Conditioning in optimization and variational analysis

Javier Peña
Carnegie Mellon University

November, 2003
Oxford University

Preamble: square matrices

Condition number of $A \in \mathbb{R}^{n \times n}$:

$$\kappa(A) = \|A\| \, \|A^{-1}\|.$$

(Recall: $\|A\| = \max\{\|Ax\| : \|x\| \leq 1\}$.)

Key parameter for problem

$$Ax = b.$$

- Sensitivity to data perturbations:

$$A(x + \delta x) = b + \delta b \Rightarrow \frac{\|\delta x\|}{\|x\|} \leq \kappa(A)\frac{\|\delta b\|}{\|b\|}.$$

- Performance of iterative methods:
  For A is symmetric and positive definite, k CG iterations yield

$$\frac{\|x_k - \bar{x}\|_A}{\|x_0 - \bar{x}\|_A} \leq 2\left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}\right)^k.$$

- Effects of finite-precision arithmetic.

⋆ **Measure of well-posedness:**

$$\begin{aligned} \frac{1}{\kappa(A)} &= \frac{\text{dist}(A, \text{Sing})}{\|A\|} \\ &= \frac{\min\{\|B\| : A + B \text{ is singular}\}}{\|A\|}. \end{aligned}$$

The latter follows from this classical identity:

Thm 1 (Eckart and Young) *Assume* $A \in \mathbb{R}^{n \times n}$ *is non-singular. Then*

$$\min\{\|B\| : A + B \text{ is singular}\} = \frac{1}{\|A^{-1}\|} = \max\{\delta : \delta \mathbb{B}_{\mathbb{R}^n} \subseteq A \mathbb{B}_{\mathbb{R}^n}\}.$$

Notation: $\mathbb{B}_{\mathbb{R}^d} := \{x \in \mathbb{R}^d : \|x\| \leq 1\}$.

AGENDA:

▷ Conditioning of *conic systems*

$$\begin{aligned} Ax &= b \\ x &\in C, \end{aligned}$$

where $A \in \mathbb{R}^{m \times n}$, and $C \subseteq \mathbb{R}^n$ is a closed convex cone.

- Consider *data structure*, e.g., sparsity or block-structure.

- Conditioning of *generalized equations*

$$b \in F(x),$$

where $F : X \rightrightarrows Y$ is a set-valued mapping.

Well-posedness

Let $C \subseteq \mathbb{R}^n$ be a closed convex cone.

Definition: $A \in \mathbb{R}^{m \times n}$ is *well-posed* if

$$AC = \mathbb{R}^n.$$

Notice

- A well-posed iff

  $$Ax = b, \; x \in C$$

  feasible for all $b \in \mathbb{R}^n$.

- If $m = n$ and $C = \mathbb{R}^n$ then

  $$A \text{ well-posed } \Leftrightarrow A \text{ non-singular.}$$

Write

$A \in \mathcal{W}$ if A well-posed,
$A \in \mathcal{I}$ if A ill-posed.

Generalization of E-Y identity:

Thm 2 (Renegar) *Assume $A \in \mathcal{W}$. Then*

$$\min\{\|B\| : \; A + B \in \mathcal{I}\} = \max\{\delta : \delta\mathbb{B}_{\mathbb{R}^m} \subseteq A(\mathbb{B}_{\mathbb{R}^n} \cap C)\}.$$

Notice:

$$\max\{\delta : \delta\mathbb{B}_{\mathbb{R}^m} \subseteq A(\mathbb{B}_{\mathbb{R}^n} \cap C)\} = \frac{1}{\max\limits_{v \in \mathbb{B}_{\mathbb{R}^m}} \min\{\|x\| : Ax = v, \, x \in C\}}$$

sort of $\frac{1}{\text{``}\|A^{-1}\|\text{''}}$.

Definition (Renegar)

Condition number := reciprocal of

$$\frac{\text{dist}(A, \mathcal{I})}{\|A\|} := \frac{\min\{\|B\| : \ A + B \in \mathcal{I}\}}{\|A\|}.$$

This condition number is useful to study

- sensitivity to data perturbations
- performance of iterative (interior-point) methods
- effects of finite-precision arithmetic

for the *conic optimization problem:*

$$\begin{array}{ll} \min & \langle c, x \rangle \\ & Ax = b \\ & x \in C. \end{array}$$

(Renegar, Filipowski, Vera, Freund, Epelman, Nuñez, Ordoñez, Cucker, Cheung, P.,...)

Limitation: some results are too conservative.

Idea: consider *structured* conditioning.

Suppose $\mathbf{\Delta} \subseteq \mathbb{R}^{m \times n}$ is given, e.g., some sparsity pattern.

Consider

$$\text{dist}_{\mathbf{\Delta}}(A, \mathcal{I}) := \inf\{\|B\|_{\mathbf{\Delta}} : B \in \mathbf{\Delta},\ A + B \in \mathcal{I}\}.$$

Unstructured case: $\mathbf{\Delta} = \mathbb{R}^{m \times n}$.

Special case: when $m = n$ and $C = \mathbb{R}^n$

$\text{dist}_{\mathbf{\Delta}}(A, \mathcal{I})$ = structured dist to singularity
(Demmel, Gohberg, Higham, Rump, Qiu,...)

Also related to μ-number in robust control
(Doyle, Fan, Packard, Tits,...)

Natural question:

Is there a *structured* version of the Eckart-Young identity?

Answer: Yes when $\boldsymbol{\Delta}$ is a block-structure.

AGENDA:

- Conditioning of *conic systems*

$$\begin{aligned} Ax &= b \\ x &\in C, \end{aligned}$$

where $A \in \mathbb{R}^{m \times n}$, and $C \subseteq \mathbb{R}^n$ is a closed convex cone.

▷ Consider *data structure,* e.g., sparsity or block-structure.

- Conditioning of *generalized equations*

$$b \in F(x),$$

where $F : X \rightrightarrows Y$ is a set-valued mapping.

Simple block-structure

Perturb only certain rows and columns of A: Assume $\mathbb{R}^n = \mathbb{R}^k \times \mathbb{R}^{n-k}$, $\mathbb{R}^m = \mathbb{R}^l \times \mathbb{R}^{m-l}$ and

$$\boldsymbol{\Delta} = \left\{ \begin{bmatrix} B & 0 \\ 0 & 0 \end{bmatrix} : B \in \mathbb{R}^{k\times l} \right\}. \tag{1}$$

Thm 3 (P.) *Assume $A \in \mathcal{W}$. Then*

$$\text{dist}_{\boldsymbol{\Delta}}(A, \mathcal{I}) = \sup\{\delta : \delta\mathbb{B}_{\mathbb{R}^l} \subseteq \{Ax : x \in C, \|x_1\| \le 1\}\}.$$

Notice

$$\sup\{\delta : \delta\mathbb{B}_{\mathbb{R}^l} \subseteq \{Ax : x \in C, \|x_1\| \le 1\}\} = \frac{1}{\max\limits_{v\in\mathbb{B}_{\mathbb{R}^l}} \inf\{\|x_1\| : Ax = v,\, x \in C\}},$$

sort of $\frac{1}{``\|A^{-1}\|"}$.

Proof of Thm 3 (sketch).

Alternative (separation):

$$A \in \mathcal{I} \quad \Leftrightarrow \quad \exists y \neq 0 \text{ s.t. } A^{\mathsf{T}} y \in C^*.$$

Norm-duality (following Borwein):

$$\max_{v \in \mathbb{B}_{\mathbb{R}^l}} \inf \{\|x_1\| : Ax = v,\, x \in C\} =$$

$$\sup_{u \in \mathbb{B}_{\mathbb{R}^k},\, u \neq 0} \sup \left\{ \frac{\|y_1\|}{\|u\|} : A^{\mathsf{T}} y + u \in C^* \right\}.$$

RHS: sort of "$\|A^{-\mathsf{T}}\|$".

Rank-one construction:

Come up with $u \in \mathbb{R}^k$, $v \in \mathbb{R}^l$ such that

$$A + v\langle u, \cdot \rangle \in \mathcal{I}.$$

General block-structure

Suppose $X_i \subseteq X$, $Y_i \subseteq Y$, $i = 1, \ldots, k$. Let

$$\mathbf{\Delta} := \left\{ B : B = \sum B_i,\, B_i \in L(X_i, Y_i) \right\},$$

and for $B = \sum B_i \in \mathbf{\Delta}$, let

$$\|B\|_{\mathbf{\Delta}} := \max_i \|B_i\|.$$

Thm 4 (P.) *Assume* $A \in \mathcal{W}$. *Then*

$$\text{dist}_{\mathbf{\Delta}}(A, \mathcal{I}) = \frac{1}{\phi(A)},$$

where

$$\phi(A) =$$

$$\max_{v^j \in \mathbb{B}_{Y_j}} \inf \left\{ \max_i \frac{\|x_i\|}{z_i} : z > 0, Ax = \sum z_j v^j,\, x \in C \right\}.$$

$\phi(A)$: sort of "$\|A^{-1}\|$".

Proof of Thm 4 (sketch).

Alternative: $A \in \mathcal{I} \;\Leftrightarrow\; \exists y \neq 0$ s.t. $A^{\mathsf{T}} y \in C^*$.

Structured norm-duality:

$$\phi(A) = \sup_{u^j \in \mathbb{B}_{X_j}} \sup \left\{ \min_{i, u^i \neq 0} \frac{\|y_i\|}{\|u^i\|} : A^{\mathsf{T}} y + \sum u^j \in C^* \right\}.$$

RHS: sort of "$\|A^{-\mathsf{T}}\|$".

Rank-k construction:

Come up with $u^i \in X_i$, $v^i \in Y_i$ such that

$$A + \sum v^i \langle u^i, \cdot \rangle \in \mathcal{I}.$$

Componentwise distance to ill-posedness

Suppose $E \in \{0,1\}^{m\times n}$ defines some sparsity structure, e.g.,

$$\begin{bmatrix} \mathsf{X} & \mathsf{X} & 0 & 0 & \mathsf{X} \\ 0 & \mathsf{X} & \mathsf{X} & 0 & 0 \\ 0 & 0 & \mathsf{X} & \mathsf{X} & 0 \\ \mathsf{X} & 0 & 0 & \mathsf{X} & 0 \end{bmatrix}.$$

Consider the 1×1 blocks defined by E.
In this case

$$\text{dist}_{\boldsymbol{\Delta}}(A,\mathcal{I}) = \inf\{\delta : \exists B \text{ with } |B| \le \delta\, E \\ \text{s.t. } A + B \in \mathcal{I}\}.$$

Write $\text{dist}_E(A,\mathcal{I}) = \text{dist}_{\boldsymbol{\Delta}}(A,\mathcal{I})$ to emphasize role of E.

Componentwise version of Thm 4:

Given $B \in \mathbb{R}^{m \times n}$, let

$$\begin{array}{rl} \Phi(A,B) := \inf_{x,z} & \displaystyle\max_{j=1\ldots n} \frac{|x_j|}{z_j} \\ \text{s.t.} & Ax = Bz \\ & x \in C \\ & z > 0. \end{array}$$

Thm 5 (P.) *Assume* $A \in \mathcal{W}$ *and* $E \in \{0,1\}^{m \times n}$. *Then*

$$\begin{aligned} \text{dist}_E(A,\mathcal{I}) &= \frac{1}{\max_{|B|=E} \Phi(A,B)} \\ &= \frac{1}{\max_S \Phi(A,SE)}, \end{aligned}$$

max taken over signature *matrices:*
S *is a signature matrix iff* $|S| = I$.

Componentwise distance to singularity

Connection with eigenvalues:

Example. Assume $A \in \mathbb{R}^{n\times n}$ is non-singular and $B \in \mathbb{R}^{n\times n}$. Then

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

$\rho_0(\cdot)$ 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, $\rho_0(M) := 0$.)

Thm 6 (Rohn) *Assume* $A \in \mathbb{R}^{n\times n}$ *is non-singular and* $E \in \{0,1\}^{m\times n}$*. Then*

$$\inf\{\delta : \exists B \text{ with } |B| \le \delta\, E \text{ s.t. } A+B \in \mathrm{Sing}\} = \frac{1}{\max_{S_1,S_2} \rho_0(A^{-1}S_1ES_2)},$$

max taken over signature matrices.

Can recover Thm 6 from Thm 5.

Key step:

Thm 7 (Rump) *Assume $M \in \mathbb{R}^{n\times n}$. Then*

$$\begin{aligned}
\max_S \rho_0(MS) &= \max_{x\neq 0} \min_{x_i \neq 0} \frac{|(Mx)_i|}{|x_i|} \\
&= \max_S \inf_{z>0} \max_i \frac{|(MSz)_i|}{z_i} \\
&= \max_S \Phi(I, MS).
\end{aligned}$$

(Can be shown via LP duality.)

Thus,

$$\begin{aligned}
\max_S \Phi(A, SE) &= \max_{S_1,S_2} \Phi(I, A^{-1}S_1ES_2) \\
&= \max_{S_1,S_2} \rho_0(A^{-1}S_1ES_2).
\end{aligned}$$

Hence Thm 6 follows from Thm 5.

The structured singular value

Assume $n = m$, $\mathbb{R}^n = X_1 \times \cdots \times X_k$, and $C = \mathbb{R}^n$.

Let

$$\mathbf{\Delta} := \left\{ B : B = \sum B_j,\ B_j \in L(X_j, X_j) \right\}.$$

($\mathbf{\Delta}$: a diagonal block-structure.)

Definition (Doyle)

$$\mu_{\mathbf{\Delta}}(M) := \frac{1}{\inf\{\|B\| : B \in \mathbf{\Delta}, \det(I - MB) = 0\}}$$

$\mu_{\mathbf{\Delta}}$: important parameter in robust control (Doyle, Fan, Packard, Tits,...)

Notice: For M non-singular

$$
\begin{aligned}
\frac{1}{\mu_{\mathbf{\Delta}}(M)} &= \inf\{\|B\|_{\mathbf{\Delta}} : B \in \mathbf{\Delta}, M^{-1} - B \in \mathsf{Sing}\} \\
&= \mathsf{dist}_{\mathbf{\Delta}}(M^{-1}, \mathcal{I}).
\end{aligned}
$$

Hence can characterize $\mu_{\mathbf{\Delta}}$ by using Thm 4.

AGENDA:

- Conditioning of *conic systems*

$$\begin{array}{l} Ax = b \\ x \in C, \end{array}$$

  where $A \in \mathbb{R}^{m \times n}$, and $C \subseteq \mathbb{R}^n$ is a closed convex cone.

- Consider *data structure*, e.g., sparsity or block-structure.

▷ Conditioning of *generalized equations*

$$b \in F(x),$$

where $F : X \rightrightarrows Y$ is a set-valued mapping.

Generalized equations

Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued mapping: $F(x) \subseteq Y$ for $x \in X$.

Define graph$(F) \subseteq \mathbb{R}^n \times \mathbb{R}^m$, $F^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ as follows

$$\text{graph}(F) := \{(x, y) : y \in F(x)\},$$

$$x \in F^{-1}(y) \Leftrightarrow y \in F(x).$$

Generalized equation:

Given $b \in Y$, find $x \in X$ s.t. $b \in F(x)$.

Examples of generalized equations

- Usual equations:
$$f(x) = b$$
for $f : \mathbb{R}^n \to \mathbb{R}^m$.

- Conic systems:
$$Ax = b,\ x \in C.$$

- Complementarity problems:
$$f(x) \geq b,\ x \geq 0,\ x^{\mathsf{T}}(f(x) - b) = 0.$$

- Variational inequalities:
$$b \in f(x) + N_D(x),$$
where $D \subseteq \mathbb{R}^n$ is a closed convex set.

- Optimality conditions (e.g., KKT).

Metric regularity

Assume $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $\bar{y} \in F(\bar{x})$.

Definition. F is *metrically regular* at $\bar{x}$ for $\bar{y}$ if there exists $\kappa > 0$ such that

$$d(x, F^{-1}(y)) \leq \kappa\, d(y, F(x)) \qquad (2)$$

for all (x, y) in a neighborhood of $(\bar{x}, \bar{y})$. Let

$$\text{reg } F(\bar{x} \,|\, \bar{y}) := \inf\{\kappa : (2) \text{ holds}\}.$$

Convention: reg $F(\bar{x} \,|\, \bar{y}) := \infty$ if F is not metrically regular at $\bar{x}$ for $\bar{y}$.

Thm 8 (Dontchev, Lewis, Rockafellar)
Assume $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $\bar{y} \in F(\bar{x})$ *and graph*(F) *is locally closed at* $(\bar{x}, \bar{y})$*. Then*

$$\inf\{\|B\| : \mathrm{reg}(F+B)(\bar{x}\,|\,\bar{y}+B\bar{x}) = \infty\} = \frac{1}{\mathrm{reg}\,F(\bar{x}\,|\,\bar{y})}.$$

Connections with fundamental results in non-smooth analysis:

Lusternik-Graves, Robinson-Ursescu.

(Extensions of open mapping and closed graph thms.)

Sublinear mappings

Definition. $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *sublinear* if graph(F) is a convex cone.

Conic systems: special case of sublinear mappings.

For F sublinear define

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

Fact: If F is sublinear, then so is F^{-1} and

$$\operatorname{reg} F(0|0) = \|F^{-1}\|.$$

Thm 9 (Lewis) *Assume F is a sublinear mapping with closed graph. Then*

$$\text{inf}\{\|G\| : G \in L(X,Y),\ F{+}G \text{ not surj}\} = \frac{1}{\|F^{-1}\|}.$$

Block-structured version:

Thm 10 (Lewis, also P.) *Assume F is a sublinear mapping with closed graph. Then*

$$\text{inf}\{\|B\|_{\mathbf{\Delta}} : B \in \mathbf{\Delta},\ F + B \text{ not surj}\} = \frac{1}{\|F^{-1}\|^{-}_{\mathbf{\Delta}}}.$$

where

$$\|F^{-1}\|^{-}_{\mathbf{\Delta}} = \sup_{v^j \in \mathbb{B}_{Y_j}} \inf \left\{ \max_i \frac{\|x_i\|}{z_i} : z > 0,\ x \in F^{-1}\Big(\sum z_j v^j\Big) \right\}.$$

Proof of Thm 10 (sketch).

Adjoint: Define $F^* : Y^* \rightrightarrows X^*$ by

$$u \in F^*(v) \Leftrightarrow \langle v, y \rangle \geq \langle u, x \rangle \ \forall y \in F(x).$$

Alternative (separation):
F not surjective $\Leftrightarrow$ $\exists v \neq 0$ s.t. $0 \in F^*(v)$.

Structured norm-duality (following Borwein):

Thm 11 (P.) *Assume F a sublinear mapping with closed graph. Then*

$$\|F^{-1}\|^-_{\boldsymbol{\Delta}} = \|F^{-*}\|^+_{\boldsymbol{\Delta}},$$

where

$$\|F^{-*}\|^+_{\boldsymbol{\Delta}} = \sup_{u^j \in \mathbb{B}_{X_j^*}} \sup \left\{ \min_{i, u^i \neq 0} \frac{\|v_i\|}{\|u^i\|} : z > 0,\ v \in F^{-*}\Big(\sum z_j u^j\Big) \right\}.$$

Rank-k construction.

For X, Y Banach spaces and $F : X \rightrightarrows Y$.

Good news:

Structured norm-duality holds.

Bad news:

Alternative step fails:

F not surj $\not\Rightarrow \exists v \neq 0$ s.t. $0 \in F^*(v)$.

Thm 10 (structured case) does not hold.

Good news:

Dual counterpart holds.

Definition:

$$F^* \text{ is } \mathbf{\Delta}\text{-singular iff } \|F^{-*}\|_{\mathbf{\Delta}}^{+} = \infty.$$

For X, Y Banach spaces:

Thm 12 (P.) *Assume* $F^* : Y^* \rightrightarrows X^*$ *is not* $\mathbf{\Delta}$*-singular. Then*

$$\inf\{\|B^*\|_{\mathbf{\Delta}} : F^* + B^* \text{ is } \mathbf{\Delta}\text{-sing}\} = \frac{1}{\|F^{-*}\|_{\mathbf{\Delta}}^{+}}.$$

Good news:

Under suitable circumstances

$$F \text{ is not surj } \Leftrightarrow F^* \text{ is } \mathbf{\Delta}\text{-singular}.$$

Thus, amended version of Thm 10 holds for Banach spaces X, Y.

In particular, Thm 9 (unstructured case) holds for Banach spaces X, Y (Lewis).

Conclusions

- Eckart-Young identity extends to:
  - ∗ conic systems and set-valued mappings
  - ∗ block-structured perturbations.

- Three key steps:
  - ∗ alternative

    $$A \in \mathcal{I} \Leftrightarrow \exists y \neq 0 \text{ s.t. } A^* y \in C^*$$

  - ∗ norm-duality

    $$\text{“}\|A^{-1}\|\text{”} = \text{“}\|A^{-*}\|\text{”}$$

  - ∗ low-rank construction.

- Current work:
  - ∗ other types of structure, e.g., symmetry (cf. Rump)
  - ∗ infinite-dimensional spaces
  - ∗ structured metric regularity

- Acknowledgements: Raphael Hauser, Oxford University.