Convex Optimization

Stephen Boyd   Lieven Vandenberghe

Revised slides by Stephen Boyd, Lieven Vandenberghe, and Parth Nobel

6. Approximation and fitting

Outline

Norm and penalty approximation

Regularized approximation

Robust approximation

Norm approximation

- minimize $\|Ax - b\|$, with $A \in \mathbf{R}^{m\times n}$, $m \geq n$, $\|\cdot\|$ is any norm

- **approximation**: $Ax^\star$ is the best approximation of b by a linear combination of columns of A
- **geometric**: $Ax^\star$ is point in $\mathcal{R}(A)$ closest to b (in norm $\|\cdot\|$)
- **estimation**: linear measurement model $y = Ax + v$
  - measurement y, v is measurement error, x is to be estimated
  - implausibility of v is $\|v\|$
  - given $y = b$, most plausible 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 (in norm $\|\cdot\|$)

Examples

- Euclidean approximation ($\|\cdot\|_2$)
  - solution $x^\star = A^\dagger b$
- Chebyshev or minimax approximation ($\|\cdot\|_\infty$)
  - can be solved via 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 via 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)

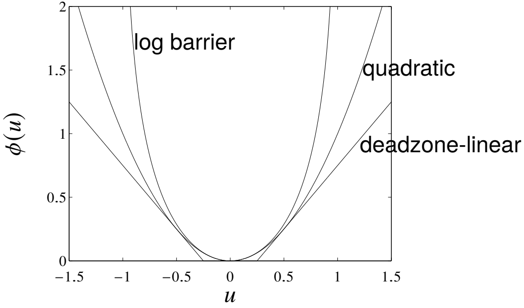
examples

- 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: histograms of residuals

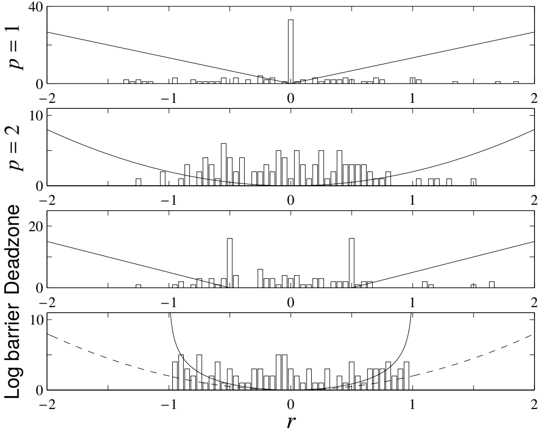
$A \in \mathbf{R}^{100\times 30}$; shape of penalty function affects distribution of residuals

absolute value $\phi(u) = |u|$

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

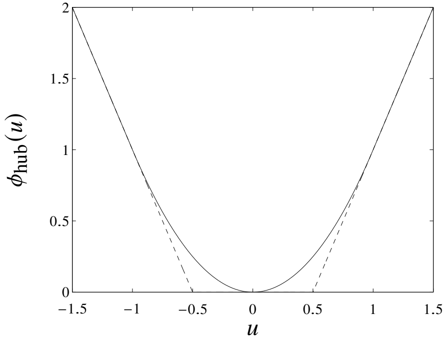
deadzone $\phi(u) = \max\{0, |u| - 0.5\}$

log-barrier $\phi(u) = -\log(1 - u^2)$



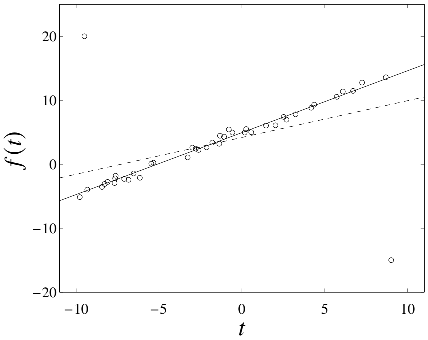
Huber penalty function

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



- linear growth for large u makes approximation less sensitive to outliers
- called a **robust penalty**

Example



- 42 points (circles) t_i, y_i, with two outliers
- affine function $f(t) = \alpha + \beta t$ fit using quadratic (dashed) and Huber (solid) penalty

Least-norm problems

- least-norm problem:

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

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

- **geometric:** $x^\star$ is smallest point in solution set $\{x \mid Ax = b\}$
- **estimation:**
  - $b = Ax$ are (perfect) measurements of x
  - $\|x\|$ is implausibility of x
  - $x^\star$ is 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 Euclidean norm ($\|\cdot\|_2$)
  - solution $x = A^\dagger b$ (assuming $b \in \mathcal{R}(A)$)
- least sum of absolute values ($\|\cdot\|_1$)
  - can be solved via 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 yield sparse $x^\star$

Outline

Norm and penalty approximation

Regularized approximation

Robust approximation

Regularized approximation

- a bi-objective problem:

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

- $A \in \mathbf{R}^{m \times n}$, 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

- minimize $\|Ax - b\| + \gamma\|x\|$
- solution for $\gamma > 0$ traces out optimal trade-off curve
- other common method: minimize $\|Ax - b\|^2 + \delta\|x\|^2$ with $\delta > 0$
- with $\|\cdot\|_2$, called **Tikhonov regularization** or **ridge regression**

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

- can be solved as a least-squares problem

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

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

Optimal input design

- **linear dynamical system** (or **convolution system**) with impulse response h:

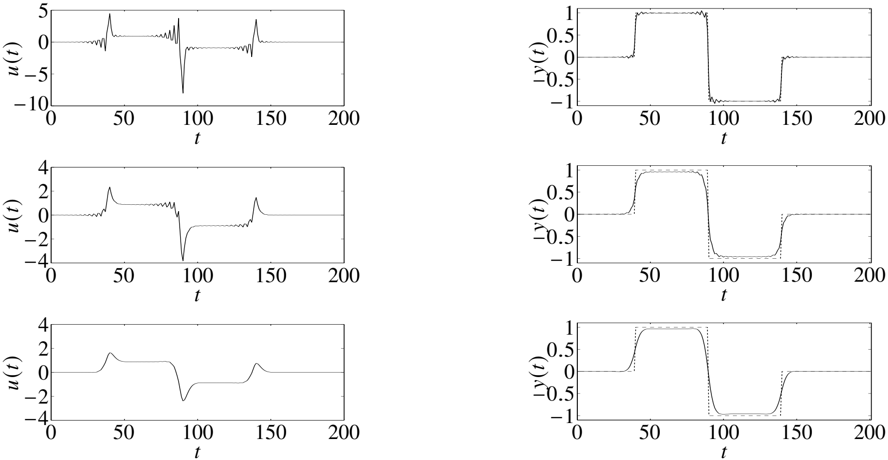
$$y(t) = \sum_{\tau=0}^{t} h(\tau)u(t-\tau), \quad t = 0, 1, \ldots, N$$

- **input design problem:** multicriterion problem with 3 objectives
  - tracking error with desired output y_{des}: $J_{\text{track}} = \sum_{t=0}^{N}(y(t) - y_{\text{des}}(t))^2$
  - input magnitude: $J_{\text{mag}} = \sum_{t=0}^{N} u(t)^2$
  - 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**: minimize $J_{\text{track}} + \delta J_{\text{der}} + \eta J_{\text{mag}}$
  - for fixed δ, η, a least-squares problem in $u(0), \ldots, u(N)$

Example

- minimize $J_{\text{track}} + \delta J_{\text{der}} + \eta J_{\text{mag}}$
- (top) $\delta = 0$, small η; (middle) $\delta = 0$, larger η; (bottom) large δ



Signal reconstruction

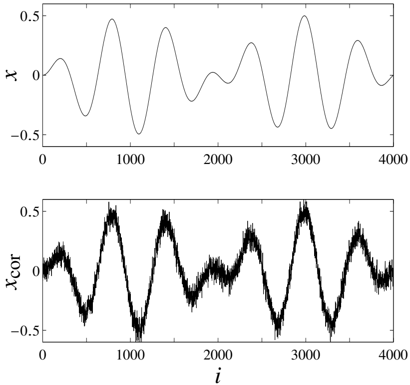
- bi-objective problem:

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

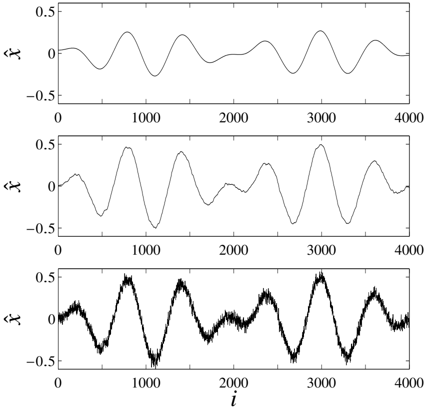
  - $x \in \mathbf{R}^n$ is unknown signal
  - $x_{\text{cor}} = x + v$ is (known) corrupted version of x, with additive noise v
  - variable $\hat{x}$ (reconstructed signal) is estimate of x
  - $\phi : \mathbf{R}^n \to \mathbf{R}$ is regularization function or smoothing objective

- **examples:**
  - quadratic smoothing, $\phi_{\text{quad}}(\hat{x}) = \sum_{i=1}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2$
  - total variation smoothing, $\phi_{\text{tv}}(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$

Quadratic smoothing example

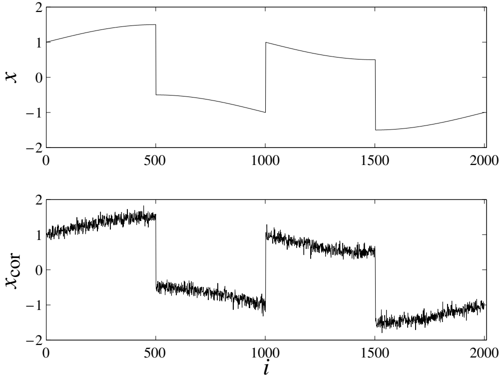


original signal x and noisy signal x_{cor}

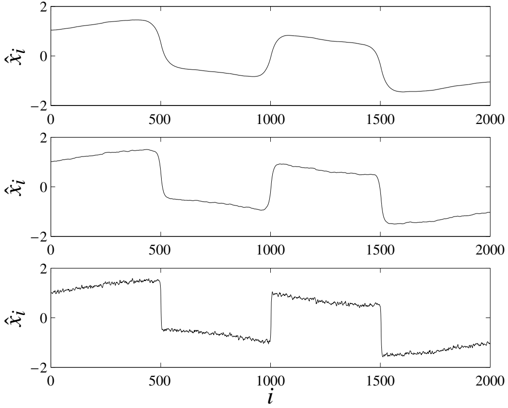


three solutions on trade-off curve
$\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

Reconstructing a signal with sharp transitions



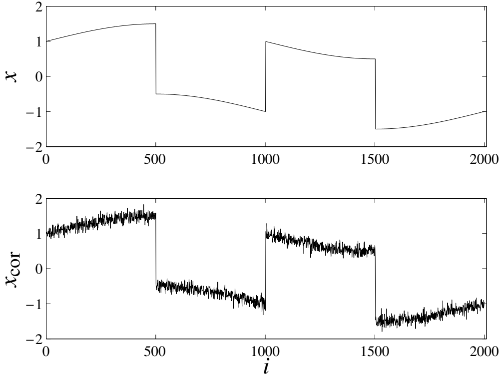
original signal x and noisy signal x_{cor}



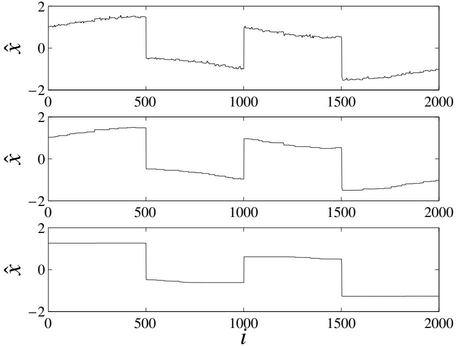
three solutions on trade-off curve
$\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

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

Total variation reconstruction



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

Outline

Norm and penalty approximation

Regularized approximation

Robust approximation

Robust approximation

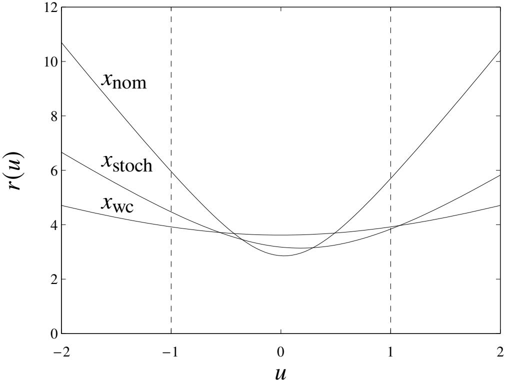
- minimize $\|Ax - b\|$ with uncertain A
- two approaches:
  - **stochastic**: assume A is random, minimize $\mathbf{E}\,\|Ax - b\|$
  - **worst-case:** set $\mathcal{A}$ of possible values of A, minimize $\sup_{A \in \mathcal{A}} \|Ax - b\|$
- tractable only in special cases (certain norms $\|\cdot\|$, distributions, sets $\mathcal{A}$)

Example

$A(u) = A_0 + uA_1$, $u \in [-1, 1]$

- $x_{\rm nom}$ 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 $\sup_{-1\leq u\leq 1} \|A(u)x - b\|_2^2$

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



Stochastic robust least-squares

- $A = \bar{A} + U$, U random, $\mathbf{E}\, U = 0$, $\mathbf{E}\, U^T U = P$
- stochastic least-squares problem: minimize $\mathbf{E}\, \|(\bar{A} + U)x - b\|_2^2$
- 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 least-squares problem is equivalent to: minimize $\|\bar{A}x - b\|_2^2 + \|P^{1/2}x\|_2^2$
- for $P = \delta I$, get Tikhonov regularized problem: minimize $\|\bar{A}x - b\|_2^2 + \delta\|x\|_2^2$

Worst-case robust least-squares

- $\mathcal{A} = \{\bar{A} + u_1 A_1 + \cdots + u_p A_p \mid \|u\|_2 \leq 1\}$ (an ellipsoid in $\mathbf{R}^{m\times n}$)
- worst-case robust least-squares problem is

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

  where $P(x) = \begin{bmatrix} A_1 x & A_2 x & \cdots & A_p x \end{bmatrix}$, $q(x) = \bar{A}x - b$
- from book appendix B, strong duality holds between the following problems

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

- hence, robust least-squares problem is equivalent to SDP

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

Example

- $r(u) = \|(A_0 + u_1 A_1 + u_2 A_2)x - b\|_2$, u uniform on unit disk
- three choices of x:
  - $x_{\rm ls}$ minimizes $\|A_0 x - b\|_2$
  - $x_{\rm tik}$ minimizes $\|A_0 x - b\|_2^2 + \delta\|x\|_2^2$ (Tikhonov solution)
  - $x_{\rm rls}$ minimizes $\sup_{A \in \mathcal{A}} \|Ax - b\|_2^2 + \|x\|_2^2$