On the distance to ill-posedness

Javier Peña
Carnegie Mellon University

Fields Institute
Toronto, Canada
October 2009

Preamble

Theorem (Eckart-Young, 1936)

Assume $A \in \mathbb{R}^{n\times n} \setminus \mathrm{Sing}$. *Then*

$$\mathrm{dist}(A, \mathrm{Sing}) = \frac{1}{\|A^{-1}\|} = \max\{\delta : \delta\mathbb{B}_{\mathbb{R}^n} \subseteq A\mathbb{B}_{\mathbb{R}^n}\}.$$

Theorem (distance to rank-deficiency)

Assume $A \in \mathbb{R}^{m\times n}$ *is of rank* $m \leq n$. *Then*

$$\begin{aligned}\mathrm{dist}(A, \Sigma) &= \max\{\delta : \delta\mathbb{B}_{\mathbb{R}^m} \subseteq A\mathbb{B}_{\mathbb{R}^n}\} \\ &= \frac{1}{\max_{v\in\mathbb{B}_{R^m}} \min\{\|x\| : Ax = v\}} = \frac{1}{\text{``}\|A^{-1}\|\text{''}}.\end{aligned}$$

Σ = rank-deficient matrices.

Proof of distance to rank-deficiency Theorem

Alternative

$$A \in \Sigma \Leftrightarrow \exists y \neq 0, A^{\mathsf{T}} y = 0.$$

Norm-duality

$$\text{``}\|A^{-1}\|\text{''} = \max_{v \in \mathbb{B}_{\mathbb{R}^m}} \min\{\|x\| : Ax = v\}$$
$$= \max_{u \in \mathbb{B}_{\mathbb{R}^n}} \max\left\{\|y\| : A^{\mathsf{T}} y + u = 0\right\} = \text{``}\|A^{-\mathsf{T}}\|\text{''}.$$

Rank-one construction

Find $v \in \mathbb{R}^m$ and $u \in \mathbb{R}^n$ such that

$$A + vu^{\mathsf{T}} \in \Sigma.$$

Theme

Extensions of the Eckart-Young Theorem:

- From linear systems of equations to linear systems of constraints
- From unstructured (arbitrary) perturbations to structured (e.g., sparse) perturbations
- Connection with “best-conditioned” solutions

Why does this matter?

- Distance to ill-posedness leads to a notion of condition number for optimization (Renegar)
- Conditioning is related to accuracy and performance of algorithms
- Work along these lines by: Belloni, Cheung, Cucker, Dunagan, Epelman, Filipowski, Freund, Renegar, Vempala, etc.

From linear equations to linear constraints

Notice:

Given $A \in \mathbb{R}^{m \times n}$ with $m \leq n$, we have $A \notin \Sigma \Leftrightarrow A\mathbb{R}^n = \mathbb{R}^m$.

Equivalently, $A \notin \Sigma$ if and only if $Ax = b$ has a solution for all $b \in \mathbb{R}^n$.

How do we extend this to constraint systems?

Assume $K \subseteq \mathbb{R}^n$ is a closed convex cone (e.g., $K = \mathbb{R}^n_+$).
Given $A \in \mathbb{R}^{m \times n}$ with $m \leq n$ consider

$$Ax = b, \; x \in K \quad (\text{e.g., } Ax = b, \; x \geq 0)$$

and

$$c - A^\mathsf{T} y \in K^* \quad (\text{e.g., } A^\mathsf{T} y \leq c)$$

for $b \in \mathbb{R}^m, \; c \in \mathbb{R}^n$.

Well-posed and ill-posed matrices

Throughout this talk:

Assume $K \subseteq \mathbb{R}^n$ is a closed convex cone (e.g., $K = \mathbb{R}^n_+$) and $m \leq n$.

Define

$$\mathcal{P} := \{A \in \mathbb{R}^{m \times n} : AK = \mathbb{R}^m\},$$
$$\mathcal{D} := \{A \in \mathbb{R}^{m \times n} : A^\mathsf{T}\mathbb{R}^m + K^* = \mathbb{R}^n\}.$$

Notice

- $A \in \mathcal{P} \Leftrightarrow Ax = b, x \in K$ has a solution for all $b \in \mathbb{R}^m$
- $A \in \mathcal{D} \Leftrightarrow c - A^\mathsf{T}y \in K^*$ has a solution for all $c \in \mathbb{R}^n$

Ill-posed instances

$$\Sigma := \mathbb{R}^{m \times n} \setminus (\mathcal{P} \cup \mathcal{D}).$$

Theorem (Renegar, 1995)

(a) *If* $A \in \mathcal{P}$ *then*

$$\operatorname{dist}(A, \Sigma) = \max\{\delta : \delta \mathbb{B}_{\mathbb{R}^m} \subseteq A(\mathbb{B}_{\mathbb{R}^n} \cap K)\}.$$

(b) *If* $A \in \mathcal{D}$ *then*

$$\operatorname{dist}(A, \Sigma) = \max\{\delta : \delta \mathbb{B}_{\mathbb{R}^n} \subseteq A^{\mathsf{T}} \mathbb{B}_{\mathbb{R}^m} + K^*\}.$$

A more general setting: sublinear mappings

Definition

$F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *sublinear* if $\mathrm{graph}(F) = \{(x, y) : y \in F(x)\}$ is a convex cone. In that case

$$\|F\|^- := \sup_{x \in \mathbb{B}_{R^n}} \inf_y \{\|y\| : y \in F(x)\}.$$

Theorem (Lewis, 1998)

Assume $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a sublinear mapping with closed graph and F is surjective. Then

$$\inf\{\|G\| : G \in \mathbb{R}^{m \times n},\ F + G \text{ is not surjective}\} = \frac{1}{\|F^{-1}\|^-}.$$

Conic systems: special case of sublinear mappings

Given $K \subseteq \mathbb{R}^n$ and $A \in \mathbb{R}^{m\times n}$ consider

$$F_{A,\mathcal{P}}(x) := \begin{cases} Ax, & \text{if } x \in K \\ \emptyset & \text{otherwise} \end{cases}$$

Then $A \in \mathcal{P} \Leftrightarrow F_{A,\mathcal{P}}$ surjective. Renegar's distance Theorem (a) follows.

Similarly, consider

$$F_{A,\mathcal{D}}(y) := A^{\mathsf{T}}y + K^*.$$

Then $A \in \mathcal{D} \Leftrightarrow F_{A,\mathcal{D}}$ surjective. Renegar's distance Theorem (b) follows.

Structured distance to ill-posedness

Observe

- Previous distance theorems assume unstructured (arbitrary) data perturbations.
- Often data perturbations are restricted to some specific structure, e.g., sparsity or slack variables.
- Ignoring such structure may lead to substantial underestimation of the sensible distance to ill-posedness.

Structured distance to ill-posedness

Example

Take $K = \mathbb{R}^n_+$ and

$$A = \begin{bmatrix} 0.1 & -1 & 0 & \cdots & 0 \\ 0 & 0.1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0.1 & -1 \end{bmatrix} \in \mathcal{P}$$

Unstructured distance to ill-posedness $= (0.1)^{n-1}$
Structured (sparse) distance $= 0.1$

Single block structure

Suppose we are only allowed to perturb a block of A: Assume $k \leq m$, $\ell \leq n$ and put

$$\Delta := \left\{ \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} : B \in \mathbb{R}^{k\times\ell} \right\}.$$

Proposition (P. 1998)

Assume $A \in \mathcal{P}$. Then

$$\begin{aligned} \mathrm{dist}_\Delta(A,\Sigma) &= \max\left\{\delta : \delta\mathbb{B}_{\mathbb{R}^k} \subseteq \{Ax : x \in K, x_{1:\ell} \in \mathbb{B}_{\mathbb{R}^\ell}\}\right\} \\ &= \frac{1}{\max\limits_{v\in\mathbb{B}_{\mathbb{R}^k}} \min\{\|x_{1:\ell}\| : Ax = v, x \in K\}} \\ &= \frac{1}{\text{``}\|A^{-1}\|\text{''}}. \end{aligned}$$

Proof of single block-structured distance Proposition

Alternative

$$A \notin \mathcal{P} \Leftrightarrow \exists y \neq 0, A^{\mathsf{T}} y \in K^*.$$

Norm-duality

$$\begin{aligned}
\text{“}\|A^{-1}\|^{-}\text{”} &= \max_{v \in \mathbb{B}_{\mathbb{R}^k}} \min\{\|x_{1:\ell}\| : Ax = v, x \in K\} \\
&= \max_{u \in \mathbb{B}_{\mathbb{R}^\ell}} \max\left\{\|y_{1:k}\| : A^{\mathsf{T}} y + u \in K^*\right\} = \text{“}\|A^{-\mathsf{T}}\|^{+}\text{”}.
\end{aligned}$$

Rank-one construction

Find $v \in \mathbb{R}^\ell$ and $u \in \mathbb{R}^k$ such that

$$A + \begin{bmatrix} vu^{\mathsf{T}} & 0 \\ 0 & 0 \end{bmatrix} \notin \mathcal{P}.$$

Sublinear mappings: special case of conic systems

Given a sublinear mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, put

$$K_F := \text{graph}(F) \subseteq \mathbb{R}^{n+m} \text{ and } A_F := \begin{bmatrix} 0 & I_m \end{bmatrix} \in \mathbb{R}^{m\times(n+m)}.$$

Observe

- F surjective $\Leftrightarrow A_F \in \mathcal{P}$
- For $B \in \mathbb{R}^{m\times n}$, $F + B$ not surjective $\Leftrightarrow A_F + \begin{bmatrix} B & 0 \end{bmatrix} \notin \mathcal{P}$.

Lewis's distance theorem then follows from P's single block-structured distance proposition for conic systems.

Multiple block structure

Suppose $X_j \subseteq \mathbb{R}^n$, $Y_j \subseteq \mathbb{R}^m$, $j = 1, \ldots, k$. Let

$$\Delta := \left\{ B : B = \sum_j B_j,\ B_j \in L(X_j, Y_j) \right\},$$

and for $B = \sum_i B_j \in \Delta$, let $\|B\|_\Delta := \max_j \|B_j\|$.

Theorem (P. 2003)

Assume $A \in \mathcal{P}$. Then

$$\operatorname{dist}_\Delta(A, \Sigma) =$$
$$\left(\sup_{v^j \in \mathbb{B}_{Y_j}} \inf_{x,z} \left\{ \max_i \frac{\|x_i\|}{z_i} : z > 0, Ax = \sum z_j v^j,\ x \in K \right\} \right)^{-1}$$

Right-hand side: sort of "$1/\|A^{-1}\|$".

Proof of multiple block-structured distance Theorem

Alternative

$$A \notin \mathcal{P} \Leftrightarrow \exists y \neq 0, A^{\mathsf{T}} y \in K^*.$$

Norm-duality

$$\text{``}\|A^{-1}\|^{-}\text{''} = \sup_{v^j \in \mathbb{B}_{Y_j}} \inf_{x,z} \left\{ \max_i \frac{\|x_i\|}{z_i} : z > 0, Ax = \sum z_j v^j,\ x \in K \right\}$$
$$= \sup_{u^j \in \mathbb{B}_{X_j}} \sup \left\{ \min_{i, u^i \neq 0} \frac{\|y_i\|}{\|u^i\|} : A^{\mathsf{T}} y + \sum u^j \in K^* \right\} = \text{``}\|A^{-\mathsf{T}}\|^{+}\text{''}.$$

Rank-k construction

Find $v_j \in Y_j,\ u_j \in X_j,\ j = 1, \ldots, k$ such that

$$A + \sum_j v_j u_j^{\mathsf{T}} \notin \mathcal{P}.$$

Componentwise distance to singularity

Observation

Assume $A \in \mathbb{R}^{n\times n} \setminus \mathrm{Sing}$ and $B \in \mathbb{R}^{n\times n}$. Then

$$\inf\{|\delta| : A + \delta B \in \mathrm{Sing}\} = \frac{1}{\rho_0(A^{-1}B)}.$$

$\rho_0(\cdot)$ is the *real spectral radius:*

$$\rho_0(M) := \max\{|\lambda| : \lambda \text{ is a } \textit{real} \text{ eigenvalue of } M\}.$$

(If M has no real eigenvalues, $\rho_0(M) := 0$.)

Componentwise distance to singularity

Theorem (Rohn, 1989)

Assume $A \in \mathbb{R}^{n\times n} \setminus \mathrm{Sing}$ *and* $E \in \{0,1\}^{n\times n}$. *Then*

$$\inf\{\delta :\ \exists B \text{ with } |B| \le \delta\, E, A + B \in \mathrm{Sing}\} = \frac{1}{\max_{S_1,S_2} \rho_0(A^{-1}S_1 E S_2)},$$

max taken over signature matrices.
$S \in \{-1,1\}^{n\times n}$ *is a signature matrix if* $|S| = I$.

Rohn's Theorem can be recovered from multiple block-structured distance Theorem.

Distance to ill-posedness and best-conditioned solutions

For the remaining of this presentation

Assume $K = \mathbb{R}^n_+$, and given $A \in \mathbb{R}^{m\times n}$, write $A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$.

Goffin-Cheung-Cucker's condition number

Assume $a_i \neq 0,\ i = 1, \dots, n$. Define

$$v(A) := \max_{\|y\|=1} \min_{j=1,\dots,n} \frac{a_j^\mathsf{T} y}{\|a_j\|},\ \mathscr{C}(A) := \frac{1}{|v(A)|}.$$

Notice

- $A \in \mathcal{D} \Leftrightarrow v(A) > 0$
- $A \in \mathcal{P} \Leftrightarrow v(A) < 0$

Distance to ill-posedness and best-conditioned solutions

Geometric interpretation

When $A \in \mathcal{D}$, $v(A)$ is a measure of "thickness" of the cone

$$\{y : A^\mathsf{T} y \geq 0\}.$$

$v(A)$ is also a measure of the "best-conditioned" solution to

$$A^\mathsf{T} y \geq 0.$$

Distance to ill-posedness and best-conditioned solutions

Theorem (Cheung & Cucker, 2001)

Assume $a_i \neq 0,\ i = 1, \ldots, n$. *Then*

$$|v(A)| = \inf \left\{ \max_{i=1,\ldots,n} \frac{\|a_i - \tilde{a}_i\|}{\|a_i\|} : \tilde{A} \in \Sigma \right\}.$$

Remark

- This gives an identity between the best-conditioned solution and distance to ill-posedness of the system of constraints.
- The above distance theorem can be related to the block-structured distance theorem: The right hand side is a certain $\mathrm{dist}_\Delta(A, \Sigma)$.

Stratified distance to ill-posedness

Can we restrict the distance to ill-posedness to Σ?

Motivation

- When $K = \mathbb{R}^n$, $\Sigma =$ rank-deficient matrices.
- The set of ill-posed instances Σ can be written as

$$\Sigma = \Sigma_{m-1} \cup \Sigma_{m-2} \cup \cdots \cup \Sigma_1 \cup \Sigma_0$$

  $\Sigma_r =$ matrices with rank at most r.
- Given $A \in \Sigma_i \setminus \Sigma_{i-1}$,

$$\mathrm{dist}_{\Sigma_i}(A, \Sigma_{i-1}) = \sigma_i(A).$$

  $\sigma_i(A)$: i-th (smallest positive) singular value of A.

Stratified distance to ill-posedness

Consider again $K = \mathbb{R}^n_+$.

How can we stratify Σ?

Answer: Use a "canonical" partition $\mathscr{P}(A) = \{B, N\}$ of $\{1, \ldots, n\}$.

Proposition

Assume $A \in \mathbb{R}^{m \times n}$. Then there exists a unique partition $B \cup N = \{1, \ldots, n\}$ such that for some $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$

$$A_B x_B = 0,\ x_B > 0,\ A_B^\mathsf{T} y = 0,\ A_N^\mathsf{T} y > 0.$$

Observe

- $A \in \mathcal{D} \Leftrightarrow B = \emptyset$
- $A \in \mathcal{P} \Leftrightarrow N = \emptyset$ and $\operatorname{rank}(A) = m$.

Stratified distance to ill-posedness

Assume $A \in \mathbb{R}^{m\times n}$ and $\mathscr{P}(A) = \{B, N\}$. Define

$$L = \ker(A_B^\mathsf{T}) \subseteq \mathbb{R}^m, \text{ and } L_\perp = \mathrm{range}(A_B) \subseteq \mathbb{R}^m.$$

If $N \neq \emptyset$, define

$$v_N(A) := \max_{\substack{y\in L\\ \|y\|=1}} \min_{j\in N} \frac{a_j^\mathsf{T} y}{\|a_j\|}.$$

If $B \neq \emptyset$, define

$$v_B(A) = \max_{\substack{y\in L_\perp\\ \|y\|=1}} \min_{j\in B} \frac{a_j^\mathsf{T} y}{\|a_j\|}.$$

Stratified distance to ill-posedness

Theorem (Cheung-Cucker-P., 2008)

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

$$v_N(A) = \min_{\substack{\mathscr{P}(\tilde{A})\neq\mathscr{P}(A)\\ \tilde{A}_B=A_B}} \max_{j\in N} \frac{\|\tilde{a}_j - a_j\|}{\|a_j\|}$$

and

$$|v_B(A)| = \min_{\substack{\mathscr{P}(\tilde{A})\neq\mathscr{P}(A)\\ \tilde{A}_N=A_N\\ \ker(\tilde{A}_B^{\mathsf{T}})\supseteq L}} \max_{j\in B} \frac{\|\tilde{a}_j - a_j\|}{\|a_j\|}.$$

Conclusions

- Ill-posed matrices (for systems of constraints) are an extension of rank-deficient matrices (for systems of equations)
- The Eckart-Young distance Theorem and its proof extend to the distance to ill-posedness
- Similar distance theorems hold when restricted to certain manifolds.
- Relationships between distance to ill-posedness and "best-conditioned" solutions