06. Approximation and fitting

By Yang Lin[1] (2023 秋季, @NJU)

Contents

1 Norm approximation
2 Least-norm problems
3 Regularized approximation
4 Robust approximation

[1]Institute: Nanjing University. Email: linyang@nju.edu.cn.

1 Norm approximation

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

($A \in \mathbf{R}^{m \times n}$ with $m \geq n$)

Interpretations of solution $x = \arg\min_x \|Ax - b\|$:

1. **geometric**: Ax^* is point in $\mathcal{R}(A)$ closest to b

2. **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^*

3. **Optimal design:** x are design variables(input), Ax is result(output)

x^* is design that best approximates desired result b

1 Norm approximation

Examples.

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

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

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

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

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

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

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

1 Norm approximation

Penalty function approximation.

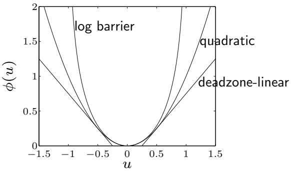
$$\begin{array}{ll} \texttt{minimize} & \phi(r_1) + \cdots \phi(r_m) \\ \texttt{subject to} & r = Ax - b \end{array}$$

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

Examples

1. quadratic: $\phi(u) = u^2$
2. deadzone-linear with width a: $\phi(u) = \max\{0, |u| - a\}$
3. log-barrier with limit a:

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

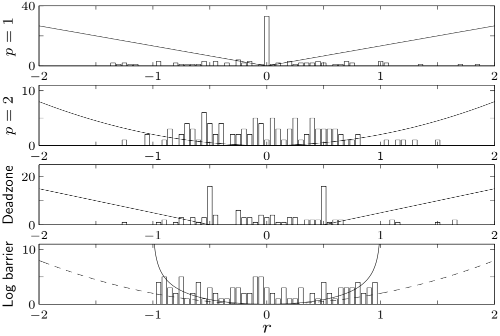


1 Norm approximation

example (m =100, n =30): histogram ofresiduals for penalties

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

Shape of penalty function has large effect on distribution of residuals

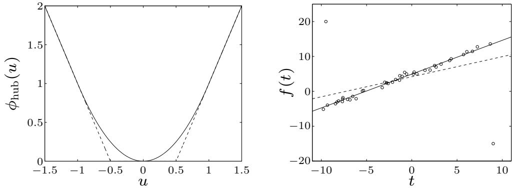


1 Norm approximation

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

2 Least-Norm problems

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

($A \in \mathbf{R}^{m \times n}$ with $m \leq n$)

Interpretations of solution $x = \arg\min_{Ax=b} \|x\|$:

1. geometric: x is point in affine set $\{x | Ax = b\}$ with minimum distance to 0

2. estimation: $b = Ax$ are (perfect) measurements of x; x is smallest (' most plausible') estimate consistent with measurements

3. x are design variables (inputs); b are required results(outputs) x^* is smallest (' most efficient') design that satisfies requirements

2 Least-Norm problems

Examples.

least-squares solution of linear equations ($\|\cdot\|_2$):
can be solved via optimality conditions

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

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

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

tends to produce sparse solution x^*

extension: least-penalty problem

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

$\phi : \mathbf{R} \to \mathbf{R}$ is convex penalty function

3 Regularized approximation

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

($A \in \mathbf{R}^{m\times n}$)

interpretation: find good approximation $Ax \approx b$ with small x

1. **estimation**: linear measurement model $y = Ax + v$, with prior knowledge that $\|x\|$ is small
2. **optimal design**: small x is cheaper or more efficient, or the linear model $y = Ax$ is only valid for small x
3. **robust approximation**: good approximation $Ax \approx b$ with small x is less sensitive to errors in A than good approximation with large x

3 Regularized approximation

Scalarized problem

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

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

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

can be solved as a least-squares problem

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

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

3 Regularized approximation

Optimal input design

linear dynamical system with impulse response h:

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

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

regularized least-squares formulation

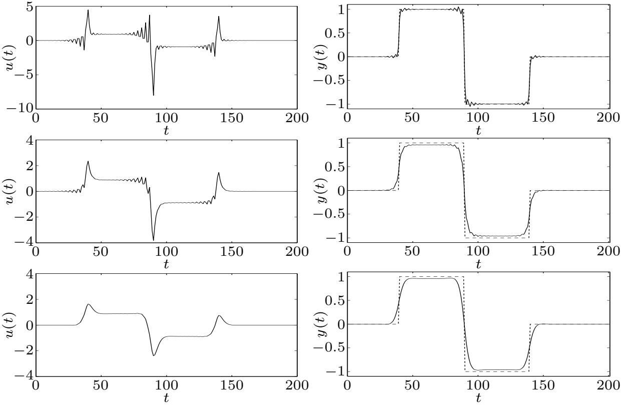
$$\texttt{minimize} \quad J_{\text{track}} + \delta J_{\text{der}} + \eta J_{\text{mag}}$$

for fixed δ, η, a least-squares problem in $\mu(0), \ldots, \mu(N)$

3 Regularized approximation

example: 3 solutions on optimal trade-off curve

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



3 Regularized approximation

Signal reconstruction

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

$x \in \mathbf{R}^n$ is unknown signal

$x_{\mathrm{cor}} = x + v$ is (known) corrupted version of x, with additive noise v

variable $\hat{x}$ (reconstructed signal) is estimate of x

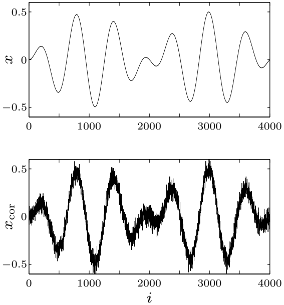
$\phi : \mathbf{R}^n \to \mathbf{R}$ is regularization function or smoothing objective

examples: quadratic smoothing, total variation smoothing:

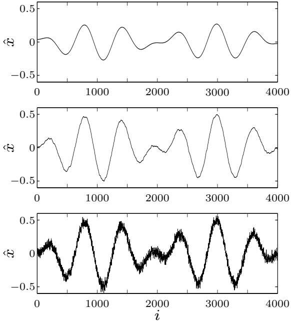
$$\phi_{\mathrm{quad}}(\hat{x}) = \sum_{i=1}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2, \quad \phi_{\mathrm{tv}}(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|$$

3 Regularized approximation

quadratic smoothing example



original signal x and noisy signal x_{cor}

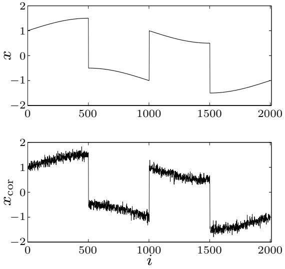


three solutions on trade-off curve $\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

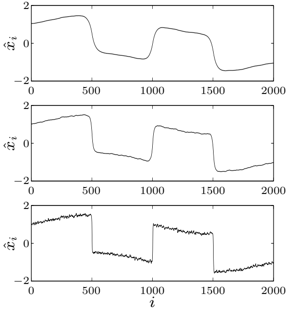
3 Regularized approximation

total variation reconstruction example

quadratic smoothing smooths out noise and sharp transitions in signal



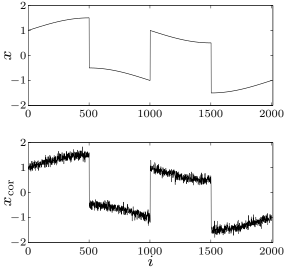
original signal x and noisy signal x_{cor}



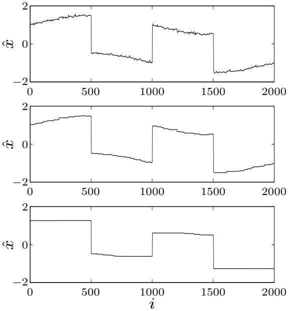
three solutions on trade-off curve $\|\hat{x} - x_{\mathrm{cor}}\|_2$ versus $\phi_{\mathrm{quad}}(\hat{x})$

3 Regularized approximation

total variation smoothing preserves sharp transitions in signal



original signal x and noisy signal x_{cor}



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

4 Robust approximation

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

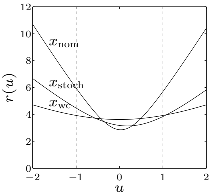
two approaches:

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

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

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

1. x_{norm} minimizes $\|A_0x-b\|_2^2$
2. x_{stoch} minimizes $\mathbf{E}\|A(u)x-b\|_2^2$ with u uniform on $[-1,1]$
3. x_{wc} minimizes $\sup_{-1\le u\le 1}\|A(u)x-b\|_2^2$ figure shows $r(u) = \|A(u)x-b\|_2$



4 Robust approximation

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

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

1. explicit expression for objective:

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

2.hence, robust LS problem is equivalent to LS problem

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

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

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

4 Robust approximation

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

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

where $P(x) = [A_1 x \ A_2 x \ \cdots \ A_p x]$, $q(x) = \bar{A}x - b$

1. strong duality holds between the following problems

$$\begin{array}{ll} \texttt{minimize} & \|Pu + q\|_2^2 \\ \texttt{subject to} & \|u\|_2^2 \le 1 \end{array}$$

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

2. hence, robust LS problem is equivalent to SDP

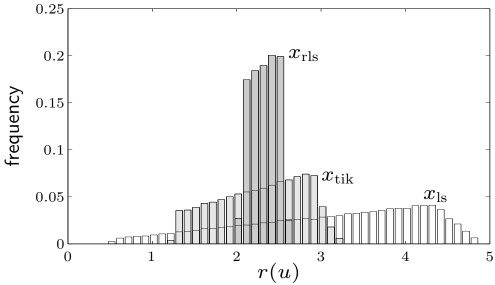
$$\begin{array}{ll} \texttt{minimize} & t + \lambda \\ \texttt{subject to} & \begin{bmatrix} I & P(x) & q(x) \\ P(x)^T & \lambda I & 0 \\ q(x)^T & 0 & t \end{bmatrix} \succeq 0 \end{array}$$

4 Robust approximation

example

$$r(u) = \|(A_0 + u_1A_1 + u_2A_2)x - b\|2$$

with u uniformly distributed on unit disk, for three values of x



1. x_{ls} minimizes $\|A_0x - b\|_2$
2. x_{tik} minimizes $\|A_0x - b\|_2^2 + \delta\|x\|_2^2$ (Tikhonov solution)
3. x_{wc} minimizes $\sup_{\|u\|_2 \le 1} \|A_0x - b\|_2^2 + \|x\|_2^2$