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 $x^* = \operatorname{argmin}_x \|Ax - b\|$

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

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

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

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

\((A \in \mathbb{R}^{m \times n}, \phi : \mathbb{R} \rightarrow \mathbb{R} \text{ is a convex penalty function})\)

**Examples**

- quadratic: \(\phi(u) = u^2\)
- deadzone-linear with width \(a\):
  \[
  \phi(u) = \max \{0, |u| - a\}
  \]
- log-barrier with limit \(a\):
  \[
  \phi(u) = \begin{cases} 
  -a^2 \log(1 - (u/a)^2) & |u| < a \\
  \infty & \text{otherwise}
  \end{cases}
  \]
Comparison

Example \((m = 100, n = 30)\): histogram of residuals for penalties

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

The shape of penalty function has a large effect on the distribution of residuals.

Approximation and fitting 6.5
**Huber penalty**

**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

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

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

**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

  \[ 2x + A^T v = 0, \quad Ax = b \]

• 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 $Ax = b$

$\phi : \mathbb{R} \rightarrow \mathbb{R}$ is convex penalty function
Regularized approximation

minimize (w.r.t. $\mathbb{R}^2_+$) $\left( \|Ax - b\|, \|x\| \right)$

$A \in \mathbb{R}^{m \times n}$, norms on $\mathbb{R}^m$ and $\mathbb{R}^n$ 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 \( \delta > 0 \)

**Tikhonov regularization**

minimize \[ ||Ax - b||_2^2 + \delta ||x||_2^2 \]

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 surface

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

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

- \( x \in \mathbb{R}^n \) is unknown signal
- \( x_{cor} = x + v \) is (known) corrupted version of \( x \), with additive noise \( v \)
- variable \( \hat{x} \) (reconstructed signal) is estimate of \( x \)
- \( \phi : \mathbb{R}^n \rightarrow \mathbb{R} \) 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

$\|\hat{x} - x_{\text{cor}}\|_2 \text{ versus } \phi_{\text{quad}}(\hat{x})$
Total variation reconstruction example

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

quadratic smoothing smooths out noise and sharp transitions in signal

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

Approximation and fitting 6.16
Robust approximation

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

two approaches:

- **stochastic**: assume $A$ is random, minimize $\mathbb{E} \|Ax - b\|$
- **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**: $A(u) = A_0 + uA_1$

- $x_{\text{nom}}$ minimizes $\|A_0x - b\|_2^2$
- $x_{\text{stoch}}$ minimizes $\mathbb{E} \|A(u)x - b\|_2^2$
  with $u$ uniform on $[-1, 1]$
- $x_{\text{wc}}$ 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 \( A = \tilde{A} + U \), \( U \) random, \( \mathbf{E} U = 0 \), \( \mathbf{E} U^T U = P \)

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

- explicit expression for objective:

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

- hence, robust LS problem is equivalent to LS problem

\[
\text{minimize} \quad \| \tilde{A}x - b \|_2^2 + P^{1/2}x \|_2^2
\]

- for \( P = \delta I \), get Tikhonov regularized problem

\[
\text{minimize} \quad \| \tilde{A}x - b \|_2^2 + \delta \| x \|_2^2
\]
Worst-case robust LS

with \( \mathcal{A} = \{ \tilde{A} + u_1 A_1 + \cdots + u_p A_p \mid \|u\|_2 \leq 1 \} \)

\[
\begin{align*}
\text{minimize} & \quad \sup_{A \in \mathcal{A}} \|Ax - b\|_2^2 = \sup_{\|u\|_2 \leq 1} \|P(x)u + q(x)\|_2^2
\end{align*}
\]

where \( P(x) = \begin{bmatrix} A_1 x & A_2 x & \cdots & A_p x \end{bmatrix} \), \( q(x) = \tilde{A}x - b \)

- from page 5.16, strong duality holds between the following problems

\[
\begin{align*}
\text{maximize} & \quad \|Pu + q\|_2^2 \\
\text{subject to} & \quad \|u\|_2^2 \leq 1
\end{align*}
\]

\[
\begin{align*}
\text{minimize} & \quad t + \lambda \\
\text{subject to} & \quad \begin{bmatrix} I & P & q \\ P^T & \lambda I & 0 \\ q^T & 0 & t \end{bmatrix} \succeq 0
\end{align*}
\]

- hence, robust LS problem is equivalent to SDP

\[
\begin{align*}
\text{minimize} & \quad t + \lambda \\
\text{subject to} & \quad \begin{bmatrix} I & P(x) & q(x) \\ P(x)^T & \lambda I & 0 \\ q(x)^T & 0 & t \end{bmatrix} \succeq 0
\end{align*}
\]
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 \( \|A_0x - b\|_2 \)
- \( x_{tik} \) minimizes \( \|A_0x - b\|_2^2 + \delta \|x\|_2^2 \) (Tikhonov solution)
- \( x_{rls} \) minimizes \( \sup_{A \in \mathcal{A}} \|Ax - b\|_2^2 + \|x\|_2^2 \)