ILLUSTRATIONS OF METRIC REGULARITY

Adrian Lewis

Cornell University

April 2, 2008

Beijing, 2008

Joint work with: J. Bolte, J. Burke
A. Daniilidis, A. Dontchev, R. Luke
J. Malick, M. Overton, J. Pang
R.T. Rockafellar, M. Shiota

1. OUTLINE

1. OUTLINE

- Demmel's paradigm and metric regularity:
ill-conditioning $\leftrightarrow$ distance to ill-posedness $\leftrightarrow$ algorithms.

1. OUTLINE

- Demmel's paradigm and metric regularity:
  ill-conditioning $\leftrightarrow$ distance to ill-posedness $\leftrightarrow$ algorithms.
- Von Neumann's method of alternating projections:
  nonconvex versions in image-processing and control design.

1. OUTLINE

- Demmel's paradigm and metric regularity:
ill-conditioning $\leftrightarrow$ distance to ill-posedness $\leftrightarrow$ algorithms.
- Von Neumann's method of alternating projections:
nonconvex versions in image-processing and control design.
- Linear convergence of alternating nonconvex projections,
and Mordukhovich's "extremal principle".

1. OUTLINE

- Demmel's paradigm and metric regularity:
  ill-conditioning $\leftrightarrow$ distance to ill-posedness $\leftrightarrow$ algorithms.
- Von Neumann's method of alternating projections:
  nonconvex versions in image-processing and control design.
- Linear convergence of alternating nonconvex projections,
  and Mordukhovich's "extremal principle".
- Pseudospectra and robust modelling of dynamics.

1. OUTLINE

- Demmel's paradigm and metric regularity: ill-conditioning $\leftrightarrow$ distance to ill-posedness $\leftrightarrow$ algorithms.
- Von Neumann's method of alternating projections: nonconvex versions in image-processing and control design.
- Linear convergence of alternating nonconvex projections, and Mordukhovich's "extremal principle".
- Pseudospectra and robust modelling of dynamics.
- Lipschitz behavior of pseudospectra, resolvent-critical points, and Ioffe's semi-algebraic Sard theorem.

2. PROJECTION

2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

$$d_S(y) \;=\; \min_{x \in S} \|x - y\|.$$

2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

$$\begin{aligned} d_S(y) &= \min_{x \in S} \|x - y\|. \\ P_S(y) &= \{\text{nearest points}\}. \end{aligned}$$

2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

$$\begin{aligned} d_S(y) &= \min_{x \in S} \|x - y\|. \\ P_S(y) &= \{\text{nearest points}\}. \end{aligned}$$

S is **prox-regular** if P_S single-valued locally.

2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

$$\begin{aligned} d_S(y) &= \min_{x \in S} \|x - y\|. \\ P_S(y) &= \{\text{nearest points}\}. \end{aligned}$$

S is **prox-regular** if P_S single-valued locally. Eg: convex sets, smooth manifolds. . . .

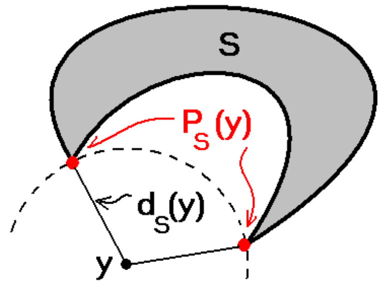
2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

$$d_S(y) = \min_{x \in S} \|x - y\|.$$

$$P_S(y) = \{\text{nearest points}\}.$$

S is **prox-regular** if P_S single-valued locally. Eg: convex sets, smooth manifolds....

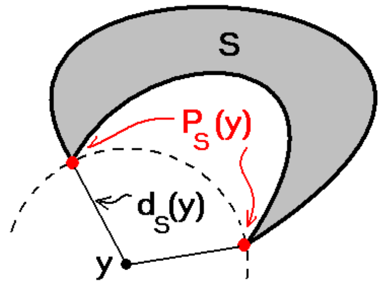


2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

$$d_S(y) = \min_{x \in S} \|x - y\|.$$

$$P_S(y) = \{\text{nearest points}\}.$$



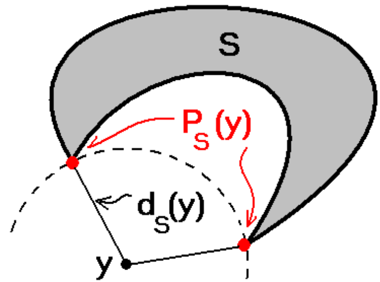
S is **prox-regular** if P_S single-valued locally. Eg: convex sets, smooth manifolds. . . .

For closed S, T, to find a point in $S \cap T$,

2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

$$
\begin{aligned}
d_S(y) &= \min_{x \in S} \|x - y\|. \\
P_S(y) &= \{\text{nearest points}\}.
\end{aligned}
$$



S is **prox-regular** if P_S single-valued locally. Eg: convex sets, smooth manifolds. . . .

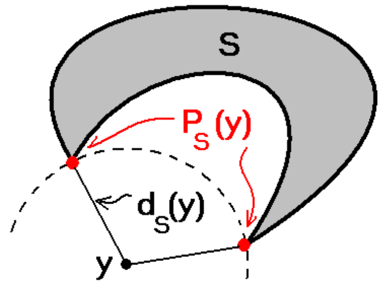
For closed S, T, to find a point in $S \cap T$,
repeat:

$$x \leftarrow P_S(x)$$

2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

$$d_S(y) \;=\; \min_{x \in S} \|x - y\|.$$
$$P_S(y) \;=\; \{\text{nearest points}\}.$$



S is **prox-regular** if P_S single-valued locally. Eg: convex sets, smooth manifolds....

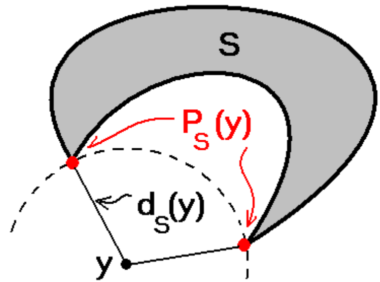
For closed S, T, to find a point in $S \cap T$,
repeat:

$$x \;\leftarrow\; P_S(x)$$
$$x \;\leftarrow\; P_T(x).$$

2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

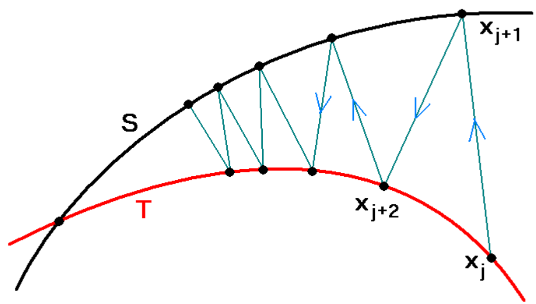
$$d_S(y) = \min_{x \in S} \|x - y\|.$$
$$P_S(y) = \{\text{nearest points}\}.$$



S is **prox-regular** if P_S single-valued locally. Eg: convex sets, smooth manifolds. . . .

For closed S, T, to find a point in $S \cap T$,
repeat:

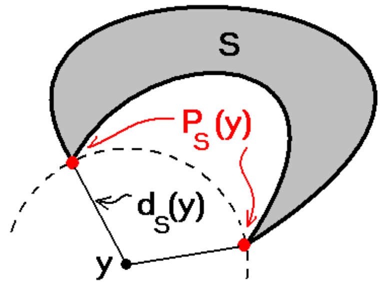
$$x \leftarrow P_S(x)$$
$$x \leftarrow P_T(x).$$



2. PROJECTION

Distance from y to $S \subset \mathbb{R}^n$ is

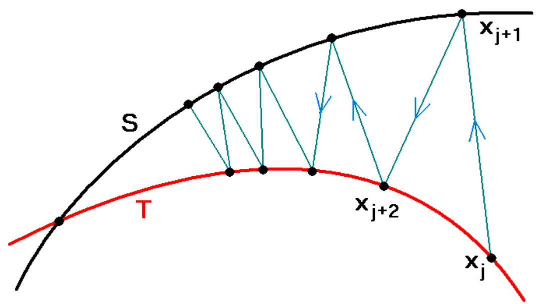
$$d_S(y) \;=\; \min_{x \in S} \|x - y\|.$$
$$P_S(y) \;=\; \{\text{nearest points}\}.$$



S is **prox-regular** if P_S single-valued locally. Eg: convex sets, smooth manifolds. . . .

For closed S, T, to find a point in $S \cap T$,
repeat:

$$x \;\leftarrow\; P_S(x)$$
$$x \;\leftarrow\; P_T(x).$$



$d_{S \cap T}(x) \to 0$ if S and T **convex** (von Neumann '33) . . . , linearly if $\mathrm{ri}\, S \cap \mathrm{ri}\, T \neq \emptyset$.

3. NONCONVEX ALTERNATING PROJECTIONS

3. NONCONVEX ALTERNATING PROJECTIONS

Projection may be numerically easy even for nonconvex sets, so alternating projections also tempts as a **nonconvex** heuristic.

3. NONCONVEX ALTERNATING PROJECTIONS

Projection may be numerically easy even for nonconvex sets, so alternating projections also tempts as a **nonconvex** heuristic.

Eg (Grigoriadis-Skelton '96)... **Low-order control**:

$$\Big\{\boldsymbol{X} \text{ positive semidefinite} : \text{rank}\, \boldsymbol{X} \leq \boldsymbol{r}\Big\} \cap \Big\{\boldsymbol{X} : \boldsymbol{\mathcal{A}}(\boldsymbol{X}) = \boldsymbol{b}\Big\}.$$

3. NONCONVEX ALTERNATING PROJECTIONS

Projection may be numerically easy even for nonconvex sets, so alternating projections also tempts as a **nonconvex** heuristic.

Eg (Grigoriadis-Skelton '96)... **Low-order control**:

$$\Big\{\boldsymbol{X} \text{ positive semidefinite} : \text{rank}\, \boldsymbol{X} \leq \boldsymbol{r}\Big\} \cap \Big\{\boldsymbol{X} : \boldsymbol{\mathcal{A}}(\boldsymbol{X}) = \boldsymbol{b}\Big\}.$$

Eg (Bauschke-Combettes-Luke '02)... **Phase retrieval**:
For linear $\boldsymbol{A} \colon \mathbb{C}^{n} \longrightarrow \mathbb{C}^{m}$,

$$\Big\{(\boldsymbol{x}, \boldsymbol{z}) : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{z}\Big\} \;\cap\; \Big\{(\boldsymbol{x}, \boldsymbol{z}) : |\boldsymbol{z}_j| = \boldsymbol{b}_j \;\forall \boldsymbol{j}\Big\}.$$

. . . Numerical example . . .

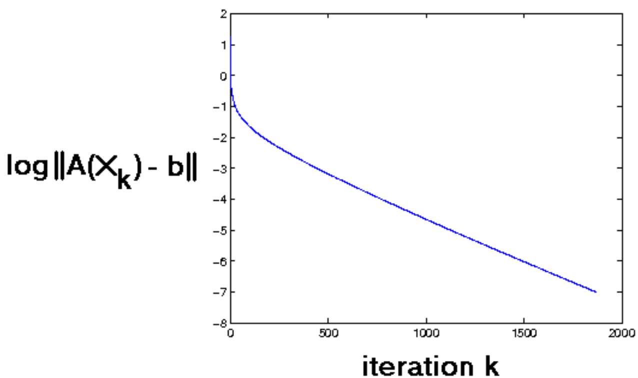
. . . Numerical example . . .

We seek a 100-by-110 matrix $\boldsymbol{X}$ of **rank 4**, satisfying 450 linear equations:

$$\boldsymbol{A}(\boldsymbol{X}) = \boldsymbol{b}.$$

We solve by alternately projecting onto

- $\{\boldsymbol{X} : \boldsymbol{A}(\boldsymbol{X}) = \boldsymbol{b}\}$ (by solving the normal equations);
- $\{\boldsymbol{X} : \text{rank}\, \boldsymbol{X} = 4\}$ (via the singular value decomposition).



Why does alternating nonconvex projections work?

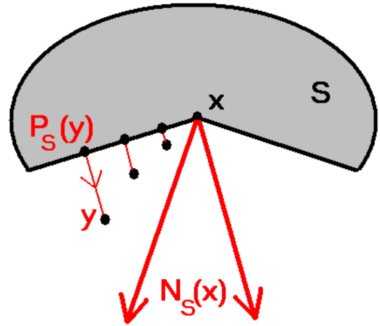
4. KEY IDEA: THE NORMAL CONE

4. KEY IDEA: THE NORMAL CONE

$\boldsymbol{N_S(x)}$ at $\boldsymbol{x} \in \boldsymbol{S}$ consists of limits of directions to points $\boldsymbol{y}$ near $\boldsymbol{x}$ from $\boldsymbol{P_S(y)}$.

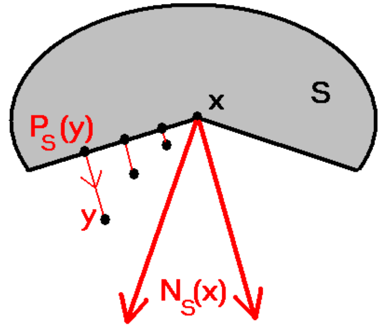
4. KEY IDEA: THE NORMAL CONE

$N_S(x)$ at $x \in S$ consists of limits of directions to points y near x from $P_S(y)$.



4. KEY IDEA: THE NORMAL CONE

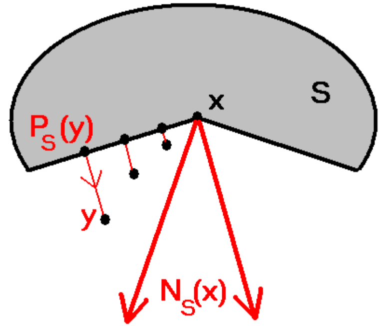
$N_S(x)$ at $x \in S$ consists of limits of directions to points y near x from $P_S(y)$.



Extremal principle (Mordukhovich)

4. KEY IDEA: THE NORMAL CONE

$\boldsymbol{N_S(x)}$ at $\boldsymbol{x} \in \boldsymbol{S}$ consists of limits of directions to points $\boldsymbol{y}$ near $\boldsymbol{x}$ from $\boldsymbol{P_S(y)}$.



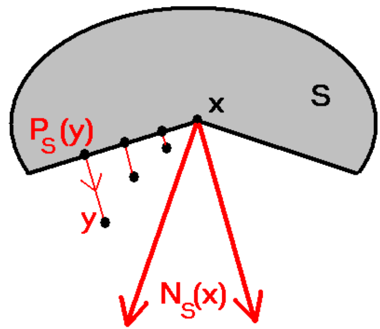
Extremal principle (Mordukhovich)

$$(*) \qquad \boldsymbol{N_S(x) \cap -N_T(x) \; = \; \{0\}}$$

$\Rightarrow$ small translations preserve a local intersection.

4. KEY IDEA: THE NORMAL CONE

$N_S(x)$ at $x \in S$ consists of limits of directions to points y near x from $P_S(y)$.



Extremal principle (Mordukhovich)

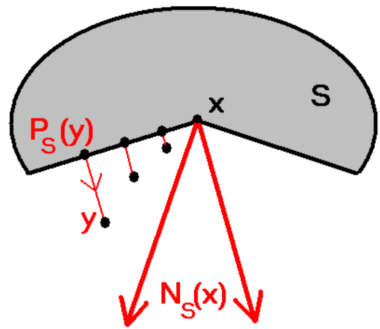
$$(*) \qquad N_S(x) \cap -N_T(x) \; = \; \{0\}$$

$\Rightarrow$ small translations preserve a local intersection.

Algorithmic proof via. . .

4. KEY IDEA: THE NORMAL CONE

$N_S(x)$ at $x \in S$ consists of limits of directions to points y near x from $P_S(y)$.



Extremal principle (Mordukhovich)

$$(*) \qquad N_S(x) \cap -N_T(x) \;=\; \{0\}$$

$\Rightarrow$ small translations preserve a local intersection.

Algorithmic proof via...

Theorem (L/Luke/Malick '07) If S or T is prox-regular, $(*) \;\Rightarrow\;$ alternating projections converges linearly near x.

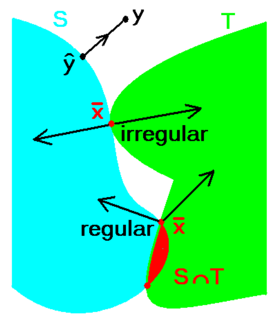
. . . Sketch proof . . .

... Sketch proof ...

Closed S, T have **regular** intersection at $\bar{x}$:

$$N_S(\bar{x}) \cap -N_T(\bar{x}) = \{0\}.$$

So for some $\theta > 0$, y, z near $\bar{x}$, $\hat{y} \in P_S(y)$, $\hat{z} \in P_T(z)$ implies angle $\geq \theta$ between normal vector $y - \hat{y}$ and $\hat{z} - z$.

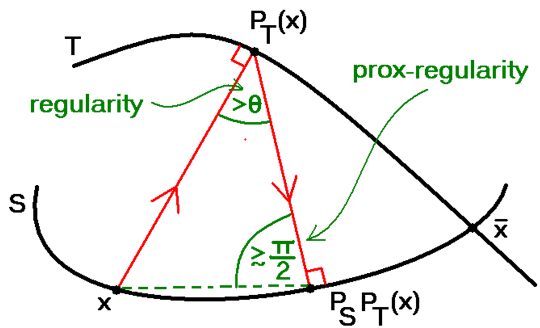


Convergence proof

For $x \in S$ near $\bar{x}$,

$$\frac{\|P_S P_T(x) - P_T(x)\|}{\|P_T(x) - x\|}$$

isn't much larger than $\cos \theta$.



5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

Find $\boldsymbol{x} \in \boldsymbol{S} \cap \boldsymbol{T}$. Equivalently, solve $\boldsymbol{d_S}(\boldsymbol{x}) + \boldsymbol{d_T}(\boldsymbol{x}) = \boldsymbol{0}$.
Three equivalent properties of instances:

5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

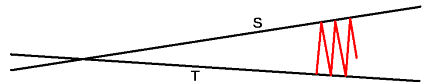
Find $\boldsymbol{x} \in \boldsymbol{S} \cap \boldsymbol{T}$. Equivalently, solve $\boldsymbol{d_S}(\boldsymbol{x}) + \boldsymbol{d_T}(\boldsymbol{x}) = \mathbf{0}$.
Three equivalent properties of instances:

- "Basic" algorithms converge slowly.

5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

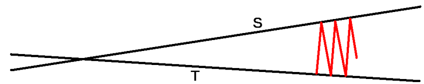
Find $\boldsymbol{x} \in \boldsymbol{S} \cap \boldsymbol{T}$. Equivalently, solve $\boldsymbol{d_S}(\boldsymbol{x}) + \boldsymbol{d_T}(\boldsymbol{x}) = \boldsymbol{0}$.
Three equivalent properties of instances:

- "Basic" algorithms converge slowly.



5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

Find $\boldsymbol{x} \in \boldsymbol{S} \cap \boldsymbol{T}$. Equivalently, solve $\boldsymbol{d_S}(\boldsymbol{x}) + \boldsymbol{d_T}(\boldsymbol{x}) = \boldsymbol{0}$.
Three equivalent properties of instances:



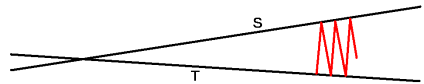
- "Basic" algorithms converge slowly.

- Small perturbations cause "ill-posedness".

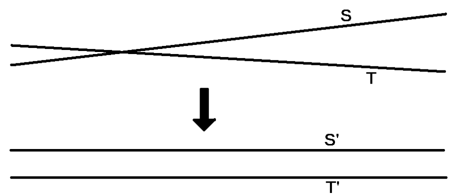
5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

Find $\boldsymbol{x} \in \boldsymbol{S} \cap \boldsymbol{T}$. Equivalently, solve $\boldsymbol{d_S}(\boldsymbol{x}) + \boldsymbol{d_T}(\boldsymbol{x}) = \mathbf{0}$.
Three equivalent properties of instances:

- "Basic" algorithms converge slowly.



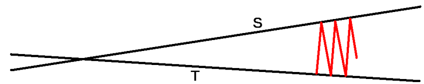
- Small perturbations cause "ill-posedness".



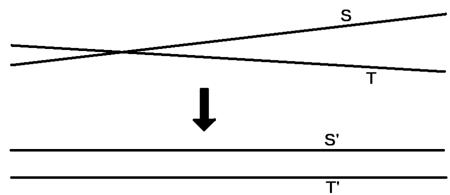
5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

Find $x \in S \cap T$. Equivalently, solve $d_S(x) + d_T(x) = 0$.
Three equivalent properties of instances:

- "Basic" algorithms converge slowly.



- Small perturbations cause "ill-posedness".

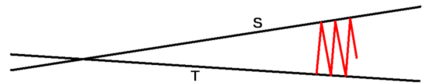


- Weak "error bounds":

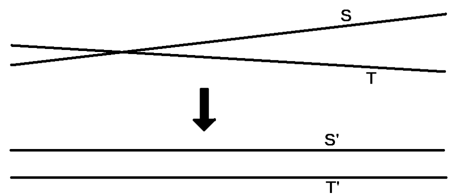
5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

Find $x \in S \cap T$. Equivalently, solve $d_S(x) + d_T(x) = 0$.
Three equivalent properties of instances:

- "Basic" algorithms converge slowly.



- Small perturbations cause "ill-posedness".

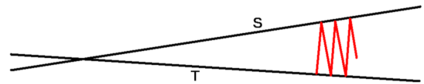


- Weak "error bounds": $d_{S \cap T} \leq k(d_S + d_T)$ needs k large.

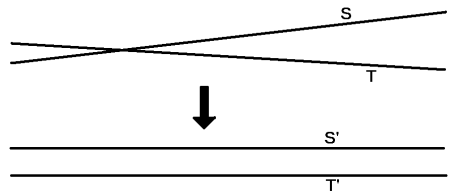
5. DEMMEL'S IDEA (1987): ILL-CONDITIONED $\equiv$ NEARLY ILL-POSED $\equiv$ SLOW CONVERGENCE

Find $x \in S \cap T$. Equivalently, solve $d_S(x) + d_T(x) = 0$.
Three equivalent properties of instances:

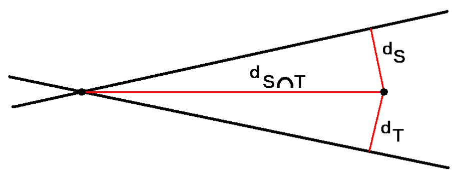
- "Basic" algorithms converge slowly.



- Small perturbations cause "ill-posedness".



- Weak "error bounds": $d_{S \cap T} \leq k(d_S + d_T)$ needs k large.



6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$\boldsymbol{Ax = b.}$$

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$\boldsymbol{Ax = b.}$$

- Small distance to singularity $\boldsymbol{\lambda}_{\mathsf{min}}(\boldsymbol{A})$.

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$\boldsymbol{Ax = b.}$$

- Small distance to singularity $\boldsymbol{\lambda}_{\mathsf{min}}(\boldsymbol{A})$.
- Weak error bound:

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$\boldsymbol{Ax = b}.$$

- Small distance to singularity $\boldsymbol{\lambda}_{\mathsf{min}}(\boldsymbol{A})$.
- Weak error bound:

$$\|\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\| \;\leq\; \frac{\mathbf{1}}{\boldsymbol{\lambda}_{\mathsf{min}}}\|\boldsymbol{Ax} - \boldsymbol{b}\|.$$

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$\boldsymbol{Ax = b.}$$

- Small distance to singularity $\boldsymbol{\lambda}_{\min}(\boldsymbol{A})$.
- Weak error bound:

$$\|\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\| \;\leq\; \frac{\boldsymbol{1}}{\boldsymbol{\lambda}_{\min}}\|\boldsymbol{Ax} - \boldsymbol{b}\|.$$

- Slow linear rate of basic algorithms.

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$\boldsymbol{Ax = b.}$$

- Small distance to singularity $\boldsymbol{\lambda}_{\min}(\boldsymbol{A})$.
- Weak error bound:

$$\|\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\| \;\leq\; \frac{\mathbf{1}}{\boldsymbol{\lambda}_{\min}}\|\boldsymbol{Ax} - \boldsymbol{b}\|.$$

- Slow linear rate of basic algorithms. Eg:

$$\left(\frac{\boldsymbol{\lambda}_{\max} - \boldsymbol{\lambda}_{\min}}{\boldsymbol{\lambda}_{\max} + \boldsymbol{\lambda}_{\min}}\right)^{\mathbf{2}}$$

  for steepest descent.

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$\boldsymbol{Ax = b.}$$

- Small distance to singularity $\boldsymbol{\lambda}_{\mathsf{min}}(\boldsymbol{A})$.
- Weak error bound:

$$\|\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\| \;\leq\; \frac{\mathbf{1}}{\boldsymbol{\lambda}_{\mathsf{min}}}\|\boldsymbol{Ax} - \boldsymbol{b}\|.$$

- Slow linear rate of basic algorithms. Eg:

$$\left(\frac{\boldsymbol{\lambda}_{\mathsf{max}} - \boldsymbol{\lambda}_{\mathsf{min}}}{\boldsymbol{\lambda}_{\mathsf{max}} + \boldsymbol{\lambda}_{\mathsf{min}}}\right)^{\mathbf{2}}$$

for steepest descent. Conjugate gradients analogous.

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$Ax = b.$$

- Small distance to singularity $\lambda_{\text{min}}(A)$.
- Weak error bound:

$$\|x - A^{-1}b\| \;\leq\; \frac{1}{\lambda_{\text{min}}}\|Ax - b\|.$$

- Slow linear rate of basic algorithms. Eg:

$$\left(\frac{\lambda_{\text{max}} - \lambda_{\text{min}}}{\lambda_{\text{max}} + \lambda_{\text{min}}}\right)^2$$

for steepest descent. Conjugate gradients analogous.

Extends to linear programming via interior points (Renegar '95).

6. EXAMPLE: POSITIVE-DEFINITE SYSTEMS

$$\boldsymbol{Ax = b.}$$

- Small distance to singularity $\boldsymbol{\lambda}_{\mathsf{min}}(\boldsymbol{A})$.
- Weak error bound:

$$\|\boldsymbol{x} - \boldsymbol{A}^{-1}\boldsymbol{b}\| \;\leq\; \frac{\mathbf{1}}{\boldsymbol{\lambda}_{\mathsf{min}}}\|\boldsymbol{Ax} - \boldsymbol{b}\|.$$

- Slow linear rate of basic algorithms. Eg:

$$\left(\frac{\boldsymbol{\lambda}_{\mathsf{max}} - \boldsymbol{\lambda}_{\mathsf{min}}}{\boldsymbol{\lambda}_{\mathsf{max}} + \boldsymbol{\lambda}_{\mathsf{min}}}\right)^{\mathbf{2}}$$

  for steepest descent. Conjugate gradients analogous.

Extends to linear programming via interior points (Renegar '95).

How general is this pattern?

7. INVERTING SET-VALUED MAPPINGS

7. INVERTING SET-VALUED MAPPINGS

Given **set-valued** $\boldsymbol{F} : \mathbb{R}^{\boldsymbol{p}} \rightrightarrows \mathbb{R}^{\boldsymbol{q}}$, suppose

- $\boldsymbol{F}(\boldsymbol{x})$ is easy to compute;

7. INVERTING SET-VALUED MAPPINGS

Given **set-valued** $\boldsymbol{F} : \mathbb{R}^{\boldsymbol{p}} \rightrightarrows \mathbb{R}^{\boldsymbol{q}}$, suppose

- $\boldsymbol{F}(\boldsymbol{x})$ is easy to compute;
- $\boldsymbol{F}^{-1}(\boldsymbol{y})$ is hard to compute.

7. INVERTING SET-VALUED MAPPINGS

Given **set-valued** $\boldsymbol{F} : \mathbb{R}^{\boldsymbol{p}} \rightrightarrows \mathbb{R}^{\boldsymbol{q}}$, suppose

- $\boldsymbol{F}(\boldsymbol{x})$ is easy to compute;
- $\boldsymbol{F}^{-1}(\boldsymbol{y})$ is hard to compute.

Problem Given $\bar{\boldsymbol{y}}$, find $\boldsymbol{x}$ so $\bar{\boldsymbol{y}} \in \boldsymbol{F}(\boldsymbol{x})$.

7. INVERTING SET-VALUED MAPPINGS

Given **set-valued** $\boldsymbol{F} : \mathbb{R}^{\boldsymbol{p}} \rightrightarrows \mathbb{R}^{\boldsymbol{q}}$, suppose

- $\boldsymbol{F}(\boldsymbol{x})$ is easy to compute;
- $\boldsymbol{F}^{-1}(\boldsymbol{y})$ is hard to compute.

Problem Given $\bar{\boldsymbol{y}}$, find $\boldsymbol{x}$ so $\bar{\boldsymbol{y}} \in \boldsymbol{F}(\boldsymbol{x})$.

Example (Set intersection) Find $\boldsymbol{x} \in \boldsymbol{S} \cap \boldsymbol{T}$.

7. INVERTING SET-VALUED MAPPINGS

Given **set-valued** $\boldsymbol{F} : \mathbb{R}^{\boldsymbol{p}} \rightrightarrows \mathbb{R}^{\boldsymbol{q}}$, suppose

- $\boldsymbol{F}(\boldsymbol{x})$ is easy to compute;
- $\boldsymbol{F}^{-1}(\boldsymbol{y})$ is hard to compute.

Problem Given $\bar{\boldsymbol{y}}$, find $\boldsymbol{x}$ so $\bar{\boldsymbol{y}} \in \boldsymbol{F}(\boldsymbol{x})$.

Example (Set intersection) Find $\boldsymbol{x} \in \boldsymbol{S} \cap \boldsymbol{T}$.
Equivalently, if

$$\boldsymbol{G}(\boldsymbol{x}) = (\boldsymbol{S} - \boldsymbol{x}) \times (\boldsymbol{T} - \boldsymbol{x}),$$

7. INVERTING SET-VALUED MAPPINGS

Given **set-valued** $\boldsymbol{F} : \mathbb{R}^{\boldsymbol{p}} \rightrightarrows \mathbb{R}^{\boldsymbol{q}}$, suppose

- $\boldsymbol{F}(\boldsymbol{x})$ is easy to compute;
- $\boldsymbol{F}^{-1}(\boldsymbol{y})$ is hard to compute.

Problem Given $\bar{\boldsymbol{y}}$, find $\boldsymbol{x}$ so $\bar{\boldsymbol{y}} \in \boldsymbol{F}(\boldsymbol{x})$.

Example (Set intersection) Find $\boldsymbol{x} \in \boldsymbol{S} \cap \boldsymbol{T}$.
Equivalently, if

$$\boldsymbol{G}(\boldsymbol{x}) = (\boldsymbol{S} - \boldsymbol{x}) \times (\boldsymbol{T} - \boldsymbol{x}),$$

solve

$$(\mathbf{0}, \mathbf{0}) \in \boldsymbol{G}(\boldsymbol{x}).$$

8. LOCAL ERROR BOUNDS

8. LOCAL ERROR BOUNDS

$\boldsymbol{F}$ is **metrically regular** for $\bar{\boldsymbol{y}} \in \boldsymbol{F}(\bar{\boldsymbol{x}})$

8. LOCAL ERROR BOUNDS

F is **metrically regular** for $\bar{y} \in F(\bar{x})$ if, for some ρ,

$$d_{F^{-1}(y)}(x) \leq \rho d_{F(x)}(y)$$

8. LOCAL ERROR BOUNDS

F is **metrically regular** for $\bar{y} \in F(\bar{x})$ if, for some ρ,

$$d_{F^{-1}(y)}(x) \leq \rho d_{F(x)}(y) \quad \text{for all } (x, y) \text{ near } (\bar{x}, \bar{y}).$$

8. LOCAL ERROR BOUNDS

F is **metrically regular** for $\bar{y} \in F(\bar{x})$ if, for some ρ,

$$d_{F^{-1}(y)}(x) \leq \rho d_{F(x)}(y) \quad \text{for all } (x, y) \text{ near } (\bar{x}, \bar{y}).$$

If not, $\bar{y}$ is **critical**.

8. LOCAL ERROR BOUNDS

$\boldsymbol{F}$ is **metrically regular** for $\bar{y} \in F(\bar{x})$ if, for some ρ,

$$d_{F^{-1}(y)}(x) \leq \rho d_{F(x)}(y) \quad \text{for all } (x, y) \text{ near } (\bar{x}, \bar{y}).$$

If not, $\bar{y}$ is **critical**. Smallest such ρ is **modulus**.

8. LOCAL ERROR BOUNDS

F is **metrically regular** for $\bar{y} \in F(\bar{x})$ if, for some ρ,

$$d_{F^{-1}(y)}(x) \leq \rho d_{F(x)}(y) \text{ for all } (x, y) \text{ near } (\bar{x}, \bar{y}).$$

If not, $\bar{y}$ is **critical**. Smallest such ρ is **modulus**. **Examples**:

- Smooth F is metrically regular wherever ∇F onto:

8. LOCAL ERROR BOUNDS

F is **metrically regular** for $\bar{y} \in F(\bar{x})$ if, for some ρ,

$$d_{F^{-1}(y)}(x) \leq \rho d_{F(x)}(y) \text{ for all } (x, y) \text{ near } (\bar{x}, \bar{y}).$$

If not, $\bar{y}$ is **critical**. Smallest such ρ is **modulus**. **Examples**:

- Smooth F is metrically regular wherever ∇F onto:

$$\text{modulus} = \frac{1}{\sigma_{\min}(\nabla F(\bar{x}))}.$$

8. LOCAL ERROR BOUNDS

F is **metrically regular** for $\bar{y} \in F(\bar{x})$ if, for some ρ,

$$d_{F^{-1}(y)}(x) \leq \rho d_{F(x)}(y) \text{ for all } (x, y) \text{ near } (\bar{x}, \bar{y}).$$

If not, $\bar{y}$ is **critical**. Smallest such ρ is **modulus**. **Examples**:

- Smooth F is metrically regular wherever ∇F onto:

$$\text{modulus} = \frac{1}{\sigma_{\min}(\nabla F(\bar{x}))}.$$

- For "set intersection", $G(x) = (S - x) \times (T - x)$,

$$\text{modulus} = \frac{1}{\sqrt{1 - \cos\theta}},$$

8. LOCAL ERROR BOUNDS

F is **metrically regular** for $\bar{y} \in F(\bar{x})$ if, for some ρ,

$$d_{F^{-1}(y)}(x) \leq \rho d_{F(x)}(y) \text{ for all } (x, y) \text{ near } (\bar{x}, \bar{y}).$$

If not, $\bar{y}$ is **critical**. Smallest such ρ is **modulus**. **Examples**:

- Smooth F is metrically regular wherever ∇F onto:

$$\text{modulus} = \frac{1}{\sigma_{\min}(\nabla F(\bar{x}))}.$$

- For "set intersection", $G(x) = (S - x) \times (T - x)$,

$$\text{modulus} = \frac{1}{\sqrt{1 - \cos\theta}},$$

where θ is minimal angle between $N_S(\bar{x})$ and $-N_T(\bar{x})$.

9. CONDITIONING AND SPEED

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{\boldsymbol{y}} \in \boldsymbol{F}(\bar{\boldsymbol{x}})$,

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{y} \in F(\bar{x})$, is

$$\inf_{\text{linear } A} \Big\{ \|A\| : \text{irregularity for } \bar{y} + A\bar{x} \in (F + A)(\bar{x}) \Big\}.$$

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{\boldsymbol{y}} \in \boldsymbol{F}(\bar{\boldsymbol{x}})$, is

$$\inf_{\text{linear } \boldsymbol{A}} \Big\{ \|\boldsymbol{A}\| : \text{irregularity for } \bar{\boldsymbol{y}} + \boldsymbol{A}\bar{\boldsymbol{x}} \in (\boldsymbol{F} + \boldsymbol{A})(\bar{\boldsymbol{x}}) \Big\}.$$

Examples

- For smooth $\boldsymbol{F}$,

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{y} \in F(\bar{x})$, is

$$\inf_{\text{linear } A} \left\{ \|A\| : \text{irregularity for } \bar{y} + A\bar{x} \in (F + A)(\bar{x}) \right\}.$$

Examples

- For smooth F, distance is $\sigma_{\min}(\nabla F(\bar{x}))$:

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{y} \in F(\bar{x})$, is

$$\inf_{\text{linear } A} \Big\{ \|A\| : \text{irregularity for } \bar{y} + A\bar{x} \in (F + A)(\bar{x}) \Big\}.$$

Examples

- For smooth F, distance is $\sigma_{\text{min}}(\nabla F(\bar{x}))$:
  underlies solution speed for $F(x) = 0$.

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{y} \in F(\bar{x})$, is

$$\inf_{\text{linear } A} \Big\{ \|A\| : \text{irregularity for } \bar{y} + A\bar{x} \in (F + A)(\bar{x}) \Big\}.$$

Examples

- For smooth F, distance is $\sigma_{\min}(\nabla F(\bar{x}))$: underlies solution speed for $F(x) = 0$.
- For set intersection,

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{y} \in F(\bar{x})$, is

$$\inf_{\text{linear } A} \Big\{ \|A\| : \text{irregularity for } \bar{y} + A\bar{x} \in (F + A)(\bar{x}) \Big\}.$$

Examples

- For smooth F, distance is $\sigma_{\min}(\nabla F(\bar{x}))$: underlies solution speed for $F(x) = 0$.
- For set intersection, distance is $\sqrt{1 - \cos\theta}$.

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{y} \in F(\bar{x})$, is

$$\inf_{\text{linear } A} \Big\{ \|A\| : \text{irregularity for } \bar{y} + A\bar{x} \in (F + A)(\bar{x}) \Big\}.$$

Examples

- For smooth F, distance is $\sigma_{\text{min}}(\nabla F(\bar{x}))$: underlies solution speed for $F(x) = 0$.
- For set intersection, distance is $\sqrt{1 - \cos\theta}$. Alternating projections converges linearly, rate $\cos\theta$.

9. CONDITIONING AND SPEED

The **distance to ill-posedness** for $\bar{y} \in F(\bar{x})$, is

$$\inf_{\text{linear } A} \Big\{ \|A\| : \text{irregularity for } \bar{y} + A\bar{x} \in (F + A)(\bar{x}) \Big\}.$$

Examples

- For smooth F, distance is $\sigma_{\min}(\nabla F(\bar{x}))$:
  underlies solution speed for $F(x) = 0$.
- For set intersection, distance is $\sqrt{1 - \cos\theta}$.
  Alternating projections converges linearly, rate $\cos\theta$.

Theorem (Dontchev/L/Rockafellar '03) For **any** closed F,

$$\textbf{distance to ill-posedness} \quad = \quad \frac{1}{\textbf{modulus}}.$$

. . . Other examples . . .

...Other examples ...

The modulus often controls local linear convergence rates:

- Proximal point methods: Iusem-Pennanen-Svaiter '03, Aragón-Artacho-Dontchev-Geoffroy '05.
- Klatte/Kummer '07 several conceptual algorithms.
- Luo-Tseng '93: error bounds and descent methods.

Demmel '87: for a square matrix,

$$\text{spectrum hard to compute} \\ \Updownarrow \\ \text{defective matrices are nearby.}$$

10. PSEUDOSPECTRA (Trefethen-Embree ’05)

10. PSEUDOSPECTRA (Trefethen-Embree '05)

Asymptotics of $\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}$ depend on spectrum $\boldsymbol{\Lambda}(\boldsymbol{A})$.

10. PSEUDOSPECTRA (Trefethen-Embree '05)

Asymptotics of $\dot{x} = Ax$ depend on spectrum $\Lambda(A)$.

But **transient** stability depends on the **pseudospectra**

$$\Lambda_\epsilon(A) \;=\; \{z : \sigma_{\mathsf{min}}(A - zI) \le \epsilon\}.$$

10. PSEUDOSPECTRA (Trefethen-Embree '05)

Asymptotics of $\dot{x} = Ax$ depend on spectrum $\Lambda(A)$.

But **transient** stability depends on the **pseudospectra**

$$\Lambda_\epsilon(A) \ = \ \{z : \sigma_{\mathsf{min}}(A - zI) \leq \epsilon\}.$$

Kreiss Matrix Theorem (1962) Equivalent properties:

- $\dot{x} = Ax$ has big transient peaks

10. PSEUDOSPECTRA (Trefethen-Embree '05)

Asymptotics of $\dot{x} = Ax$ depend on spectrum $\Lambda(A)$.

But **transient** stability depends on the **pseudospectra**

$$\Lambda_\epsilon(A) \ = \ \{z : \sigma_{\mathsf{min}}(A - zI) \leq \epsilon\}.$$

Kreiss Matrix Theorem (1962)   Equivalent properties:

- $\dot{x} = Ax$ has big transient peaks
- $\Lambda_\epsilon(A)$ grows quickly into right halfplane as $0 < \epsilon \uparrow$

10. PSEUDOSPECTRA (Trefethen-Embree '05)

Asymptotics of $\dot{x} = Ax$ depend on spectrum $\Lambda(A)$.

But **transient** stability depends on the **pseudospectra**

$$\Lambda_\epsilon(A) \;=\; \{z : \sigma_{\mathsf{min}}(A - zI) \le \epsilon\}.$$

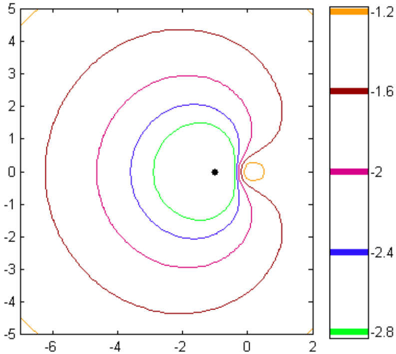
Kreiss Matrix Theorem (1962) Equivalent properties:

- $\dot{x} = Ax$ has big transient peaks
- $\Lambda_\epsilon(A)$ grows quickly into right halfplane as $0 < \epsilon \uparrow$

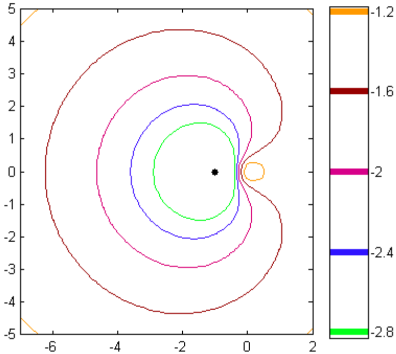
Demmel's example: $A = -\begin{bmatrix} 1 & 5 & 5^2 & 5^3 & 5^4 \\ 0 & 1 & 5 & 5^2 & 5^3 \\ 0 & 0 & 1 & 5 & 5^2 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \ldots$

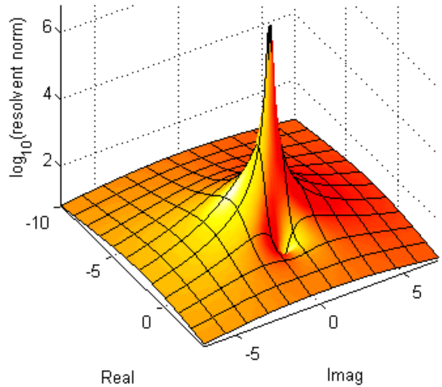
11. EIGTOOL PLOTS (T. Wright)

11. EIGTOOL PLOTS (T. Wright)

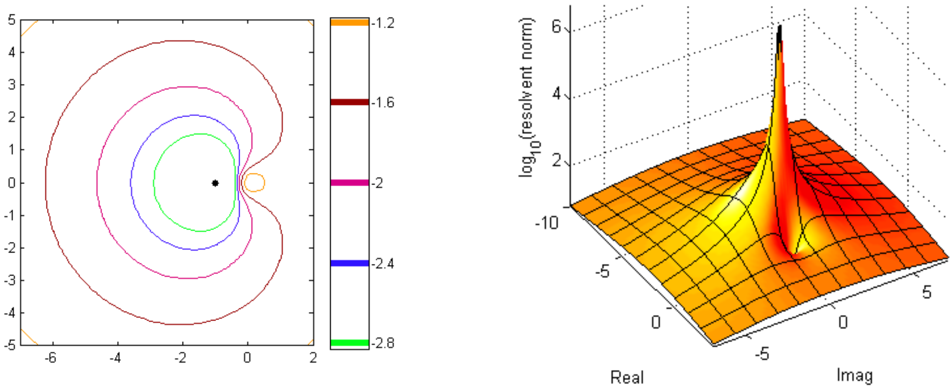


11. EIGTOOL PLOTS (T. Wright)



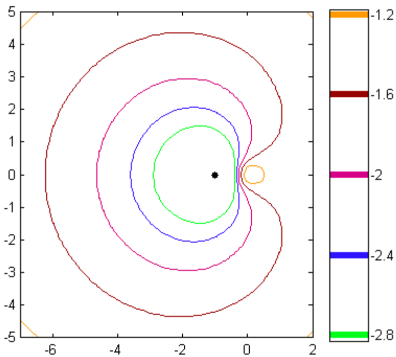


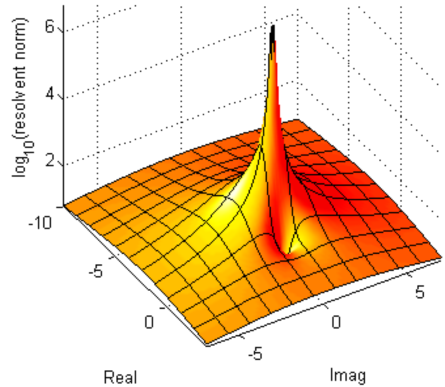
11. EIGTOOL PLOTS (T. Wright)



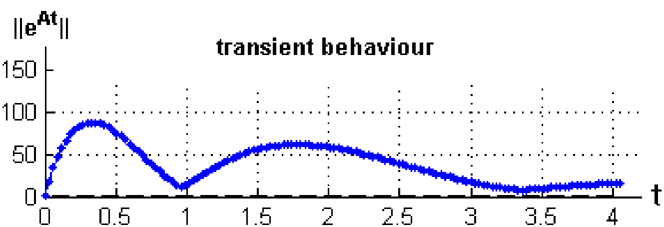
$\mathbf{\Lambda}_{.01}(\boldsymbol{A})$ intersects right halfplane, so
an unstable $\boldsymbol{X}$ satisfies $\|\boldsymbol{X}-\boldsymbol{A}\| \le \mathbf{.01}$.

11. EIGTOOL PLOTS (T. Wright)





$\Lambda_{.01}(A)$ intersects right halfplane, so an unstable X satisfies $\|X - A\| \leq .01$.



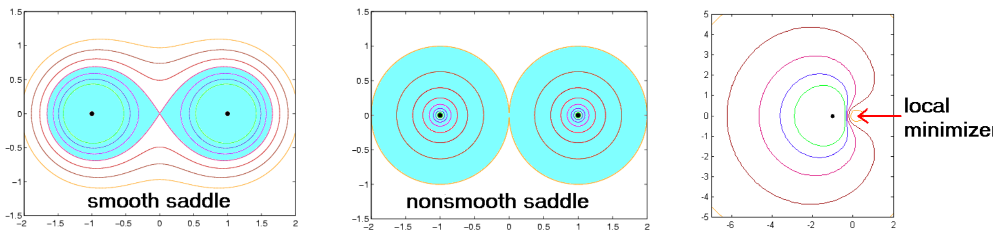
. . . Resolvent-critical points . . .

. . . Resolvent-critical points . . .

On $\mathbb{C}$, the resolvent norm

$$\frac{1}{\sigma_{\min}(A - zI)} = \|(A - zI)^{-1}\|$$

can't have local maxima (by the Maximum Modulus Principle), but can have other smooth and nonsmooth critical points.



(**Aside:** Largest saddle value $\times$ distance to defectiveness $= 1$.)

12. COMPUTING WITH PSEUDOSPECTRA

12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\left\{\mathbf{Re}\,\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \boldsymbol{\Lambda}_\epsilon(A)\right\}.$$

12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\left\{\mathbf{Re}\,\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \boldsymbol{\Lambda}_\epsilon(A)\right\}.$$

Key pseudospectral properties:

12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\Big\{\textbf{Re}\,\lambda : \lambda \in \Lambda_\epsilon(A)\Big\}.$$

Key pseudospectral properties:

- Components contain eigenvalues.

12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\Big\{\mathbf{Re}\,\lambda : \lambda \in \Lambda_\epsilon(A)\Big\}.$$

Key pseudospectral properties:

- Components contain eigenvalues.
- Hamiltonian eigensolvers (Benner et al. '00) compute intersections with lines.

12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\left\{\mathbf{Re}\,\lambda : \lambda \in \Lambda_\epsilon(A)\right\}.$$

Key pseudospectral properties:

- Components contain eigenvalues.
- Hamiltonian eigensolvers (Benner et al. '00) compute intersections with lines.

Hence `eigtool`'s reliable **criss-cross algorithm** for α_ϵ.

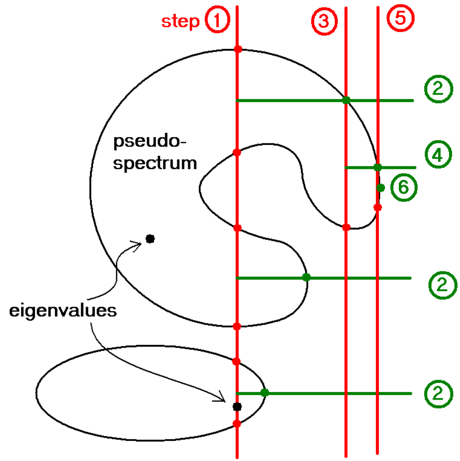
12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\left\{\mathbf{Re}\,\boldsymbol{\lambda} : \boldsymbol{\lambda} \in \Lambda_\epsilon(A)\right\}.$$

Key pseudospectral properties:

- Components contain eigenvalues.
- Hamiltonian eigensolvers (Benner et al. '00) compute intersections with lines.

Hence `eigtool`'s reliable **criss-cross algorithm** for α_ϵ.



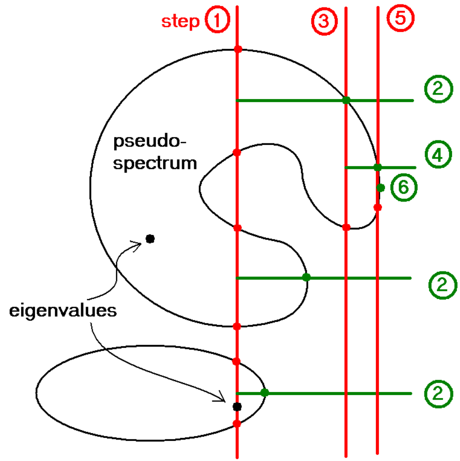
12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\Big\{\mathbf{Re}\,\lambda : \lambda \in \Lambda_\epsilon(A)\Big\}.$$

Key pseudospectral properties:

- Components contain eigenvalues.
- Hamiltonian eigensolvers (Benner et al. '00) compute intersections with lines.

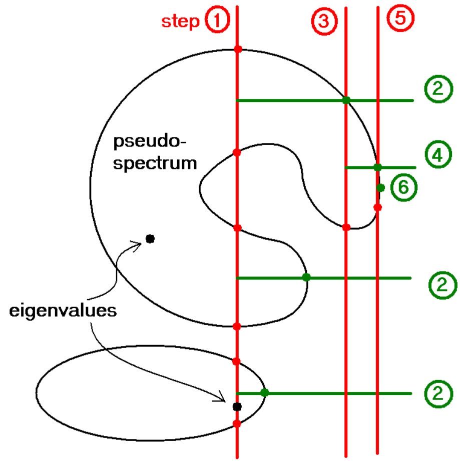
Hence `eigtool`'s reliable **criss-cross algorithm** for α_ϵ.



Global and quadratic convergence: (Burke/L/Overton '03).

12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\Big\{\mathbf{Re}\,\lambda : \lambda \in \Lambda_\epsilon(A)\Big\}.$$



Key pseudospectral properties:

- Components contain eigenvalues.
- Hamiltonian eigensolvers (Benner et al. '00) compute intersections with lines.

Hence `eigtool`'s reliable **criss-cross algorithm** for α_ϵ.

Global and quadratic convergence: (Burke/L/Overton '03).

Nonsmooth methods can optimize α_ϵ

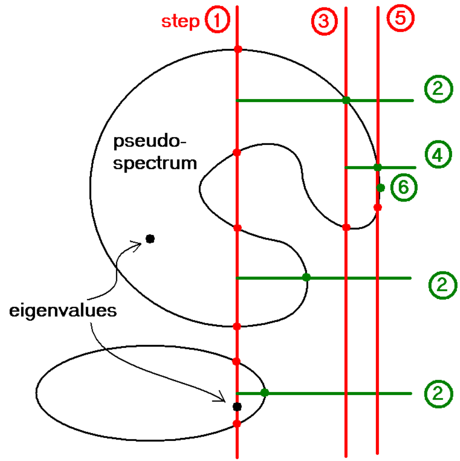
12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\left\{\mathbf{Re}\,\lambda : \lambda \in \Lambda_\epsilon(A)\right\}.$$

Key pseudospectral properties:

- Components contain eigenvalues.
- Hamiltonian eigensolvers (Benner et al. '00) compute intersections with lines.

Hence eigtool's reliable **criss-cross algorithm** for α_ϵ.



Global and quadratic convergence: (Burke/L/Overton '03).

Nonsmooth methods can optimize α_ϵ (and $\mathbf{H}^\infty$ norm:

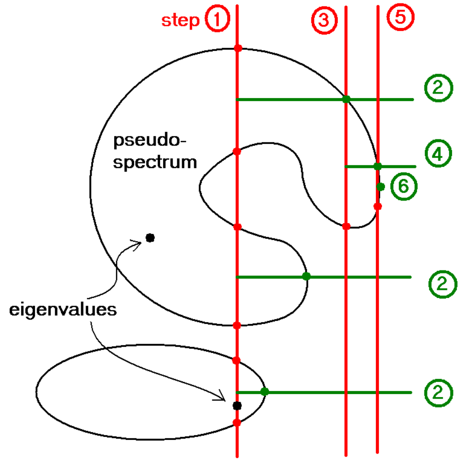
12. COMPUTING WITH PSEUDOSPECTRA

Example: $$\alpha_\epsilon(A) = \max\left\{\mathbf{Re}\,\lambda : \lambda \in \Lambda_\epsilon(A)\right\}.$$

Key pseudospectral properties:

- Components contain eigenvalues.
- Hamiltonian eigensolvers (Benner et al. '00) compute intersections with lines.

Hence `eigtool`'s reliable **criss-cross algorithm** for α_ϵ.



Global and quadratic convergence: (Burke/L/Overton '03).

Nonsmooth methods can optimize α_ϵ (and $\mathbf{H}^\infty$ norm: HIFOO).

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $\boldsymbol{b} \in \boldsymbol{F}(\boldsymbol{x})$, then value $\boldsymbol{b}$ is **critical**.

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $\boldsymbol{b} \in \boldsymbol{F}(\boldsymbol{x})$, then value $\boldsymbol{b}$ is **critical**.

Theorem (Sard '42) Sufficiently smooth $\boldsymbol{F}$ have **almost no** critical values, so $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{b}$ "typically" behaves well.

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $\boldsymbol{b} \in \boldsymbol{F}(\boldsymbol{x})$, then value $\boldsymbol{b}$ is **critical**.

Theorem (Sard '42) Sufficiently smooth $\boldsymbol{F}$ have **almost no** critical values, so $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{b}$ "typically" behaves well.

Example (Whitney '35) $\boldsymbol{F} : \mathbb{R}^{\mathbf{2}} \longrightarrow \mathbb{R}$ is $\boldsymbol{C}^{\mathbf{1}}$ yet nonconstant along an arc of critical points ($\Rightarrow$ **interval** of critical values).

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $b \in F(x)$, then value b is **critical**.

Theorem (Sard '42) Sufficiently smooth F have **almost no** critical values, so $F(x) = b$ "typically" behaves well.

Example (Whitney '35) $F\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ is C^1 yet nonconstant along an arc of critical points ($\Rightarrow$ **interval** of critical values).

Concrete examples often have **semi-algebraic** graphs:

$$\bigcup_i \bigcap_j \left\{ x \in \mathbb{R}^n : p_{ij}(x) \binom{=}{<} 0 \right\}$$

for polynomials p_{ij}.

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $\boldsymbol{b} \in \boldsymbol{F}(\boldsymbol{x})$, then value $\boldsymbol{b}$ is **critical**.

Theorem (Sard '42) Sufficiently smooth $\boldsymbol{F}$ have **almost no** critical values, so $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{b}$ "typically" behaves well.

Example (Whitney '35) $\boldsymbol{F} : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is $\boldsymbol{C}^1$ yet nonconstant along an arc of critical points ($\Rightarrow$ **interval** of critical values).

Concrete examples often have **semi-algebraic** graphs:

$$\bigcup_i \bigcap_j \left\{ \boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{p}_{ij}(\boldsymbol{x}) \binom{=}{<} \boldsymbol{0} \right\}$$

for polynomials $\boldsymbol{p}_{ij}$. **Eg:** pseudospectrum $\boldsymbol{\Lambda}_\epsilon : \mathbb{M}^n \rightrightarrows \mathbb{C}$,

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $b \in F(x)$, then value b is **critical**.

Theorem (Sard '42) Sufficiently smooth F have **almost no** critical values, so $F(x) = b$ "typically" behaves well.

Example (Whitney '35) $F : \mathbb{R}^2 \to \mathbb{R}$ is C^1 yet nonconstant along an arc of critical points ($\Rightarrow$ **interval** of critical values).

Concrete examples often have **semi-algebraic** graphs:

$$\bigcup_i \bigcap_j \left\{ x \in \mathbb{R}^n : p_{ij}(x) \binom{=}{<} 0 \right\}$$

for polynomials p_{ij}. **Eg:** pseudospectrum $\Lambda_\epsilon : \mathbb{M}^n \rightrightarrows \mathbb{C}$, but **not** Whitney's example.

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $b \in F(x)$, then value b is **critical**.

Theorem (Sard '42) Sufficiently smooth F have **almost no** critical values, so $F(x) = b$ "typically" behaves well.

Example (Whitney '35) $F : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is C^1 yet nonconstant along an arc of critical points ($\Rightarrow$ **interval** of critical values).

Concrete examples often have **semi-algebraic** graphs:

$$\bigcup_i \bigcap_j \left\{ x \in \mathbb{R}^n : p_{ij}(x) \binom{=}{<} 0 \right\}$$

for polynomials p_{ij}. **Eg:** pseudospectrum $\Lambda_\epsilon : \mathbb{M}^n \rightrightarrows \mathbb{C}$, but **not** Whitney's example.

Semi-algebraic sets are easy to recognize (Tarski-Seidenberg),

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $b \in F(x)$, then value b is **critical**.

Theorem (Sard '42) Sufficiently smooth F have **almost no** critical values, so $F(x) = b$ "typically" behaves well.

Example (Whitney '35) $F : \mathbb{R}^2 \to \mathbb{R}$ is C^1 yet nonconstant along an arc of critical points ($\Rightarrow$ **interval** of critical values).

Concrete examples often have **semi-algebraic** graphs:

$$\bigcup_i \bigcap_j \left\{ x \in \mathbb{R}^n : p_{ij}(x) \binom{=}{<} 0 \right\}$$

for polynomials p_{ij}. **Eg:** pseudospectrum $\Lambda_\epsilon : \mathbb{M}^n \rightrightarrows \mathbb{C}$, but **not** Whitney's example.

Semi-algebraic sets are easy to recognize (Tarski-Seidenberg), and well-behaved.

13. CRITICAL VALUES OF SET-VALUED MAPPINGS

If metric regularity fails for $\boldsymbol{b} \in \boldsymbol{F}(\boldsymbol{x})$, then value $\boldsymbol{b}$ is **critical**.

Theorem (Sard '42) Sufficiently smooth $\boldsymbol{F}$ have **almost no** critical values, so $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{b}$ "typically" behaves well.

Example (Whitney '35) $\boldsymbol{F} : \mathbb{R}^{\mathbf{2}} \longrightarrow \mathbb{R}$ is $\boldsymbol{C}^{\mathbf{1}}$ yet nonconstant along an arc of critical points ($\Rightarrow$ **interval** of critical values).

Concrete examples often have **semi-algebraic** graphs:

$$\bigcup_{i} \bigcap_{j} \left\{ \boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}} : \boldsymbol{p}_{ij}(\boldsymbol{x}) \binom{=}{<} \mathbf{0} \right\}$$

for polynomials $\boldsymbol{p}_{ij}$. **Eg:** pseudospectrum $\boldsymbol{\Lambda}_{\epsilon} : \mathbb{M}^{\boldsymbol{n}} \rightrightarrows \mathbb{C}$, but **not** Whitney's example.

Semi-algebraic sets are easy to recognize (Tarski-Seidenberg), and well-behaved. Many applications in optimization.

14. A TAME SARD THEOREM

14. A TAME SARD THEOREM

Theorem (Ioffe '07) For any semi-algebraic $\boldsymbol{F} \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$,

$$\text{dim}\,\{\text{critical values}\} < \boldsymbol{m}.$$

14. A TAME SARD THEOREM

Theorem (Ioffe '07) For any semi-algebraic $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$,

$$\dim\{\text{critical values}\} < m.$$

So the system $b \in F(x)$ is well-posed for almost all b.

Application Is pseudospectrum $A \mapsto \Lambda_\epsilon(A)$ well-behaved?

14. A TAME SARD THEOREM

Theorem (Ioffe '07) For any semi-algebraic $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$,

$$\dim \{\text{critical values}\} < m.$$

So the system $b \in F(x)$ is well-posed for almost all b.

Application Is pseudospectrum $A \mapsto \Lambda_\epsilon(A)$ well-behaved?

Using (Ioffe '07) (or (Bolte/Daniilidis/L/Shiota '05)),

$$z \;\mapsto\; \sigma_{\min}(A - zI)$$

has only finitely many critical values.

14. A TAME SARD THEOREM

Theorem (Ioffe '07) For any semi-algebraic $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$,

$$\dim \{\text{critical values}\} < m.$$

So the system $b \in F(x)$ is well-posed for almost all b.

Application Is pseudospectrum $A \mapsto \Lambda_\epsilon(A)$ well-behaved?

Using (Ioffe '07) (or (Bolte/Daniiliidis/L/Shiota '05)),

$$z \;\mapsto\; \sigma_{\min}(A - zI)$$

has only finitely many critical values. Hence (via variational calculus and Toeplitz-Hausdorff Theorem):

$$\Lambda_\epsilon \text{ is Lipschitz around } A \text{ for all small } \epsilon > 0$$

(L/Pang '06).

14. A TAME SARD THEOREM

Theorem (Ioffe '07) For any semi-algebraic $\boldsymbol{F} \colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$,

$$\dim \{\text{critical values}\} < \boldsymbol{m}.$$

So the system $\boldsymbol{b} \in \boldsymbol{F}(\boldsymbol{x})$ is well-posed for almost all $\boldsymbol{b}$.

Application Is pseudospectrum $\boldsymbol{A} \mapsto \boldsymbol{\Lambda}_{\epsilon}(\boldsymbol{A})$ well-behaved?

Using (Ioffe '07) (or (Bolte/Daniiliidis/L/Shiota '05)),

$$\boldsymbol{z} \;\mapsto\; \boldsymbol{\sigma}_{\mathbf{min}}(\boldsymbol{A} - \boldsymbol{z}\boldsymbol{I})$$

has only finitely many critical values. Hence (via variational calculus and Toeplitz-Hausdorff Theorem):

$$\boldsymbol{\Lambda}_{\epsilon} \text{ is Lipschitz around } \boldsymbol{A} \text{ for all small } \boldsymbol{\epsilon} > \mathbf{0}$$

(L/Pang '06). Unlike the spectrum...

15. SUMMARY

15. SUMMARY

Central theoretical ideas:

$$
\begin{array}{ccccc}
 & \text{metric regularity} & \leftrightarrow & \text{distance to ill-posedness} & \\
 & \updownarrow & & & \updownarrow \\
\text{Lipschitz behavior} & \leftrightarrow & \text{critical points} & \leftrightarrow & \text{speed of algorithms}
\end{array}
$$

15. SUMMARY

Central theoretical ideas:

$$
\begin{array}{ccccc}
 & \text{metric regularity} & \longleftrightarrow & \text{distance to ill-posedness} & \\
 & \updownarrow & & & \updownarrow \\
\text{Lipschitz behavior} & \longleftrightarrow & \text{critical points} & \longleftrightarrow & \text{speed of algorithms}
\end{array}
$$

Examples:

- alternating projections
- pseudospectra

15. SUMMARY

Central theoretical ideas:

$$
\begin{array}{ccccc}
 & \text{metric regularity} & \longleftrightarrow & \text{distance to ill-posedness} & \\
 & \updownarrow & & & \updownarrow \\
\text{Lipschitz behavior} & \longleftrightarrow & \text{critical points} & \longleftrightarrow & \text{speed of algorithms}
\end{array}
$$

Examples:

- alternating projections
- pseudospectra

people.orie.cornell.edu/∼aslewis