ROBUSTNESS IN OPTIMIZATION AND CONTROL

Adrian Lewis

Simon Fraser University
Canada

March 10, 2003

Including joint work with
J.V. Burke and M.L. Overton

OUTLINE

- Robust feasibility
- Examples:
  - robust quadratics
  - distances in robust control
  - distance to ill-posedness
- Characterizations, algorithms, variational analysis, optimization
- Robustly regularized functions
- The pseudospectral abscissa

ROBUST FEASIBILITY

Constraint:

variable $x \in S$ (feasible set $\subset \mathbf{R}^n$).

To take into account

- uncertain data
- implementation error
- finite precision,

strengthen to $x + \epsilon B \subset S$

$$\uparrow$$

(unit ball)

or equivalently, distance function

$$\text{dist}_{S^c}(x) \geq \epsilon.$$

Simplest example of

"robust feasibility/optimization".

(cf. Ben-Tal/Nemirovski '01)

EXAMPLES IN ROBUST CONTROL

Dynamical system

$$\frac{dw}{dt} = Aw \quad (A \text{ square}).$$

Desirable sets S of matrices A:

- invertible A
- stable A
  (eigenvalue real parts < 0:
  trajectories $\to 0$ exponentially),

or, for control system $\dot{w} = Aw + Bu$,

- controllable $[A, B]$
  ($[A - zI, B]$ full row rank $\forall z \in \mathbf{C}$:
  all endpoints have interpolating
  trajectories).

ROBUSTNESS

Ensuring matrix $A \in S$ robustly needs dist_{S^c} (measured in operator norm).

Distance to singularity =
smallest singular value $\sigma_{\min}(A)$
(Eckart-Young '39).

Distance to instability =

$$\min_{s \in \mathbf{R}} \sigma_{\min}(A - siI)$$

(if A stable—Van Loan '85).

Distance to uncontrollability =

$$\min_{z \in \mathbf{C}} \sigma_{\min}[A - zI, B]$$

(Eising '84).

RELATED IDEAS: WELL-POSEDNESS

Distance to singularity of square A

$$\updownarrow$$

difficulty of solving $Ax = b$.

Generally, for sublinear set-valued

$$F : \mathbf{R}^n \rightrightarrows \mathbf{R}^m$$

($\{(x,y) : y \in F(x)\}$ a convex cone),
if F surjective ($F(\mathbf{R}^n) = \mathbf{R}^m$),

difficulty of solving $b \in F(x)$

$$\updownarrow$$

"distance to nonsurjectivity"

$$\min_{\text{linear } T} \{\|T\| : F + T \text{ nonsurjective}\}.$$

Eg: (Renegar '95)

$$F(x) = \begin{cases} \{Ax\} & (x \geq 0) \\ \emptyset & \text{(else).} \end{cases}$$

QUESTIONS

For such quantities, we'd like:

- characterizations
- algorithms
- variational properties
- to optimize over variable matrices A.

Corresponding "structured" questions: eg, for given matrices P_i, Q_i,

$$\min_{T_i} \{ \max_i \|T_i\| : \\ A + \sum_{i=1}^{k} P_i T_i Q_i \text{ singular}\}.$$

cf. "structured singular values" (Doyle '82).

DISTANCE TO NONSURJECTIVITY

Following Renegar '95:

Thm For closed sublinear surjective F

$$\min_{\text{linear } T}\{\|T\| : F + T \text{ nonsurjective}\}$$
$$= \min_{\|y\|\leq 1} \max_{x} \left\{ \frac{1}{\|x\|} : y \in F(x) \right\}.$$

Structured version (following Peña '03):

Thm For closed sublinear surjective F

$$\min_{\text{linear } T_i}\{\|T\| : F + \sum_i P_i T_i Q_i \text{ nonsurjec.}\}$$

$$= \min_{\|v_i\|\leq 1} \sup_{\substack{x \\ w_i > 0}} \left\{ \min_i \frac{w_i}{\|Q_i x\|} : \right.$$
$$\left. \sum_i w_i P_i v_i \in F(x) \right\}.$$

COMPUTING DISTANCES

Distance to singularity $\sigma_{\text{min}}(A)$ easy.

Distance to instability

$$\min_{s \in \mathbf{R}} \sigma_{\text{min}}(A - siI)$$

efficiently computable via globally and quadratically convergent algorithm of Boyd et al. '90. ($O(n^3)$ for n-by-n A.) Extends to $\mathbf{H}^{\infty}$ norm

$$\max_{s \in \mathbf{R}} \sigma_{\text{max}}(C(A - siI)^{-1}B - D).$$

Distance to uncontrollability

$$\min_{z \in \mathbf{C}} \sigma_{\text{min}}[A - zI, B]$$

tractable by $O(n^6)$ linearly convergent method following Gu ('00).

OPTIMIZATION

Above distances nonsmooth nonconvex.
How could we optimize over variable A?

Minimization method for a.e. differentiable $f(x)$ (BLO '02):

Gradient-sampling

- Given x_{old}, pick nearby random x^j s;
- Find smallest $d \in \text{conv}\{\nabla f(x^j)\}$;
- Update $x_{\text{new}} = x_{\text{old}} - td$ for $t \geq 0$ chosen by linesearch.

Under reasonable conditions

$\rightarrow$ "Clarke stationary" $\bar{x}$.

i.e.

$$(*) \qquad \limsup_{x \to \bar{x}} \nabla f(x)^T d \geq 0 \quad \forall d.$$

NONSMOOTH REGULARITY

Lipschitz functions f

(i.e. $|f(x) - f(y)| \leq k\|x - y\| \; \forall x, y$)

are a.e. differentiable (Rademacher).

But we need regularity (Clarke '75), i.e.

$$\limsup_{x \to \bar{x}} \nabla f(x)^T d = f'(\bar{x}; d) \quad \forall d,$$

to deduce effective necessary condition:

$$f'(\bar{x}; d) \geq 0 \quad \forall d.$$

Eg: $f(x) = -|x|$
satisfies
condition $(*)$
at $\bar{x} = 0$.

REGULARITY IN ROBUST CONTROL

Idea "Subsmooth" functions

$$f(x) = \max_{u \in U} f_u(x)$$

where U compact and

$(x, u) \mapsto (f_u(x), \nabla f_u(x))$ continuous,

are Lipschitz, regular (Rockafellar '82)

Hence minus distances to

- singularity
- instability
- uncontrollability

(and $\mathbf{H}^\infty$ norm) are Lipschitz and regular usually. Eg: if A invertible

$$-\sigma_{\mathsf{min}}(A) = \max_{\|u\|=1} \{-\|Au\|\}$$

subsmooth (locally).

TURBO-GENERATOR EXAMPLE I

"Complex stability radius" (distance to instability) of turbo-generator model (Hung/MacFarlane '82), before and after optimization (BLO '03).

ROBUST REGULARIZATION

$x \in S = \{x : f(x) \leq 0\}$ robustly means $x + \epsilon B \subset S$, i.e.

$$f_\epsilon(x) = \sup_{x+\epsilon B} f \leq 0.$$

"Robust regularization" $f \mapsto f_\epsilon$ preserves many properties:

- convexity
- continuity (and lower semicontinuity)
- Lipschitzness, subsmoothness.

Approximation: if f continuous at x,

$$f_\epsilon(x) \downarrow f(x) \text{ as } \epsilon \downarrow 0.$$

Minimization: if ϵ small, f_ϵ has a local min near any strict local min of f.

Multiple constraints are easy because

$$(\max\{f, g\})_\epsilon = \max\{f_\epsilon, g_\epsilon\}.$$

EXAMPLE: NORMS OF AFFINES

Given matrix A and vector b, consider

$$f(x) = \|Ax + b\|_2.$$

Apply $\mathcal{S}$-lemma and Schur complement:

$$t \geq f_\epsilon(x) \quad \Leftrightarrow \quad \exists \mu \in \mathbf{R} \;\; \text{with}$$

$$\begin{bmatrix} tI & Ax+b & \epsilon A \\ (Ax+b)^T & t-\mu & 0 \\ \epsilon A^T & 0 & \mu I \end{bmatrix} \succeq 0.$$

(cf. Boyd et al. '94.)

Hence f_ϵ "semidefinite-representable": tractable to compute or use in SDPs. (Similarly for convex quadratics.)

What about

$$f(x) = \|Ax + b\|_2 + c^T x?$$

EXAMPLE: SPECTRAL ABSCISSA

For square A, trajectories of dynamical system $\frac{dw}{dt} = Aw$ decay like $e^{\alpha(A)t}$, where

$$\begin{aligned} \alpha(A) &= \max \mathsf{Re}\, \Lambda(A) \\ \Lambda(A) &= \{\text{eigenvalues of } A\}. \end{aligned}$$

$\alpha(\cdot)$ is nonconvex and nonlipschitz:

$$\alpha \begin{bmatrix} 0 & x \\ 1 & 0 \end{bmatrix}$$

But α <u>is</u> regular (Burke/Overton '01) whenever all eigenvalues have geometric multiplicity one (as usually occurs—BLO '01).

Gradient sampling works well (BLO '02).

PSEUDOSPECTRAL ABSCISSA

Robust regularization of spectral abscissa α is pseudospectral abscissa

$$\begin{aligned}\alpha_\epsilon(A) &= \max\{\alpha(Y) : \|Y - A\| \le \epsilon\} \\ &= \max \mathsf{Re}\, \Lambda_\epsilon(A)\end{aligned}$$

where pseudospectrum

$$\Lambda_\epsilon(A) = \bigcup_{\|Y-A\|\le\epsilon} \Lambda(Y)$$

$$= \{z \in \mathbf{C} : \sigma_{\mathsf{min}}(A - zI) \le \epsilon\}.$$

Note: distance to instability $\le \epsilon$

$$\Updownarrow$$

$\alpha_\epsilon(A) \ge 0$.

PSEUDOSPECTRAL GEOMETRY

Key ideas:

- Each point in pseudospectrum is accessible from an eigenvalue (using max modulus principle etc...).
- Finding intersections of lines with pseudospectral boundary easy via Hamiltonian eigenvalue computation (Van Loan '84, Benner '99).

COMPUTING $\alpha_\epsilon(A)$

(cf. $\mathbf{H}^\infty$-norm algorithm of Boyd et al.)

vertical sweeps

bisection

horizontal sweeps

- Converges globally and (at least generically) quadratically.
- 100-by-100 matrix takes seconds.
- Available in T. Wright's `eigtool`.
- Easy to deduce $\nabla \alpha_\epsilon(A)$ (if it exists).
- $\rightarrow$ optimization (gradient sampling).

TURBO-GENERATOR EXAMPLE II

Spectral abscissa and pseudospectral abscissa of turbo-generator, before and after optimization (BLO '03).

VARIATIONAL ANALYSIS OF α_ϵ

Confidence in gradient sampling algorithm for α_ϵ depends on Lipschitzness and regularity.

Thm (BLO '02) If all eigenvalues of A have geometric multiplicity one, then $\forall \epsilon > 0$ small, α_ϵ is Lipschitz and regular around A.

- Regularity not surprising (perhaps), since α regular around such A.
- But why is α_ϵ Lipschitz?

(α isn't.)

LIPSCHITZ REGULARIZATIONS

Thm Suppose "exceptional set" $E \subset \mathbf{R}^n$ semi-algebraic (defined by polynomial inequalities). Suppose

- f locally Lipschitz on E^c
- f grows sharply away from E: tangents to E can be approximated by directions of linear growth for f.

Then, around
any point,
f_ϵ Lipschitz
$\forall \epsilon > 0$ small.

For spectral abscissa case ($f = \alpha$), E consists of matrices with a multiple active eigenvalue.

SUMMARY

- In optimization generally, and robust control particularly, we study

$$\text{distance to } \{\text{“ill-posed” matrices}\}$$
$$\updownarrow$$
$$\text{robustness.}$$

- We can optimize such functions (locally) by “gradient sampling” (particularly if Lipschitz and regular).
- “Robust regularization” enforces constraints robustly and sometimes tractably:
  - convex quadratics
  - spectral abscissa.
- Pseudospectral abscissa is Lipschitz and regular—we can optimize it.