Robust Solutions to l_1, l_2, and l_∞ Uncertain Linear Approximation Problems using Convex Optimization

Haitham Hindi Stephen Boyd
Information Systems Lab, Stanford University

June 22, 1998

Contents

Motivation

- **Original Problem**

$$A\,x \approx b$$

- **Perturbed Problem**

$$(A + \Delta A)x \approx (b + \Delta b)$$

- Roughly speaking, want:
  1. **Desensitize** solution x to perturbations $i.e.$, don't want wild variations in x with small $[\Delta A \Delta b]$
  2. **Guarantee performance** in face of uncertainty in data $i.e.$, don't want wild variations in error with small $[\Delta A \Delta b]$

- Apps: static control, signal processing, system ID

Other Research

- **Total Least Squares**
  ‘80: Golub & Van Loan, ...

- **Robust Least Squares** (Minimax l_2-Norm)
  ’97 El-Ghaoui & Lebret
  ’97 Sayed, Golub, Chandrasekaran, Nascimento

- **Our Contribution**
  Generalize minimax formulation to l_1 and l_∞ norms
  Introduce and solve stochastic case

Norms of Vectors and Matrices

- **Vectors** $x \in \mathbf{R}^n$

$$\|x\|_1 \triangleq \sum_{i=1}^{n} |x_i|$$

$$\|x\|_2 \triangleq \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

$$\|x\|_\infty \triangleq \max_{i=1,\ldots,n} |x_i|$$

- **Matrices** $A \in \mathbf{R}^{m\times n}$

$$\|A\|_1 \triangleq \max_{j=1,\ldots,n} \sum_{i=1}^{m} |a_{ij}| \quad = \quad \text{max-col-sum}$$

$$\|A\|_2 \triangleq \sigma_{\max}(A) \quad = \quad \text{max sing. value}$$

$$\|A\|_\infty \triangleq \max_{i=1,\ldots,m} \sum_{j=1}^{n} |a_{ij}| \quad = \quad \text{max-row-sum}$$

Why All These Formulations?

- From control theory, know choice of norm (l_1, H_2, H_∞) and formulation (stochastic vs deterministic) has big influence on design - same for $\mathbf{R}^n$.

- $\mathbf{l_1}$: good when data contains large **outliers** since it does not put relatively more weight on large errors

- $\mathbf{l_2}$: "Default norm" in most engineering applications

- $\mathbf{l_\infty}$: allows us to handle case of **truly independent** perturbations esp. in **structured** case

- **Stochastic**: useful when have statistics of perturbations

Solving $A\,x \approx b$

Least Squares Approximation:

1. define **least squares error** function

$$\boldsymbol{\varepsilon}_{\mathrm{ls}}(x) \stackrel{\Delta}{=} \|A\,x - b\|_2$$

2. solve: $$\min_x \ \boldsymbol{\varepsilon}_{\mathrm{ls}}(x)$$

Robust Linear Approximation:

1. define **robustified error** function based on information about perturbations $[\Delta A\,\Delta b]$

$$\boldsymbol{\varepsilon}_{\mathrm{rob}}(x)$$

2. solve: $$\min_x \ \boldsymbol{\varepsilon}_{\mathrm{rob}}(x)$$

For Robust Approx approach to be useful, need:

- $\boldsymbol{\varepsilon}_{\mathrm{rob}}(x)$ must be **efficiently computable**
- $\min_x \ \boldsymbol{\varepsilon}_{\mathrm{rob}}(x)$ must be **efficiently computable**

Minimax Approach

- If $[\Delta A\, \Delta b]$ are known to lie in some **compact set** Ω, can

  1. define

  $$\boldsymbol{\varepsilon}_{\Omega p}(x) \stackrel{\Delta}{=} \max_{[\Delta A\ \Delta b] \in\ \Omega} \|(A+\Delta A)x-(b+\Delta b)\|_p$$

  2. solve

  $$\min_x \max_{[\Delta A\ \Delta b] \in\ \Omega} \|(A+\Delta A)x-(b+\Delta b)\|_p$$

- For **fixed** $[\Delta A\ \Delta b]$, $\|(A+\Delta A)x-(b+\Delta b)\|_p$ is **convex** in x.
- Hence $\boldsymbol{\varepsilon}_{\Omega p}$ is **convex** function of x.
- But that doesn't mean it's easy to compute!

Unstructured vs Structured Uncertainty Models

- If very little known about $[\Delta A\, \Delta b]$, can use model

  $$\mathcal{U} = \{\ [\Delta A\, \Delta b]\ \mid\ \|[\Delta A\, \Delta b]\|_p \le \rho\ \}$$

  and

  $$\boldsymbol{\varepsilon}_{up}(x) = \max_{\|[\Delta A\ \Delta b]\|_p \le \rho} \|(A + \Delta A)x - (b + \Delta b)\|_p$$

- If there's **structure** $e.g. A$ is sparse or Toeplitz, can **reduce conservatism** by using

  $$\mathcal{S} = \{\ [\Delta A\, \Delta b]\ \mid\ \Delta A = \sum_{i=1}^{L} A_i\, \delta_i,\quad \Delta b = \sum_{i=1}^{L} b_i\, \delta_i,\quad \|\delta\|_p \le \rho\ \}$$

  which gives

  $$\boldsymbol{\varepsilon}_{sp}(x) = \max_{\|\delta\|_p \le \rho} \left\| (A_0 + \sum_{i=1}^{L} A_i \delta_i)x - (b_0 + \sum_{i=1}^{L} b_i \delta_i) \right\|_p$$

Stochastic Formulation

- If know statistics of $[\Delta A \, \Delta b]$, can

  1. define **stochastic** error function

  $$\varepsilon_{us}(x) \stackrel{\Delta}{=} \mathop{\mathbf{E}}_{[\Delta A \, \Delta b]} \|(A + \Delta A)x - (b + \Delta b)\|_2^2$$

  2. solve

  $$\min_x \mathop{\mathbf{E}}_{[\Delta A \, \Delta b]} \|(A + \Delta A)x - (b + \Delta b)\|_2^2$$

- For **structured** perturbations, if know statistics of δ, can

  1. define **structured stochastic** error function

  $$\varepsilon_{ss}(x) \stackrel{\Delta}{=} \mathop{\mathbf{E}}_{\delta} \|(A_0 + \sum_{i=1}^{L} A_i \, \delta_i)x - (b_0 + \sum_{i=1}^{L} b_i \, \delta_i)\|_2^2$$

  2. solve

  $$\min_x \mathop{\mathbf{E}}_{\delta} \|(A_0 + \sum_{i=1}^{L} A_i \, \delta_i)x - (b_0 + \sum_{i=1}^{L} b_i \, \delta_i)\|_2^2$$

- Both ε_{us} and ε_{ss} are **convex** in x (positively weighted integrals of convex functions).

Computing ε_{up}

- Recall that

$$\varepsilon_{up}(x) = \max_{\|[\Delta A \ \Delta b]\|_p \le \rho} \|(A + \Delta A)x - (b + \Delta b)\|_p$$

- To evaluate ε_{up} for **fixed** x, must solve **maximization** problem in $[\Delta A \ \Delta b]$.

- **triangle inequality** $+$ **norm defs** $\rightarrow$ **upper bound** for any $\|[\Delta A \ \Delta b]\|_p \le \rho$:

$$\|(A + \Delta A)x - (b + \Delta b)\|_p \le \|Ax - b\|_p + \rho\|(x, -1)\|_p$$

$$\max_{\|[\Delta A \ \Delta b]\|_p \le \rho} \|(A + \Delta A)x - (b + \Delta b)\|_p \le \|Ax - b\|_p + \rho\|(x, -1)\|_p$$

- Using properties of norm, exhibit **worst case perturbation** $[\Delta A \ \Delta b]_{\text{wc}}$ which achieves **equality**

$\Longrightarrow$ upper bound is **maximum**

Result: Unstructured Minimax

Theorem:
For $p = 1, 2, \infty$ we have

$$\boldsymbol{\varepsilon}_{up}(x) = \|Ax - b\|_p + \rho\|(x, -1)\|_p$$

and robust approx problem becomes

$$\min_x \{\|Ax - b\|_p + \rho\|(x, -1)\|_p\}.$$

Remarks

1. For $p = 1, \infty$, can solve as **Linear Program**
2. For $p = 2$ can solve a **second order cone Program**
3. **Tradeoff** accuracy ($\|Ax - b\|_p$) vs large solutions ($\|(x, -1)\|_p$)
4. Least squares problem $\longrightarrow$ Convex optimization problem of same size.

Computing Structured Error

- Recall that

$$\boldsymbol{\varepsilon}_{sp}(x) = \max_{\|\delta\|_p \le \rho} \left\| (A_0 + \sum_{i=1}^{L} A_i \delta_i) x - (b_0 + \sum_{i=1}^{L} b_i \delta_i) \right\|_p$$
$$= \max_{\|\delta\|_p \le \rho} \left\| \sum_{i=1}^{L} (A_i x - b_i) \delta_i + (A_0 x - b_0) \right\|_p$$

- Define

$$F(x) = [(A_1\, x - b_1) \ldots (A_L\, x - b_L)], \qquad g(x) = -(A_0\, x - b_0)$$

then

$$\boldsymbol{\varepsilon}_{sp}(x) = \max_{\|\delta\|_p \le \rho} \|F\,\delta - g\|_p$$

- So to compute $\boldsymbol{\varepsilon}_{sp}$ must solve convex **maximization** problem

- $p = 2$ use $\mathcal{S}$-procedure
$p = 1, \infty$ use fact that max cvx fn over compact set attained at extreme point (ofcourse don't search **all** vertices!).

Result: Structured Minimax

Synopsis

- For $p = 1, 2, \infty$, problems can be solved efficiently using **convex optimization** as Linear and Semidefinite programs
- The optimization problems are a factor of L larger (L is number of structured perturbations)
- Unlike unstructured case, solutions don't look alike. Because use different techniques to solve the maximization problems

Stochastic Case

Theorem:(Unstructured)
Let $R = \mathbf{E}[\Delta A\,\Delta b]^T[\Delta A\,\Delta b]$. Then

$$\boldsymbol{\varepsilon}_{us}(x) = \|A\,x - b\|_2^2 + (x, -1)^T R\,(x, -1)$$

and the robust approximation problem is solved by

$$\min_x \{\|A\,x - b\|_2^2 + (x, -1)^T R\,(x, -1)\}$$

Theorem:(Structured)
Let $R_\delta = \mathbf{E}[\delta\,\delta^T]$, $\tilde{A}_i, \tilde{b}_i$ be "reshuffled" versions of A_i, b_i, and $S = \left(\sum_{i=1}^{L}[\tilde{A}_i\,\tilde{b}_i]\right)^T R_\delta \left(\sum_{i=j}^{L}[\tilde{A}_j\,\tilde{b}_j]\right)$. Then

$$\boldsymbol{\varepsilon}_{us}(x) = \|A_0\,x - b_0\|_2^2 + (x, -1)^T S\,(x, -1)$$

and the robust approximation problem is solved by

$$\min_x \{\|A_0\,x - b_0\|_2^2 + (x, -1)^T S\,(x, -1)\}$$

- **Proofs:** direct computation, assume **zero mean**
- Note **tradeoff** again
- Both problems are **pure quadratics** - can be solved as augmented **weighted least squares** problems

Conclusion

- Uncertainty in data can be incorporated **explicitly** into algorithms for linear approximation

- Approximation problems become **convex optimization** problems which can be solved efficiently

- Rich variety of formulations: minimax, stochastic, different norms...

- Can **exploit structure** to reduce conservatism

- Stochastic and unstructured minimax yield problems of approx same size as original

- Structured minimax yield problems a factor of L larger