6. Approximation and fitting

- norm approximation
- least-norm problems
- regularized approximation
- robust approximation
Norm approximation

minimize $\|Ax - b\|$

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

interpretations of solution $x^* = \arg\min_x \|Ax - b\|$: 

- **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
  \[ A^T Ax = A^T b \]
  \(x^* = (A^T A)^{-1} A^T b\) if \(\text{rank } A = n\)

- Chebyshev approximation (\(\| \cdot \|_\infty\)): can be solved as an LP
  \[
  \begin{align*}
  \text{minimize} & \quad t \\
  \text{subject to} & \quad -t \mathbf{1} \leq Ax - b \leq t \mathbf{1}
  \end{align*}
  \]

- sum of absolute residuals approximation (\(\| \cdot \|_1\)): can be solved as an LP
  \[
  \begin{align*}
  \text{minimize} & \quad \mathbf{1}^T y \\
  \text{subject to} & \quad -y \leq Ax - b \leq y
  \end{align*}
  \]
Penalty function approximation

\[ \begin{align*} 
\text{minimize} & \quad \phi(r_1) + \cdots + \phi(r_m) \\
\text{subject to} & \quad r = Ax - b
\end{align*} \]

\( (A \in \mathbb{R}^{m \times n}, \phi : \mathbb{R} \to \mathbb{R} \) 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

\[
\begin{align*}
\text{minimize} & \quad \|x\| \\
\text{subject to} & \quad Ax = b
\end{align*}
\]

\((A \in \mathbb{R}^{m \times n} \text{ with } m \leq n, \| \cdot \| \text{ is a norm on } \mathbb{R}^n)\)

interpretations of solution \(x^* = \arg\min_{Ax=b} \|x\|:\)

- **geometric**: \(x^*\) is point in affine set \(\{x \mid Ax = b\}\) with minimum distance to 0

- **estimation**: \(b = Ax\) are (perfect) measurements of \(x\); \(x^*\) is smallest ('most plausible') estimate consistent with measurements

- **design**: \(x\) are design variables (inputs); \(b\) are required results (outputs) \(x^*\) is smallest ('most efficient') design that satisfies requirements
examples

- least-squares solution of linear equations (\(\| \cdot \|_2\)):
  can be solved via optimality conditions
  \[
  2x + A^T \nu = 0, \quad Ax = b
  \]

- minimum sum of absolute values (\(\| \cdot \|_1\)): can be solved as an LP
  minimize \(1^T y\)
  subject to \(-y \leq x \leq y, \quad Ax = b\)
  tends to produce sparse solution \(x^*\)

extension: least-penalty problem

minimize \(\phi(x_1) + \cdots + \phi(x_n)\)
subject to \(Ax = b\)

\(\phi : \mathbb{R} \to \mathbb{R}\) is convex penalty function
Regularized approximation

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

A \in \mathbb{R}^{m \times n}, \text{ norms on } \mathbb{R}^{m} \text{ and } \mathbb{R}^{n} \text{ 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: \( \text{minimize} \quad \|Ax - b\|^2 + \delta \|x\|^2 \) 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 \( x^* = (A^T A + \delta I)^{-1} A^T b \)
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 curve

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

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

- \( x \in \mathbb{R}^n \) 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 \( x \)
- \( \phi : \mathbb{R}^n \to \mathbb{R} \) is regularization function or smoothing objective

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

\[
\phi_{\text{quad}}(\hat{x}) = \sum_{i=1}^{n-1} (\hat{x}_{i+1} - \hat{x}_i)^2, \quad \phi_{\text{tv}}(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i|
\]
quadratic smoothing example

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

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

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

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

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
Robust approximation

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

two approaches:

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

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

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

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

figure shows $r(u) = \|A(u)x - b\|_2$
stochastic robust LS with $A = \bar{A} + U$, $U$ random, $\mathbf{E}U = 0$, $\mathbf{E}U^TU = P$

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

• explicit expression for objective:

\[
\mathbf{E} \|Ax - b\|^2_2 = \mathbf{E} \|\bar{A}x - b + Ux\|^2_2 \\
= \|\bar{A}x - b\|^2_2 + \mathbf{E} x^T U^T U x \\
= \|\bar{A}x - b\|^2_2 + x^T P x
\]

• hence, robust LS problem is equivalent to LS problem

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 LS** with \( A = \{ \tilde{A} + u_1A_1 + \cdots + u_pA_p \mid \|u\|_2 \leq 1 \} \)

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

where \( P(x) = \begin{bmatrix} A_1x & A_2x & \cdots & A_px \end{bmatrix} \), \( q(x) = \tilde{A}x - b \)

- from page 5–14, strong duality holds between the following problems

maximize \( \|Pu + q\|_2^2 \)  
subject to \( \|u\|_2^2 \leq 1 \)

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

- hence, robust LS problem is equivalent to SDP

minimize \( t + \lambda \)  
subject to \( \begin{bmatrix} I & P(x) & q(x) \\ P(x)^T & \lambda I & 0 \\ q(x)^T & 0 & t \end{bmatrix} \succeq 0 \)
**example:** histogram of residuals

\[ r(u) = \| (A_0 + u_1 A_1 + u_2 A_2) x - b \|_2 \]

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

- \( x_{ls} \) minimizes \( \| A_0 x - b \|_2 \)
- \( x_{tik} \) minimizes \( \| A_0 x - b \|_2^2 + \| x \|_2^2 \) (Tikhonov solution)
- \( x_{wc} \) minimizes \( \sup_{\| u \|_2 \leq 1} \| A_0 x - b \|_2^2 + \| x \|_2^2 \)