Lec 10: Application

approximation and

fitting.

- Norm approximation

$$\min \|Ax-b\| \qquad (\|\cdot\| \text{ is any norm})$$

geometric: Ax^* is point in range(A) that's closest to b
estimation: linear measurement model $y = Ax + v$
optimal design: x is design variable, b is desired result

eg. least-square ($\|\cdot\|_2$):

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

eg. chebyshev approximation ($\|\cdot\|_\infty$)

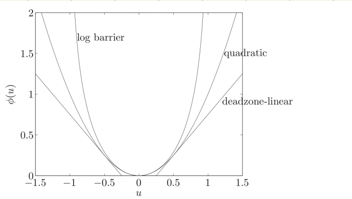
$$\min.\ \|Ax-b\|_\infty$$
$$\Downarrow$$
$$\min.\ t$$
$$\text{s.t. } -t\mathbf{1} \le Ax-b \le t\mathbf{1}$$

eg. sum of absolute residual approximation

$$\min.\ \|Ax-b\|_1$$
$$\Downarrow$$
$$\min.\ \mathbf{1}^T t$$
$$\text{s.t. } -t \le Ax-b \le t$$

- Penalty function approximation

$$\min.\ \phi(r_1) + \cdots + \phi(r_m)$$
$$\text{s.t. } r = Ax-b$$



eg.

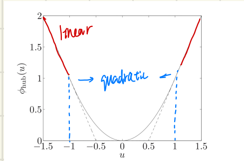
quadratic: $\phi(u) = u^2$

deadzone-linear: $\phi(u) = \max\{0, |u|-a\}$

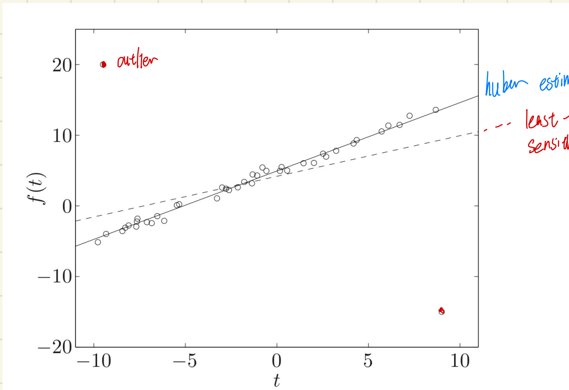
log-barrier with limit a: $\phi(u) = \begin{cases} -a^2 \cdot \log(1-(u/a)^2) & |u| < a \\ \infty & \text{otherwise} \end{cases}$

eg. Huber penalty function (with parameter M)

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



linear growth for large u makes approximation less sensitive to outliers



huber estimate

--- least-square approximation, sensitive to outliers since the loss function takes square of error

· Least norm problems

$$\begin{aligned} \min.\ & \|x\| \\ \text{s.t.}\ & Ax = b \end{aligned}$$

$\|\cdot\|$ is any norm

$A \in \mathbb{R}^{m \times n}$ $m \le n$

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

estimation: $b = Ax$ are perfect measurement of x; x^* is the smallest estimate consistent with measure

design: x are design variables; b are constraints; x^* is the most efficient design

eg least-square solution of linear equations

$$\begin{aligned} \min.\ & \|x\|_2^2 \\ \text{s.t.}\ & Ax = b \end{aligned}$$

$$L(x, v) = \|x\|_2^2 + v^T(Ax - b)$$

$$\nabla L_x = 2x + A^T v$$

$\Downarrow$ KKT condition

$$\begin{cases} Ax = b \\ 2x + A^T v = 0 \end{cases} \Rightarrow -\tfrac{1}{2} A A^T v = b \Rightarrow \begin{cases} v = -2(AA^T)^{-1} b \\ x = A^T (AA^T)^{-1} b \end{cases}$$

eg. sparse solution of linear equations $\|\cdot\|_1$

$$\begin{aligned} \min.\ & \mathbb{1}^T y \\ \text{s.t.}\ & -y \le x \le y \\ & Ax = b \end{aligned}$$

eg. least-penalty problem

$$\begin{aligned} \min.\ & \phi(x_1) + \cdots + \phi(x_n) \\ \text{s.t.}\ & Ax = b \end{aligned}$$

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

- Regularized approximation

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

$A \in \mathbb{R}^{m \times n}$, can be differem norms for $\|Ax - b\|$ and $\|x\|$
find good approximation $Ax \approx b$ with small x

size of x is related to the sensitivity of Ax to uncertainty in A

estimation: linear measurement model $y = Ax + v$, with prior knowledge that $\|x\|$ is small
design: small x is cheaper or more efficient

Scalarized problem

$$\min.\ \|Ax - b\| + \gamma \|x\|$$

Tikhonov regularization

$$\min.\ \|Ax - b\|_2^2 + \delta \|x\|_2^2 \Rightarrow \min.\ \left\| \begin{bmatrix} A \\ \sqrt{\delta} I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2$$

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

eg. Optimal input design
consider a linear dynamical system with scalar input sequence $u(0), \ldots, u(N)$;
output sequence $y(0), y(1), \cdots, y(N)$; impulse response $h(0), \cdots h(N)$

multicriterion problem with 3 objectives

1. tracking error with desired output : $J_{track} = \sum_{t=0}^{N} \left(y(t) - y_{des}(t) \right)^2$
2. input magnitude : $J_{mag} = \sum_{t=0}^{N} u(t)^2$
3. input variation : $J_{der} = \sum_{t=0}^{N-1} \left(u(t+1) - u(t) \right)^2$

$$\min. \ J_{track} + \delta J_{der} + \eta \cdot J_{mag}$$

- Signal reconstruction

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

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

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

$\hat{x}$ is estimate of x

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

eg. quadratic smoothing, total variation smoothing

$$\phi_{quad}(\hat{x}) = \sum_{i=0}^{n-1} (x_{i+1} - x_i)^2 \qquad \phi_{tv}(\hat{x}) = \sum_{i=0}^{n-1} |x_{i+1} - x_i|$$

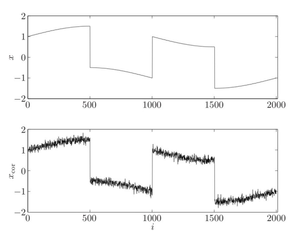
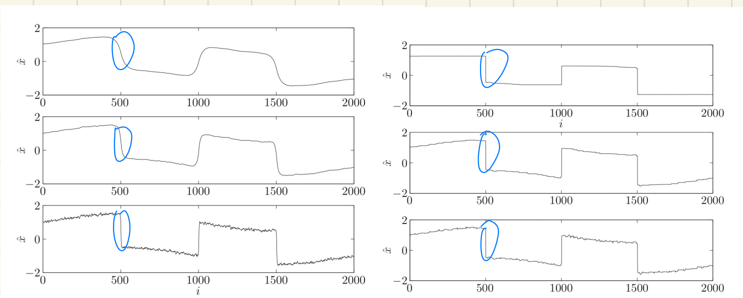


Figure 6.11 A signal $x \in \mathbf{R}^{2000}$, and the corrupted signal $x_{cor} \in \mathbf{R}^{2000}$. The noise is rapidly varying, and the signal is mostly smooth, with a few rapid variations.



quadratic smoothing loses the jump

total variation smoothing preserves the jump

- Robust approximation

min. $\|Ax - b\|$ with uncertain A

1. Stochastic approach

assume $A = \bar{A} + U$ where U is a random variable $E[U] = 0$ $E[U^T U] = P$

$$
\begin{aligned}
E\|Ax-b\|_2^2 &= E\,\|\bar{A}x-b+Ux\|_2^2 \\
&= E\,\|\bar{A}x-b\|_2^2 + 2E\,(\bar{A}x-b)^T Ux + E\,(Ux)^T Ux \\
&= \|\bar{A}x-b\|_2^2 + x^T P x \qquad (\text{quadratic regularization})
\end{aligned}
$$

$$\Downarrow$$

$$\min.\ \|\bar{A}x-b\|_2^2 + \|P^{1/2}x\|_2^2$$

$$x^* = (\bar{A}^T\bar{A}+P)^{-1}\bar{A}^T b$$

(for $P=\delta I$, it's the Tikhonov regularized problem)

$$\min\ \|\bar{A}x-b\|_2^2 + \delta\|x\|_2^2 \qquad U_{ij} \text{ zero mean, variance } \delta/m$$

2. worst-case approach

$$\mathcal{A} = \{\bar{A} + u_1A_1 + \cdots + u_pA_p \mid \|u\|_2 \le 1\}$$

(a ellipsoid: an unit ball under an affine mapping) from euclidean into matrices space

$$\Downarrow$$

$$\min.\ \sup_{A\in\mathcal{A}} \|Ax-b\|_2^2$$

$$\Downarrow$$

$$\min.\ \sup_{\|u\|_2\le 1} \|P(x)\cdot u + q(x)\|_2^2 \qquad P(x) = [A_1x\ \ A_2x\ \cdots\ A_px] \qquad q(x) = \bar{A}x - b$$

$$
\begin{aligned}
\max.\ & \|P\cdot u + q\|_2^2 \\
\text{s.t.}\ & \|u\|_2^2 \le 1
\end{aligned}
$$

(even non-convex, can be readily solved (without q term) max singular value)

$$\Downarrow \text{dual}$$

$$
\begin{aligned}
\min.\ & t+\lambda \\
\text{s.t.}\ & \begin{bmatrix} I & P & q \\ P^T & \lambda I & 0 \\ q^T & 0 & t \end{bmatrix} \succeq 0
\end{aligned}
$$

$$\Downarrow$$

$$
\begin{aligned}
\min.\ & t+\lambda \\
\text{s.t.}\ & \begin{bmatrix} I & P(x) & q(x) \\ P^T & \lambda I & 0 \\ q^T & 0 & t \end{bmatrix} \succeq 0
\end{aligned}
\qquad \text{SDP}
$$

- worst-case approximation (detailed)

describe the uncertainty by a set of possible values for A: $A \in \mathcal{A} \subseteq \mathbb{R}^{m \times n}$
(nonempty and bounded)

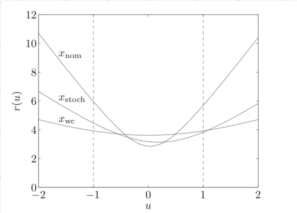
worst-case error: $e_{wc} = \sup \{ \|Ax-b\| \mid A \in \mathcal{A} \}$ (always a cvx function of x)

$\min_x \ \sup \{ \|Ax-b\| \mid A \in \mathcal{A} \}$

eg. comparison of stochastic and worst-case robust approximation
consider the least-square problem

$\min \|A(u)x - b\|$ $\qquad A(u) = A_0 + uA_1$

u is an uncertain parameter.



1. Finite set

$\mathcal{A} = \{A_1, A_2 \cdots A_k\}$

$\Downarrow$

$\min_x \ \sup_{A \in \mathcal{A}} \|Ax-b\|$ $\xLeftrightarrow{\text{equivalent}}$ $\min_x . \ \sup \{ \|Ax-b\| \mid A \in \text{Convexhull}(A_1 \cdots A_k) \}$

$\Downarrow$ epigraph form

min. t.
s.t. $\|A_i x - b\| \le t$ (SOCP for $\|\cdot\|_2$, LP for $\|\cdot\|_1$ and $\|\cdot\|_\infty$)

2. Norm bound error

here the uncertainty set $\mathcal{A}$ is a norm ball

matrix norm

$\|A\|_p = \sup_x \{ \|Ax\|_p \mid \|x\|_p = 1 \}$

max sigular value if $p=2$

$\mathcal{A} = \{ \bar{A} + U \mid \|U\| \le a \}$

$e_{wc}(x) = \sup_u \{ \|\bar{A}x - b + Ux\| \mid \|U\| \le a \}$

(i.e. euclidean norm on $\mathbb{R}^n$, max sigular value on $\mathbb{R}^{m \times n}$)

$$\max_{U}\ \|\bar{A}x-b+Ux\|_2$$
$$\|U\|_2 \le a \qquad \text{(max sigular value)}$$

$$U^* = a\cdot\frac{\bar{A}x-b}{\|\bar{A}x-b\|}\cdot\frac{x^T}{\|x\|_2}$$

$$\begin{aligned}
&\|\bar{A}x-b+Ux\|_2^2\\
&=(\bar{A}x-b+Ux)^T(\bar{A}x-b+Ux)\\
&=(\bar{A}x-b)^T(\bar{A}x-b)+2(Ux)^T(\bar{A}x-b)+(Ux)^T(Ux)
\end{aligned}$$

$$\begin{aligned}
&2(Ux)^T(\bar{A}x-b)+(Ux)^T(Ux)\\
&\le 2\|Ux\|_2\cdot\|\bar{A}x-b\|+\|Ux\|_2^2 \quad \text{(achieved when } Ux \parallel \bar{A}x-b\text{)}\\
&\le 2\|U\|_2\cdot\|x\|_2\,\|\bar{A}x-b\|+\|U\|_2^2\cdot\|x\|_2^2 \quad \text{(achieved when } x \text{ is a linear combination of basis of rowspace (U) that has the maximum Sigular value)}
\end{aligned}$$

$$\begin{array}{c} \qquad n \\ m\left[\frac{\bar{A}x-b}{\|\ \|_2}\right]\left[\ x/\|x\|_2\ \right] \end{array} \qquad \left[\ U^*\ \right]\left[\ x\ \right]$$

$$U^* = a\cdot\frac{\bar{A}x-b}{\|\bar{A}x-b\|}\cdot\frac{x^T}{\|x\|_2}$$

rank 1

only 1 Sigular value α

$$\begin{aligned}
e_{wc}(x) &= \sup_{U}\{\|\bar{A}x-b+Ux\| \mid \|U\|\le a\}\\
&= \left\|\bar{A}x-b+a\cdot\frac{\bar{A}x-b}{\|\bar{A}x-b\|_2}\cdot\frac{x^T}{\|x\|_2}x\right\|_2\\
&= \|\bar{A}x-b\|_2+a\cdot\|x\|_2
\end{aligned}$$

$$\Downarrow$$

$$\min\ \|\bar{A}x-b\|_2+a\cdot\|x\|_2 \qquad \text{(regularized norm problem)}$$

$$\Downarrow$$

$$\begin{aligned}
\min\ & t_1+at_2\\
\text{s.t. } & \|\bar{A}x-b\|_2\le t_1 \qquad \text{(SOCP)}\\
& \|x\|_2\le t_2
\end{aligned}$$

3. Uncertainty ellipsoid

describe the variation in A by giving an ellipsoid of possible values for each row

$$\mathcal{A}=\left\{\begin{bmatrix} -a_1- \\ \vdots \\ -a_m- \end{bmatrix} \;\middle|\; a_i\in\mathcal{E}_i\right\}$$

$$\mathcal{E}_i = \{ \bar{a}_i + P_i u \mid \|u\|_2 \le 1 \}$$

allow P_i to have a nontrivial nullspace

$$\sup_{a_i \in \mathcal{E}_i} |a_i^T x - b_i| = \sup \{ |\bar{a}_i^T x - b_i + (P_i u)^T x| \mid \|u\|_2 \le 1 \}$$
$$= \sup \{ |\bar{a}_i^T x - b_i + u^T (P_i^T x)| \mid \|u\|_2 \le 1 \}$$
$$= |\bar{a}_i^T x - b_i| + \|P_i^T x\|_2$$

$$e_{wc}(x) = \sup \{ \|Ax - b\|_2 \mid a_i \in \mathcal{E}_i \}$$
$$= \left[\sum_i \left(\sup \{ |a_i^T x - b_i| \mid a_i \in \mathcal{E}_i \} \right)^2 \right]^{1/2}$$
$$= \left[\sum_i \left(|\bar{a}_i^T x - b_i| + \|P_i^T x\|_2 \right)^2 \right]^{1/2}$$
$$\Downarrow$$

$$\begin{aligned} \min\ & \|t\|_2 \\ \text{s.t. } & |\bar{a}_i^T x - b_i| + \|P_i^T x\|_2 \le t_i \end{aligned}$$
$$\Downarrow$$

$$\begin{aligned} \min.\ & \|t\|_2 \\ \text{s.t. } & \bar{a}_i^T x - b_i + \|P_i^T x\|_2 \le t_i \\ & -a_i^T x + b_i + \|P_i^T x\|_2 \le t_i \end{aligned} \qquad (\text{SOCP})$$

4. Norm bound error with linear structure

$$\mathcal{A} = \{ \bar{A} + u_1 A_1 + \cdots + u_p A_p \mid \|u\| \le 1 \}$$ (a norm ball under an affine transformation)
(a generalization of norm bound $\mathcal{A} = \{ \bar{A} + U \mid \|U\| \le 1 \}$)

$$e_{wc}(x) = \sup_{A \in \mathcal{A}} \|Ax - b\|$$
$$= \sup_{\|u\| \le 1} \|(\bar{A} + u_1 A_1 + \cdots + u_p A_p) x - b\|$$
$$= \sup_{\|u\| \le 1} \|P(x) u + q(x)\| \qquad \left(P(x) = \begin{bmatrix} A_1 x & A_2 x & \cdots & A_p x \end{bmatrix} \quad q(x) = \begin{bmatrix} \bar{A} x - b \end{bmatrix} \right)$$

eg. robust Chebyshev approximation

$$\min.\ e_{wc}(x) = \sup_{\|u\|_\infty \le 1} \|(\bar{A} + u_1 A_1 + \cdots + u_p A_p) x - b\|_\infty$$

$$= \sup_{\|u\|_\infty \le 1} \| P(x) u + q(x) \|_\infty$$

$$= \max_i \sup_{\|u\|_\infty \le 1} \| p_i(x)^T u + q_i(x) \|$$

$$= \max_i \left(\|p_i(x)\|_1 + q_i(x) \right)$$