Convex Optimization
selections from Chapter 6

Audrey Hesse
Approximation & Fitting

- Norm Approximation
  - Penalty functions
- Least-norm problems
- Regularized Approximation
- Signal Reconstruction
Norm Approximation

minimize \|Ax - b\|

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

- 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 \(b\)
• A solution to norm approximation is sometimes an approximate solution
  \[ Ax \approx b \]
• Residual: vector \( r = Ax - b \)
• Residuals: the individual components of the residual associated with \( x \)
Minimize:

\[ A^T A x = A^T b \]

Minimize the sum of the squares of the residuals

\[ \| A x - b \|_2^2 = r_1^2 + r_2^2 + \ldots + r_m^2 \]

- **least-squares approximation** (\( \| \cdot \|_2 \)): solution satisfies normal equations

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

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad -t \mathbf{1} \leq A x - b \leq t \mathbf{1}
\end{align*}
\]

Minimize:

\[ \| A x - b \|_\infty = \max \{ | r_1 |, \ldots, | r_m | \} \]

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

\[
\begin{align*}
\text{minimize} & \quad \mathbf{1}^T y \\
\text{subject to} & \quad -y \leq A x - b \leq y
\end{align*}
\]

Minimize:

\[ \| A x - b \|_1 = | r_1 | + | r_2 | + \ldots + | r_m | \]
Penalty function approximation

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

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

examples

- quadratic: \( \phi(u) = u^2 \)
- deadzone-linear with width \( a \):
  \[
  \phi(u) = \max\{0, |u| - a\}
  \]
- log-barrier with limit \( a \):
  \[
  \phi(u) = \begin{cases} 
  -a^2 \log(1 - (u/a)^2) & |u| < a \\
  \infty & \text{otherwise}
  \end{cases}
  \]
example \( (m = 100, n = 30) \): histogram of residuals for penalties

\[
\phi(u) = |u|, \quad \phi(u) = u^2, \quad \phi(u) = \max\{0, |u| - a\}, \quad \phi(u) = -\log(1-u^2)
\]
Huber penalty function (with parameter $M$)

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

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

minimize \|x\|
subject to \ Ax = b

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

interpretations of solution \( x^* = \text{argmin}_{Ax=b} \|x\| \):

- **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

$$\text{minimize} \quad 1^T y$$
$$\text{subject to} \quad -y \leq x \leq y, \quad Ax = b$$

tends to produce sparse solution $x^*$

extension: least-penalty problem

$$\text{minimize} \quad \phi(x_1) + \cdots + \phi(x_n)$$
$$\text{subject to} \quad Ax = b$$

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

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

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

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

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

minimize \( J_{\text{track}} + \delta J_{\text{der}} + \eta J_{\text{mag}} \)

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

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

- $x \in \mathbb{R}^n$ 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:

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

\[ x \in \mathbb{R}^{4000} \]

original signal \( x \) and noisy signal \( x_{\text{cor}} \)

three solutions on trade-off curve

\[ \| \hat{x} - x_{\text{cor}} \|_2 \text{ versus } \phi_{\text{quad}}(\hat{x}) \]
Optimal trade-off curve

\[ \| D\hat{x} \|_2 \]

Knee at \[ \| \hat{x} - x_{\text{cor}} \| \approx 3. \]
total variation reconstruction example

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

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

quadratic smoothing smooths out noise and sharp transitions in signal
original signal $x$ and noisy signal $x_{cor}$

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

total variation smoothing preserves sharp transitions in signal