Chapter 6 Approximation and fitting

Last update on 2025-04-15 10:53:52+08:00

Table of contents

Norm approximation

Least-norm problems

Regularized approximation

Robust approximation

Function fitting and interpolation

Norm approximation

Least-norm problems

Regularized approximation

Robust approximation

Function fitting and interpolation

Norm approximation problem

let $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ and $\|\cdot\|$ norm on $\mathbb{R}^m$

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

$r = Ax - b$ is called the residual for the problem. WLOG, $m \geq n$

interpretations of solution x^*

- **approximation** x, an optimal solution, is called the regression
- **geometric** Ax^* 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^*
- **optimal design** x are design variables (input), Ax is result (output)
  x^* is design that best approximates desired result b

Examples

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

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

unique solution $x^* = (A^T A)^{-1} A^T b$ if $\mathbf{rank}\, A = n$

weighted norm approximation

$$\text{minimize} \qquad \| W(Ax - b) \|$$

the $W \in \mathbb{R}^{m \times m}$ is called the weighting matrix. The problem can be considered a norm approximation problem with the W-weighted norm

$$\|z\|_W = \|Wz\|$$

Chebyshev (minimax) 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

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

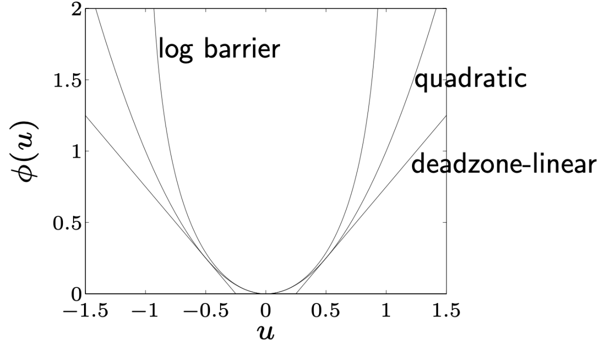
This is called a robust estimator (for reasons that will be clear later)

Penalty function approximation

let $A \in \mathbb{R}^{m \times n}$ and $\phi\colon \mathbb{R} \to \mathbb{R}$ convex penalty function

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

common penalty functions



- quadratic

$$\phi(u) = u^2$$

- deadzone-linear (with width a)

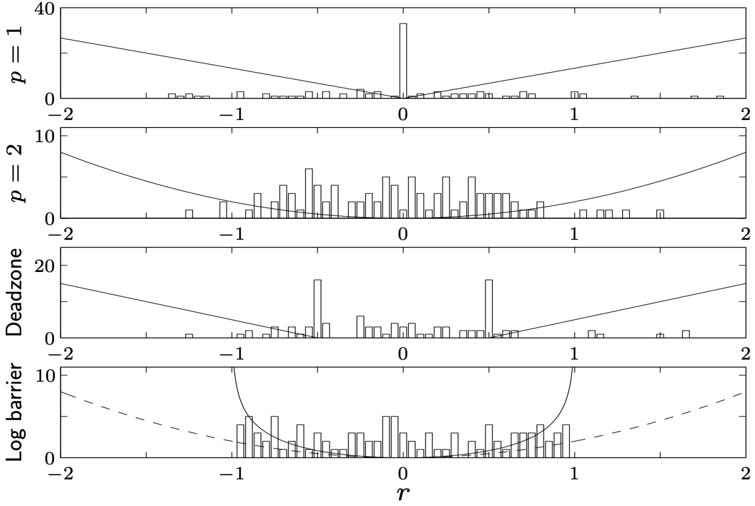
$$\phi(u) = \mathbf{max}\{0, |u| - a\}$$

- log-barrier (with limit a)

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

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

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

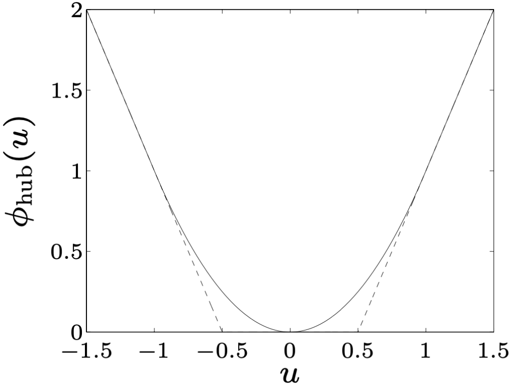


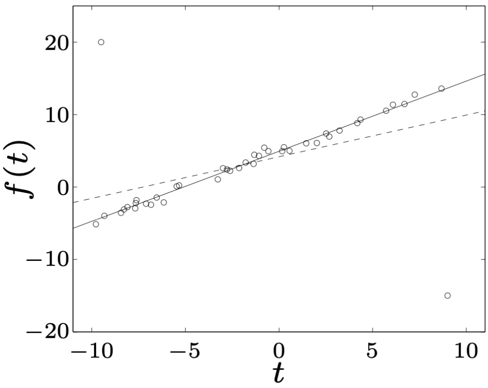
shape of penalty function has large effect on distribution of residuals

Huber penalty function (with parameter M)

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

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 (circles) using quadratic (dashed) and Huber (solid) penalty

approximation with constraints

- **nonnegative constraints on variables**

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

- **variable bounds**

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

- **probability distribution**

$$\begin{array}{ll} \text{minimize} & ||Ax - b|| \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

- **norm ball constraint**

$$\begin{array}{ll} \text{minimize} & ||Ax - b|| \\ \text{subject to} & ||x - x_0|| \leq d \end{array}$$

Norm approximation

Least-norm problems

Regularized approximation

Robust approximation

Function fitting and interpolation

Least-norm problem

let $A \in \mathbb{R}^{m \times n}$ with $m \leq n$ and $\|\cdot\|$ norm on $\mathbb{R}^n$

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

interpretation of solution x^*

- **reformulation as norm approximation problem** Let $x = x_0 + Zu$, then

$$\text{minimize} \quad \|x_0 + Zu\|$$

- **geometric** x^* is point in affine set $\{x \mid Ax = b\}$ with minimum distance to 0
- **estimation** x^* is most plausible estimate consistent with measurements $b = Ax$
- **design** x are design variables (input), b are required results (output)
  x^* is most efficient design that satisfies requirements

Examples

least-squares $\|\cdot\|_2$ can be solved via optimality conditions

$$\begin{aligned} 2x + A^T\nu &= 0 \\ Ax &= b \end{aligned}$$

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

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

extension: least penalty problem

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

Norm approximation

Least-norm problems

Regularized approximation

Robust approximation

Function fitting and interpolation

Bi-criterion formulation

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

$$\text{minimize (with respect to } \mathbb{R}^2_+) \qquad (\|Ax - b\|, \|x\|)$$

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 approximations with large x

Regularization via a scalarization method

- tracing out optimal trade-off curve

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

  for $\gamma > 0$

- Tikhonov regularization

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

  for $\delta > 0$, can be solved as a least-square problem

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

  with solution

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

- smoothing regularization

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

Optimal input design

linear dynamical system with impulse response h

$$y(t) = \sum_{\tau=0}^{t} h(\tau)u(t-\tau), \qquad t = 0, 1, \cdots, N$$

track desired output using a small and slowly varying input signal

input design problem multi-criterion problem with 3 objectives

1. tracking error with desired output y_{des}

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

2. input magnitude

$$J_{\mathrm{mag}} = \sum_{t=0}^{N} u(t)^2$$

3. input variation

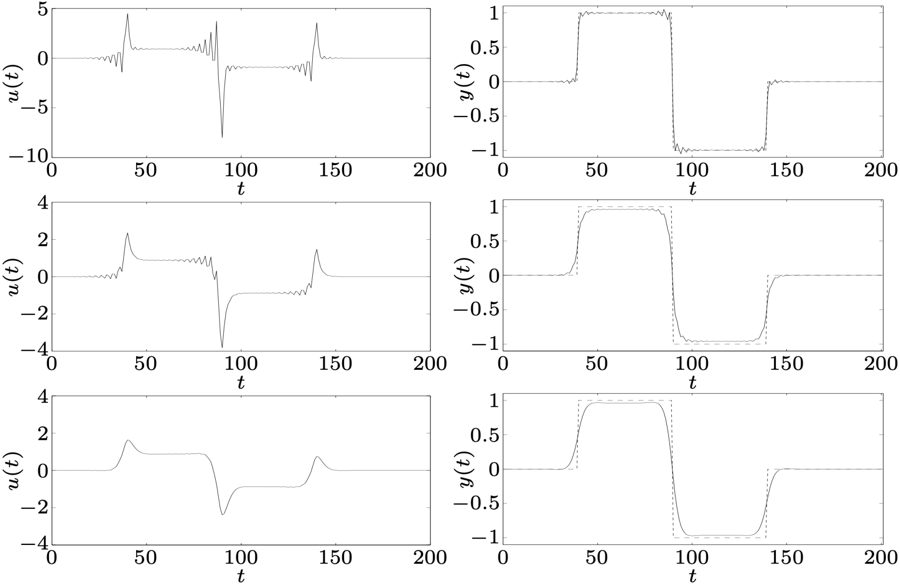
$$J_{\mathrm{der}} = \sum_{t=0}^{N-1} (u(t+1) - u(t))^2$$

regularized least-square formulation

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

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

example 3 solutions on optimal trade-off surface



top: $\delta = 0$, small $\eta = 0.005$; **middle**: $\delta = 0$, large $\eta = 0.05$; **bottom**: large $\delta = 0.3, \eta = 0.05$

ℓ_1-norm regularization

regularization with an ℓ_1-norm can be used for finding a sparse solution

$$\text{minimize} \qquad \|Ax - b\|_2 + \gamma\|x\|_1$$

by varying the parameter γ we can sweep out the optimal trade-off curve between $\|Ax - b\|_2$ and $\|x\|_1$, which serves as an approximation of the optimal trade-off curve between $\|Ax - b\|_2$ and the sparsity or cardinality of x.

the problem can be recast and solved as an SOCP

Signal reconstruction or de-noising

$$\text{minimize (with respect to } \mathbb{R}^2_+) \qquad (\|\hat{x} - x_{\text{cor}}\|_2,\ \phi(\hat{x}))$$

- $x \in \mathbb{R}^n$ 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
- $\phi \colon \mathbb{R}^n \to \mathbb{R}$ is regularization function or smoothing objective, examples include

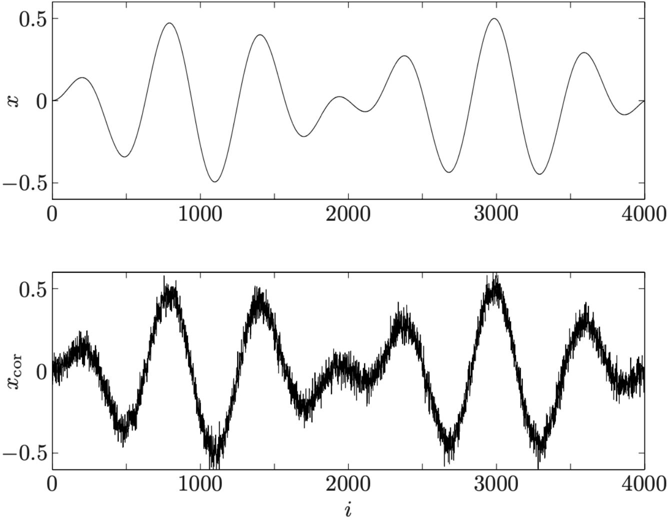
  quadratic smoothing

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

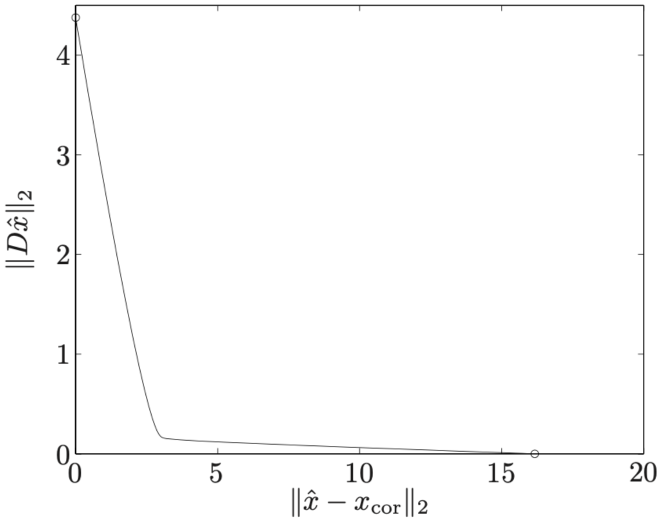
  total variation reconstruction

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

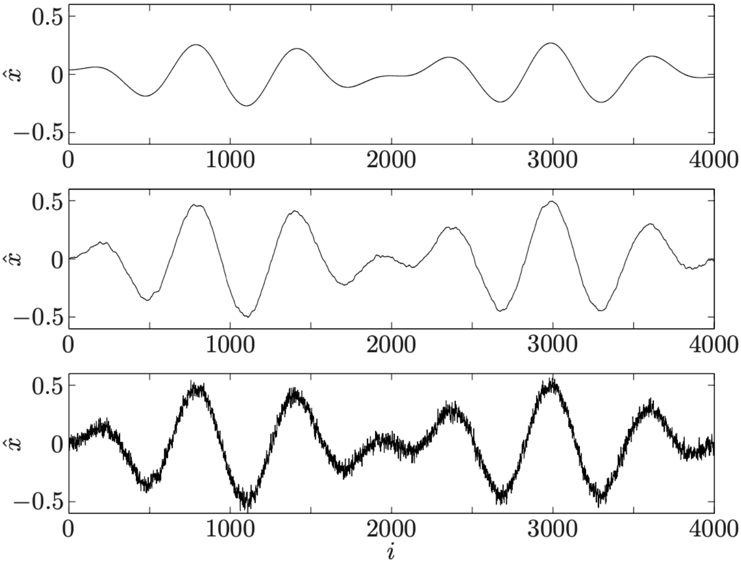
Quadratic smoothing example



original signal x and noisy signal x_{cor}

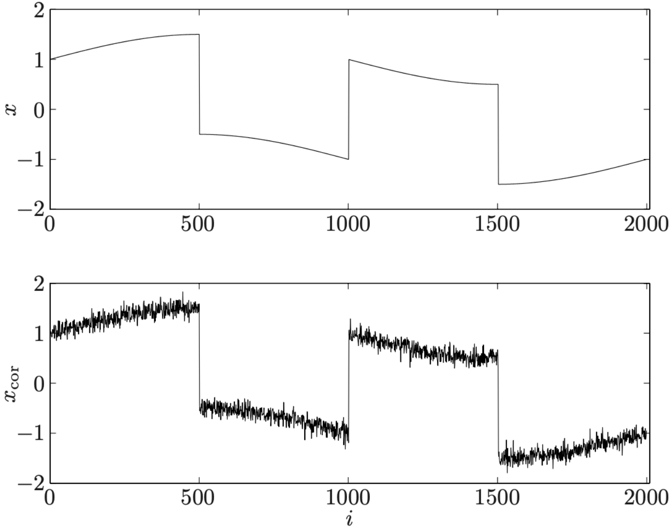


optimal trade-off curve between $(\phi_{\rm quad}(\hat{x}))^{1/2}$ and $\|\hat{x} - x_{\rm cor}\|_2$

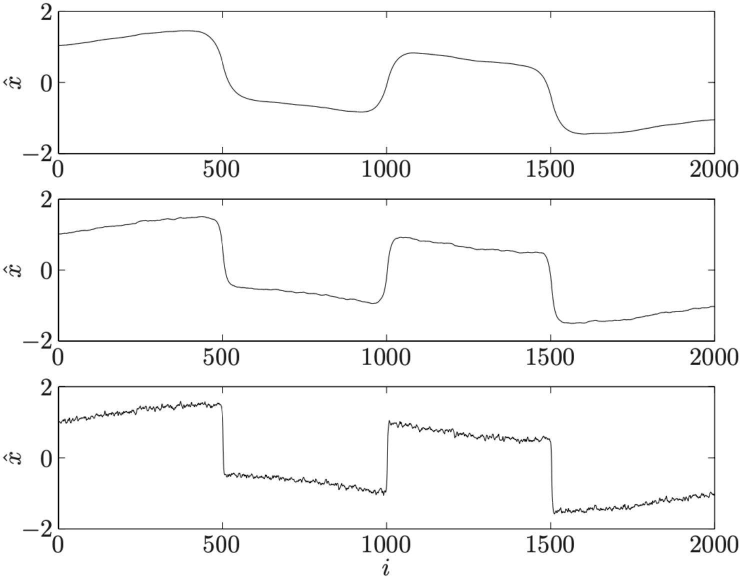


three quadratically smoothed signals $\hat{x}$ on trade-off curve

Total variation reconstruction example

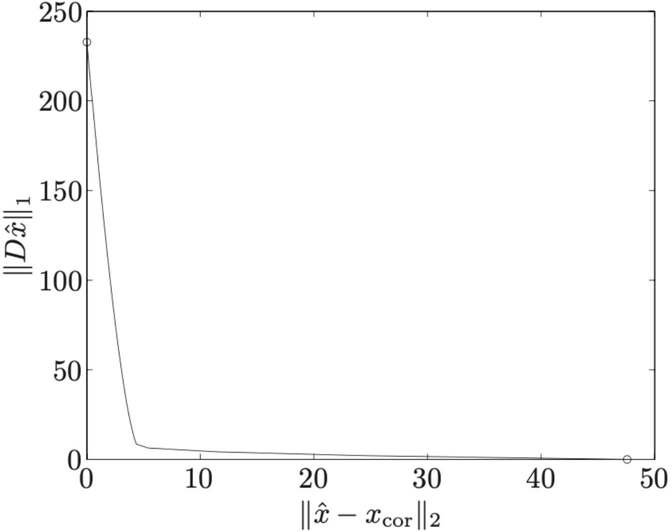


original signal x and noisy signal x_{cor}

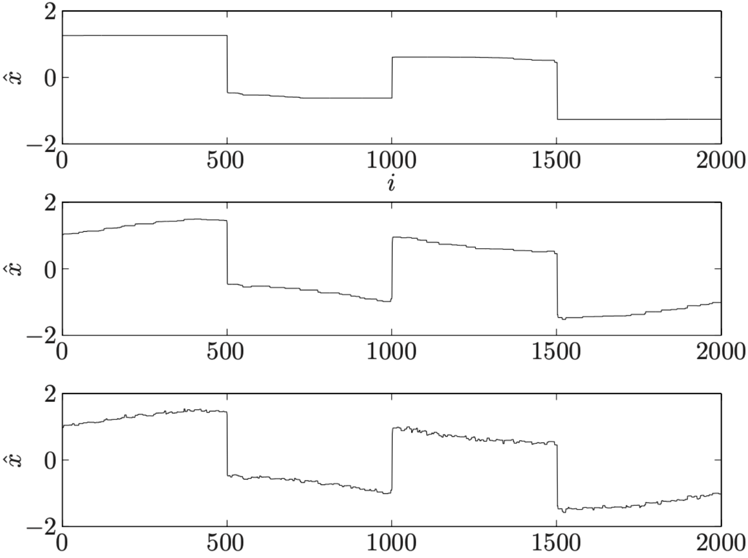


three quadratically smoothed signals $\hat{x}$

quadratic smoothing reduces noise and sharp transitions in signal



optimal trade-off curve between $\phi_{\mathrm{tv}}(\hat{x})$ and $\|\hat{x} - x_{\mathrm{cor}}\|_2$



three reconstructed signals $\hat{x}$ using total variation

total variation smoothing preserves sharp transitions in signal

Norm approximation

Least-norm problems

Regularized approximation

Robust approximation

Function fitting and interpolation

Robust approximation

$$\text{minimize} \quad \|Ax - b\| \quad \text{with uncertain } A = \overline{A} + U$$

stochastic approach assume A is random, minimize

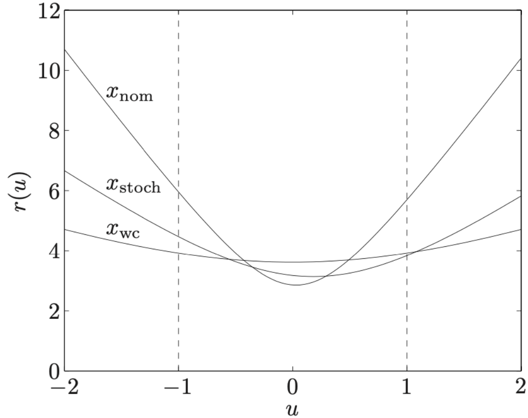
$$\mathbf{E}\|Ax - b\|$$

worst-case approach set $\mathcal{A}$ of possible values of A, minimize

$$\sup_{A \in \mathcal{A}} \|Ax - b\|$$

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

Example minimize $\|Ax - b\|$ with $A(u) = A_0 + uA_1$



residue $r(u) = \|A(u)x - b\|_2$ as a function of the uncertain parameter u

x_{nom}	minimizes	$\|A_0x - b\|_2^2$
x_{stoch}	minimizes	$\mathbf{E}\|A(u)x - b\|_2^2$ with u uniform on $[-1, 1]$
x_{wc}	minimizes	$\mathbf{sup}_{-1\leq u\leq 1}\|A(u)x - b\|_2^2$

Stochastic robust LS

assume $A = \overline{A} + U$, with U random, $\mathbf{E}U = 0$, $\mathbf{E}U^TU = P$

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

- explicit expression for objective

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

- stochastic robust LS is equivalent to standard LS

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

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

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

Worst-case robust LS

assume $\mathcal{A} = \left\{\overline{A} + u_1 A_1 + \cdots + u_p A_p \;\middle|\; \|u\|_2 \leq 1\right\}$

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

where $P(x) = \begin{bmatrix} A_1 x & A_2 x & \dots & A_p x \end{bmatrix}$ and $q(x) = \overline{A}x - b$

- we have seen strong duality between the following pair of problems

primal

$$\begin{array}{ll} \text{maximize} & \|Pu + q\|_2^2 \\ \text{subject to} & \|u\|_2^2 \leq 1 \end{array}$$

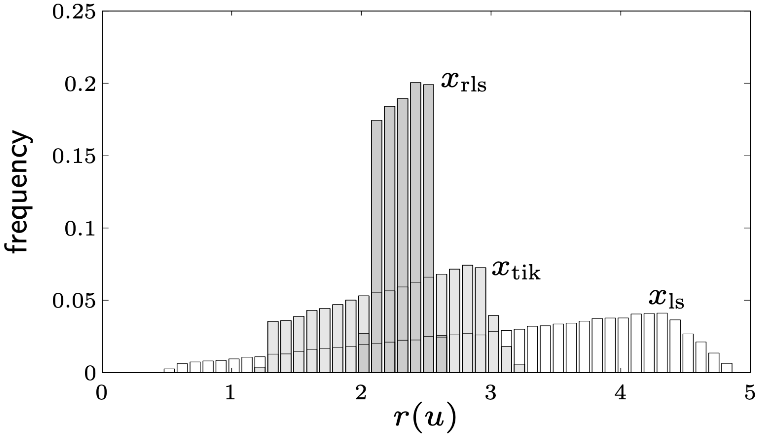
dual

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

- worst-case robust LS is equivalent to SDP

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

Example let $A = A_0 + u_1 A_1 + u_2 A_2$ with u uniformly distributed on unit disk



histogram of residues $r(u) = \|A(u)x - b\|_2$

$x_{\sf ls}$	minimizes	$\|A_0x - b\|_2^2$	(nominal)
$x_{\sf tik}$	minimizes	$\|A_0x - b\|_2^2 + \|x\|_2^2$	(Tikhonov)
$x_{\sf rls}$	minimizes	$\mathbf{sup}_{A\in\mathcal{A}}\|Ax - b\|_2^2$	(worst-case)

Norm approximation

Least-norm problems

Regularized approximation

Robust approximation

Function fitting and interpolation

function families

the function $f : \mathbb{R}^k \to \mathbb{R}$ is given by

$$f(u) = x_1 f_1(u) + \cdots + x_1 f_1(u)$$

The family $\{f_1, \cdots, f_n\}$ is called the set of basis functions and the vector $x \in \mathbb{R}^n$ is called the coefficient vector.

polynomials

- $f_i(t) = t^{i-1}$
- orthonormal polynomials wrt some positive function (or measure)
- Lagrange basis $f_1, \cdots, f_n$ associated with distinct points $t_1, \cdots, t_n$ which satisfy

$$f_i(t_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

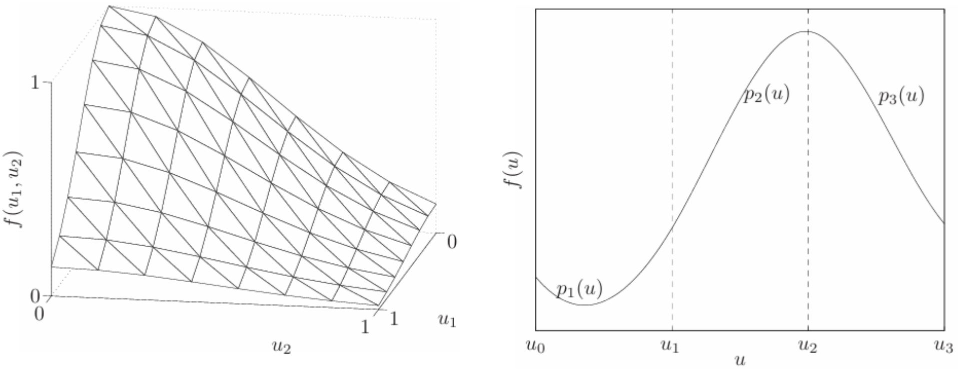
- trigonometric polynomials of degree less than n

piecewise-linear functions

- triangularization of the domain
- the basis functions f_i are affine on each simplex

piecewise polynomials and splines

- piecewise-affine functions on a triangulated domain is readily extended to piecewise polynomials and other functions
- piecewise polynomials are defined as polynomials which are continuous, i.e., the polynomials agree at the boundaries between simplexes
- restricting the piecewise polynomials to have continuous derivatives up to a certain order, we define various classes of spline functions



(left) piecewise-linear function on the unit square

(right) a cubic spline with continuous first and second derivatives

Constraints

function value interpolation and inequalities

- box constraints

$$l \leq f(v) \leq u$$

- Lipschitz constraint

$$|f(v_j) - f(v_k)| \leq L\|v_j - v_k\|, \quad j,k = 1,\ldots,m$$

- inequalities on the function values at an infinite number of points

$$f(u) \geq 0 \text{ for all } u \in D$$

derivative constraints

- the norm of the gradient at v not exceed a given limit

$$\|\nabla f(v) \leq M\|$$

- a linear matrix inequality (convex)

$$lI \preceq \nabla^2 f(v) \preceq uI$$

constraints on the derivatives at an infinite number of points

- f is monotone

$$f(u) \geq f(v) \text{ for all } u, v \in D,\ u \succeq v.$$

- the function is convex

$$f((u+v)/2) \leq (f(u) + f(v))/2 \text{ for all } u, v \in D$$

integral constraints

- moment constraint

$$\int_D t^m f(t) dt = a$$

minimum norm function fitting we are given data

$$(u_1, y_1), \ldots, (u_m, y_m),$$

and seek a function $f \in \mathcal{F}$ that matches this data as closely as possible

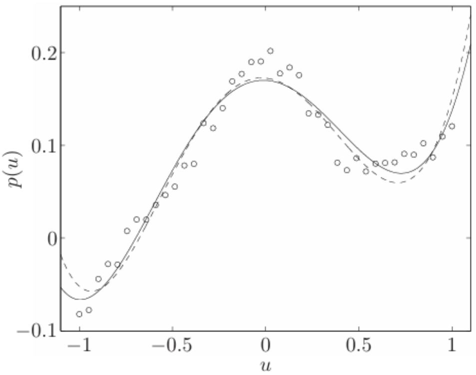
in least-squares fitting we consider the problem

$$\text{minimize} \qquad \sum_{i=1}^{m} (f(u_i) - y_i)^2$$

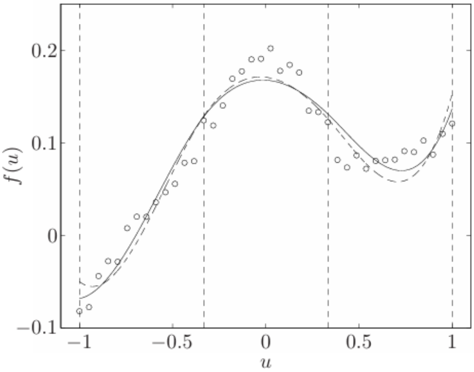
least-norm interpolation we have fewer data points than the dimension of the subspace of functions. We require that

$$f(u_i) = y_i, \qquad i = 1, \ldots, m$$

among the functions, we seek one that is smoothest, or smallest. These lead to least-norm problems



(left) polynomial fitting



(right) spline fitting

sparse descriptions and basis pursuit

in basis pursuit, there is a very large number of basis functions, and the goal is to find a good fit of the given data as a linear combination of a small number of the basis functions. sparse descriptions and basis pursuit can be used for de-noising or smoothing

we seek a function $f \in \mathcal{F}$ that fits the data well

$$f(u_i) \approx y_i, \qquad i = 1, \ldots, m,$$

with a sparse coefficient vector x, i.e., $\mathbf{card}(x)$ small.

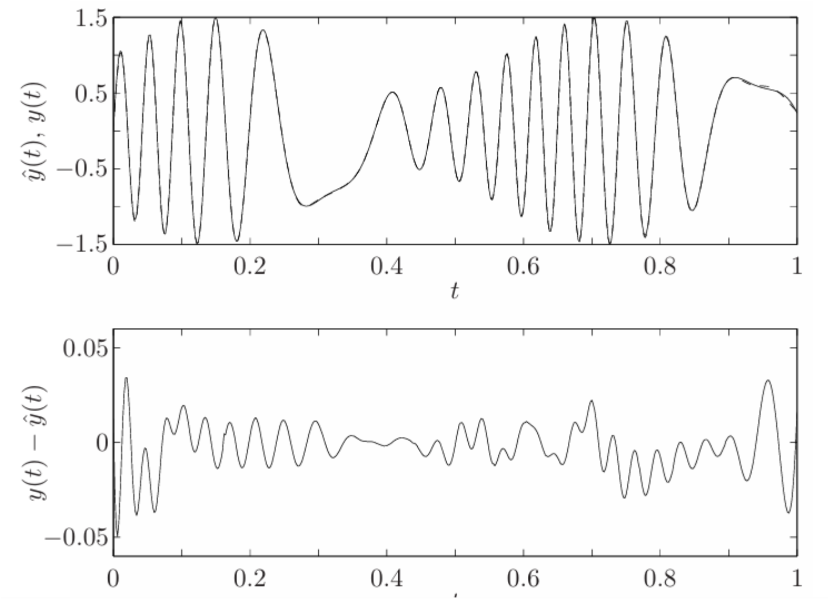
regressor selection based on ℓ_1-norm regularization

$$\text{minimize} \qquad \sum_{i=1}^{m} (f(u_i) - y_i)^2 + \gamma \|x\|_1$$

Then we solve the least-squares problem

$$\text{minimize} \qquad \sum_{i=1}^{m} (f(u_i) - y_i)^2$$

with variables $x_i, i \in \mathcal{B}$, and $x_i = 0, i \notin \mathcal{B}$



(top) original signal (solid) and approximation by basis pursuit (dashed)

(bottom) the approximation error

interpolation with convex functions

Question: When does there exist a convex function $f : \mathbb{R}^k \to \mathbb{R}$, with $\mathbf{dom}\, f = \mathbb{R}^k$, that satisfies the interpolation conditions

$$f(u_i) = y_i, \qquad i = 1, \ldots, m,$$

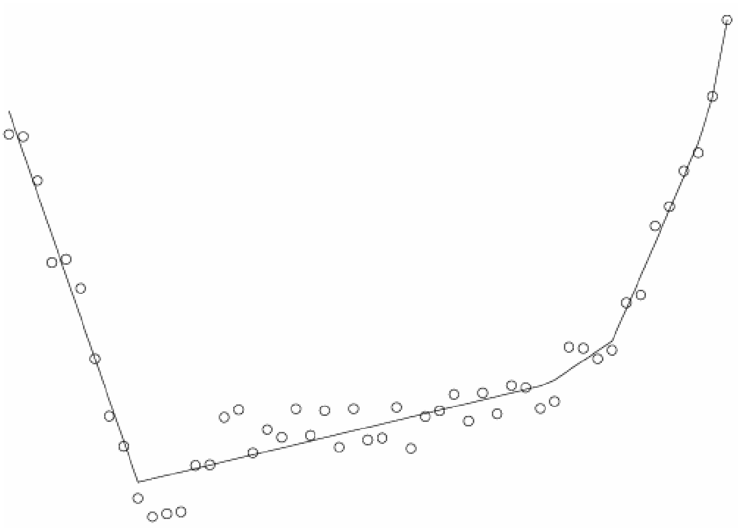
at given points $u_i \in \mathbb{R}^k$? (Here we do not restrict f to lie in any finite-dimensional subspace of functions)

Answer: if and only if there exist $g_1, \ldots, g_m$ such that

$$y_j \geq y_i + g_i^\top (u_j - u_i), \qquad i, j = 1, \ldots, m$$

if f is differentiable, we can take $g_i = \nabla f(u_i)$

in the more general case, we can construct g_i by finding a supporting hyperplane to $\mathbf{epi}\, f$ at (u_i, y_i). The vectors g_i are called **subgradients**.



least-squares fit of a convex function to data. The (piecewise-linear) function shown minimizes the sum of squared fitting error, over all convex functions