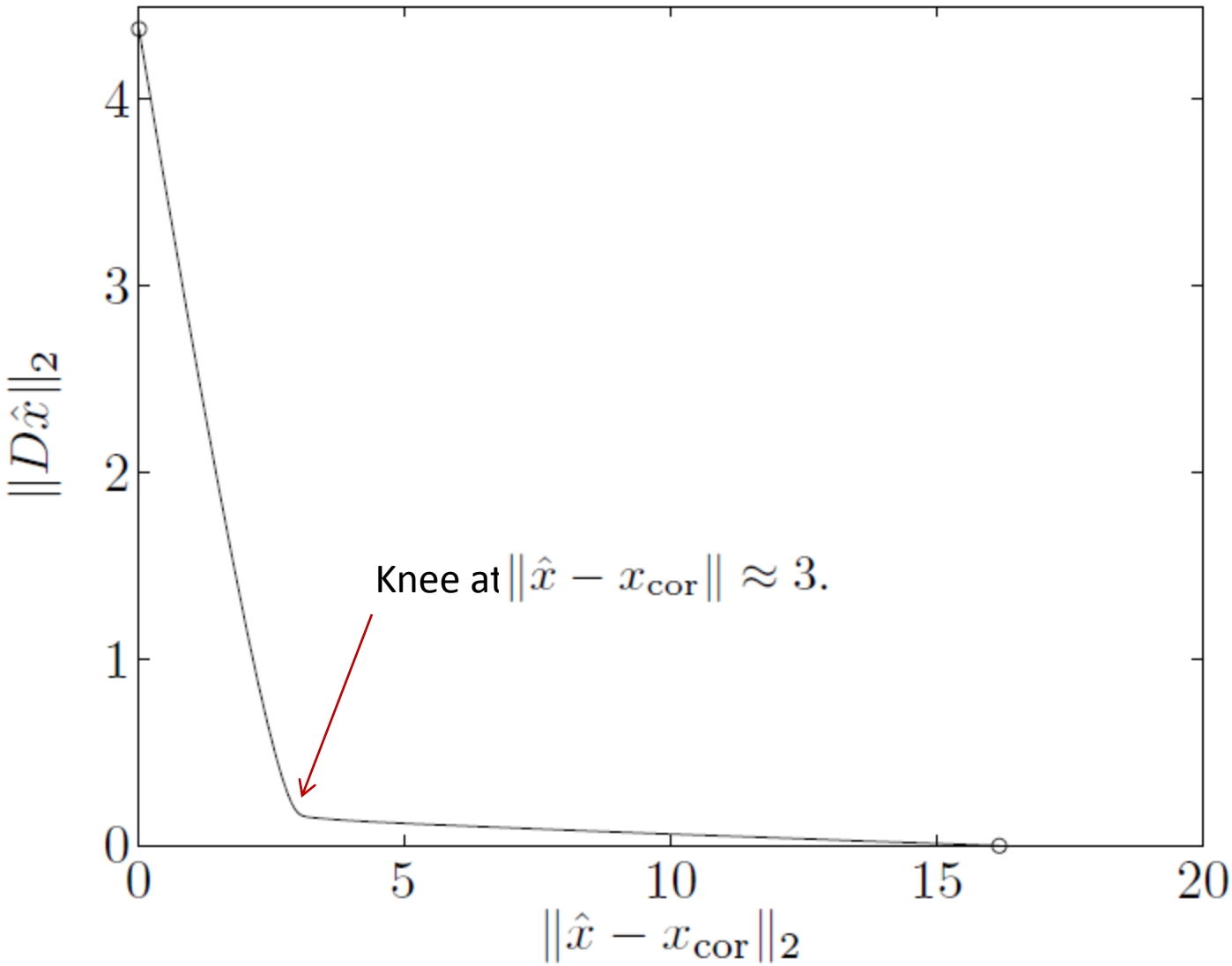



original signal $x$ and noisy signal $x_{\text{cor}}$

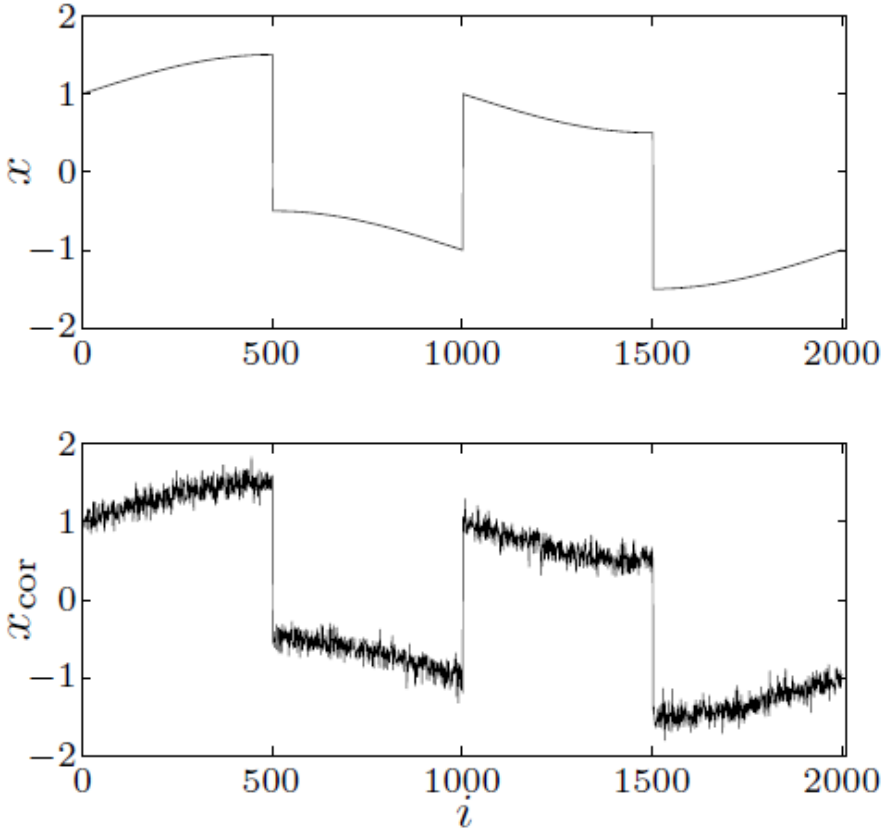



three solutions on trade-off curve $\|\hat{x} - x_{\text{cor}}\|_2$ versus $\phi_{\text{quad}}(\hat{x})$

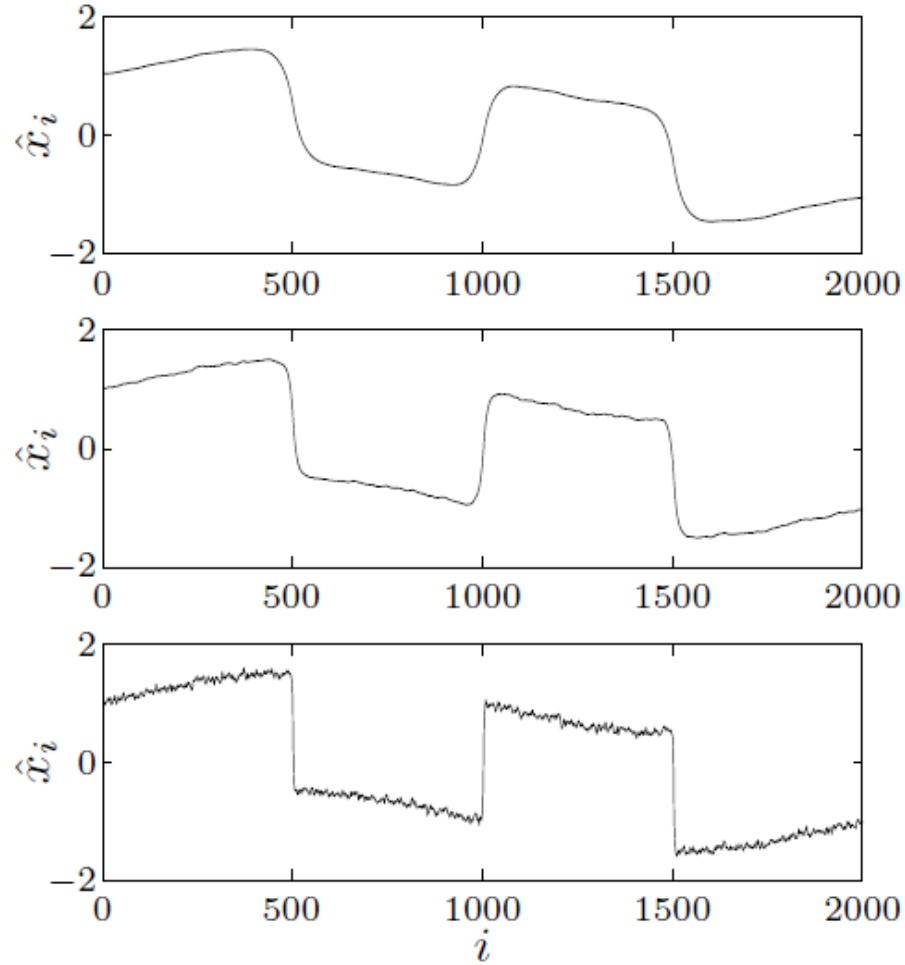
# Optimal trade-off curve



# total variation reconstruction example

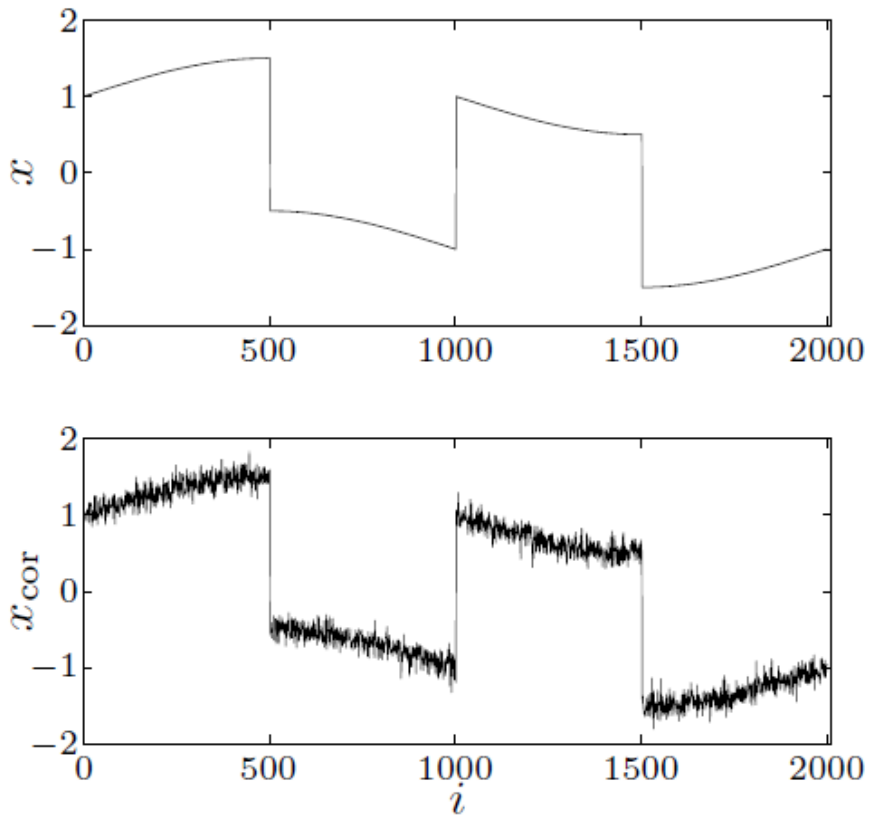



original signal $x$ and noisy signal $x_{\text{cor}}$

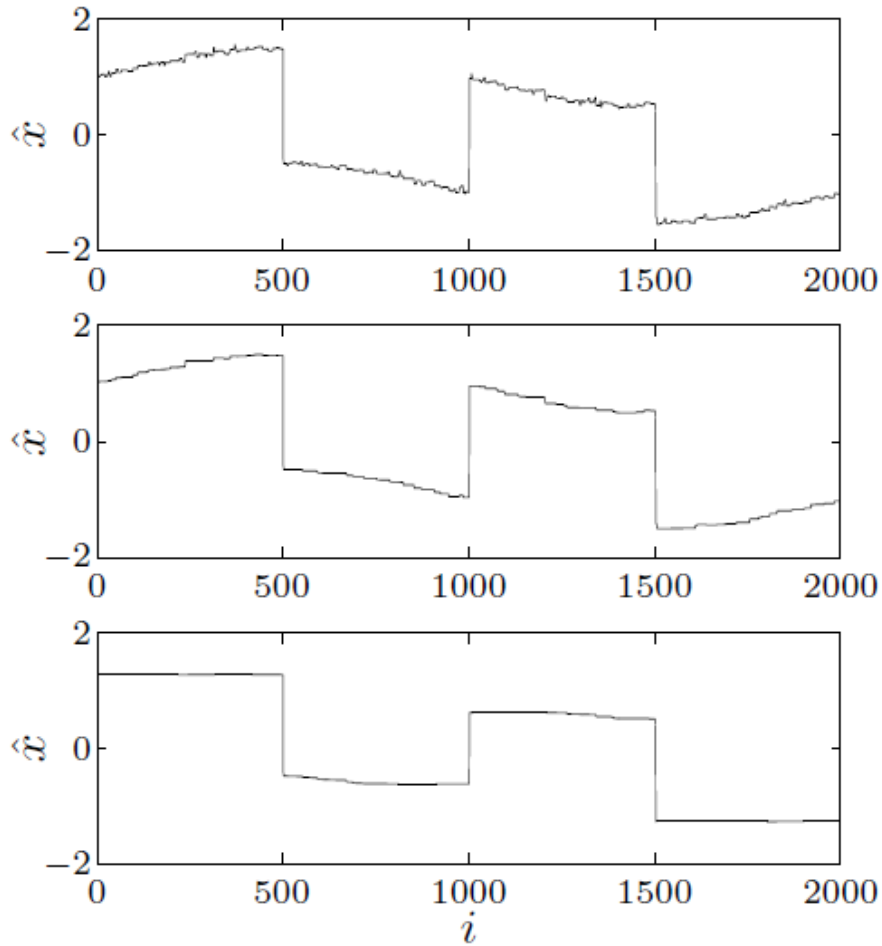



three solutions on trade-off curve $\|\hat{x} - x_{\text{cor}}\|_2$ versus $\phi_{\text{quad}}(\hat{x})$

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




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




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

total variation smoothing preserves sharp transitions in signal