NONSMOOTH OPTIMIZATION AND EIGENVALUES

Adrian Lewis

Cornell University

June 21, 2005

orie.cornell.edu/∼aslewis

1. OUTLINE

- Convexity, the Lax conjecture, and Lidskii's theorem
- Spectral duality theory
- Normal cones, Clarke regularity, and nonsmooth calculus
- Stable polynomials and matrices
- Numerical methods

2. CONVEXITY AND LIDSKII'S THEOREM

$\mathbf{S}^n = \{n \times n$ symmetric matrices$\}$ $\qquad \langle X, Y \rangle =$ trace(XY)
$\mathbf{S}^n_+ = \{$positive semidefinites$\}$ (a closed convex cone)
$\mathbf{P}^n = \{$permutations$\}$.

The determinant is a **hyperbolic polynomial** relative to I: it's homogeneous, and $\lambda \mapsto \det(X - \lambda I)$ has all real roots, the eigenvalues $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$.

Theorem $\lambda(X) - \lambda(Y) \in$ conv$(\mathbf{P}^n \lambda(X - Y))$.

Corollaries

- $\lambda : \mathbf{S}^n \to \mathbf{R}^n$ is 1-Lipschitz
- λ is monotone relative to the cones $\mathbf{S}^n_+$, $\mathbf{R}^n_+$
- $C \subset \mathbf{R}^n$ is **symmetric** if $C = \mathbf{P}^n C$. If C is also convex, then $\lambda^{-1}(C) = \{X : \lambda(X) \in C\}$ is convex.

3. THE LAX CONJECTURE (1958)

Conjecture Hyperbolic polynomials on $\mathbf{R}^3$ relative to $(1,0,0)$ have the form $p(x) = \det(x_1 I + x_2 A + x_3 B)$ with $A, B \in \mathbf{S}^n$.

True: Helton/Vinnikov 2002, Lewis/Parrilo/Ramana 2004.

Corollary (Gurvits 2004) Lidskii's theorem and its corollaries hold for **any** hyperbolic polynomial p relative to d.

Proof To prove $\lambda(x) - \lambda(y) \in \text{conv}(\mathbf{P}^n \lambda(x-y))$:

- Note
$$w \in \mathbf{R}^3 \mapsto p(w_1 d + w_2 x + w_3 y)$$
is hyperbolic relative to $(1,0,0)$.
- Appeal to the Lax conjecture.
- Apply Lidskii's theorem on $\mathbf{S}^n$. □

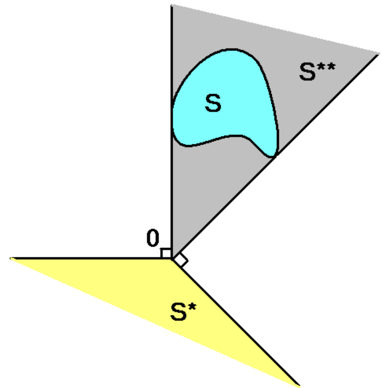
The corollaries have more direct proofs:
Gårding 1951, Bauschke/Güler/Lewis/Sendov 2002.

4. DUALITY

Key tool: for $S \subset \mathbf{R}^n$, the **polar cone** is

$$S^* = \bigcap_{x \in S} \{y : \langle x, y \rangle \leq 0\}.$$

Then $S = S^{**} \;\; \Leftrightarrow$
S is a closed convex cone.



For any symmetric $S \subset \mathbf{R}^n$,

$$(\lambda^{-1}(S))^* = \lambda^{-1}(S^*) \quad \text{in } \mathbf{S}^n$$

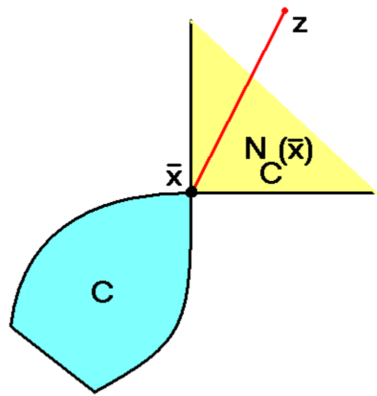
Analogous results for **Fenchel conjugacy** and in other settings:

- duals of unitarily invariant norms (von Neumann 1937)
- Euclidean Jordan algebras (Sun/Sun, Baes 2004)
- Cartan subspace of a semisimple Lie algebra

5. NORMAL CONES AND OPTIMALITY

For $\bar{x} \in$ convex $C \subset \mathbf{R}^n$, the **normal cone** is

$$N_C(\bar{x}) = (C - \bar{x})^*.$$



We have the **best approximation** property:

$\bar{x} \in \text{argmin}\{\|z - x\| : x \in C\}$
$\Leftrightarrow \quad z - \bar{x} \in N_C(\bar{x})$.

Spectral normal cone formula If $C \subset \mathbf{R}^n$ is symmetric and convex, then $X, Y \in \mathbf{S}^n$ satisfy $Y \in N_{\lambda^{-1}(C)}(X) \quad \Leftrightarrow$

$$\begin{aligned} X &= U^T(\text{Diag}\, x)U, \quad & U^T U &= I \\ Y &= U^T(\text{Diag}\, y)U, \quad & y &\in N_C(x). \end{aligned}$$

6. KEY PROPERTIES OF NORMALS

For convex $C \subset \mathbf{R}^n$, it's easy to check:

(i) $N_C(x)$ is a **convex cone**;

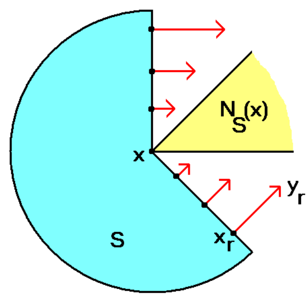
(ii) the **best approximation** property holds

(iii) $x \in C \mapsto N_C(x)$ has **closed graph**:

$$(x_r, y_r) \to (x, y) \text{ and } y_r \in N_C(x_r) \quad \Rightarrow \quad y \in N_C(x).$$

(Crucial for robust optimality conditions and algorithms.)

For **nonconvex** S, the minimal $N_S(\cdot)$ satisfying (i), (ii), (iii) is the **Clarke normal cone** (Clarke 1973).



The **spectral normal cone formula** is unchanged.

Parallel development without convexity in (i) (Mordukhovich 1976).

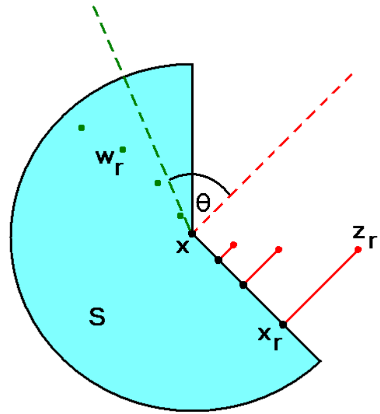
7. CLARKE REGULARITY

If $S \subset \mathbf{R}^n$ is convex, or a manifold, then it's **Clarke regular** at every $x \in S$: any tangent direction to S at x lies in $N_S(x)^*$.

Geometrically, if

- $w_r \in S$ and z_r approach x,
- z_r has nearest point x_r in S,
- the angle between $z_r - x_r$ and $w_r - x$ tends to θ,

then θ is obtuse.



S is **prox-regular** if every point near x has a unique nearest point in S ($\Rightarrow$ regular).

For symmetric $S \subset \mathbf{R}^n$,

$$\lambda^{-1}(S) \text{ regular at } X \in \mathbf{S}^n \quad \Leftrightarrow \quad S \text{ regular at } \lambda(X).$$

8. STABILITY

A polynomial $p(z)$ is **stable** if all its roots lie in left halfplane.
$A \in \mathbf{M}^n = \{n \times n$ matrices$\}$ is **stable** if $\det(A - zI)$ is stable.

Related to asymptotic stability of dynamical systems:

$$A - \alpha I \text{ stable} \quad \Leftrightarrow \quad e^{At} \to 0 \text{ like } e^{\alpha t}.$$

Theorem (Burke/Overton 2000) Everywhere regularity of

$$\textbf{stable monics} \;\; \Delta \; = \; \left\{ w \in \mathbf{C}^n : z^n + \sum_{j=0}^{n-1} w_j z^j \text{ stable} \right\}.$$

If $X \in \mathbf{M}^n$ is **nonderogatory** (1-dimensional eigenspaces),

$$\det(X - zI) = z^n + \sum \Phi(X)_j z^j$$

defines a map Φ with $\nabla\Phi(X)$ onto, so

$$\{\text{stable matrices}\} = \Phi^{-1}(\Delta)$$

is regular at X, with normal cone $\nabla\Phi(X)^* N_\Delta(\Phi(X))$.

9. BELGIAN CHOCOLATE PROBLEM

Motivated by **simultaneous plant stabilization** in control:

Problem (Blondel 1994) Find real stable polynomials p, q, and

$$r(z) = (z^2 - 2\delta z + 1)p(z) + (z^2 - 1)q(z) \qquad (*)$$

with (i) $\delta = 0.9$; (ii) maximal δ. **Prize**: 1kg of chocolate each.

Numerical experiment:

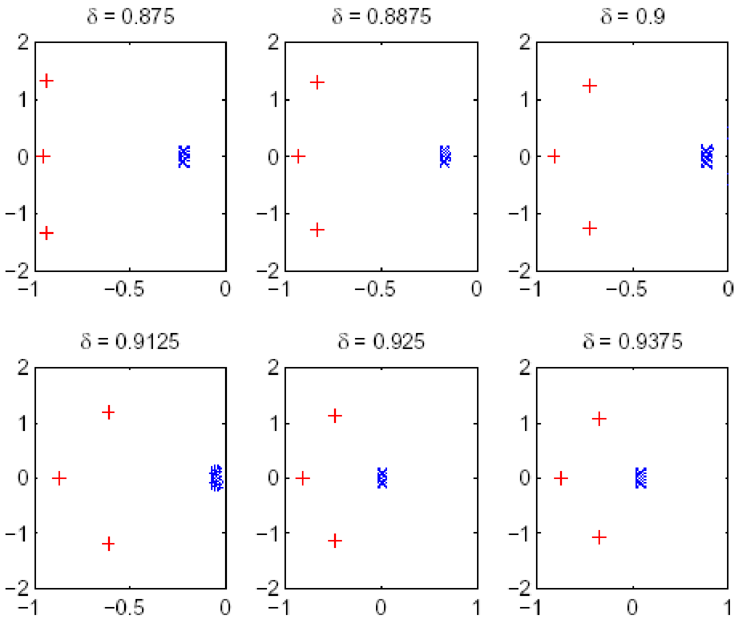
$$\begin{cases} \text{minimize} & \alpha \\ \text{subject to} & p(z-\alpha)q(z-\alpha)r(z-\alpha) \text{ stable monic} \\ & p, q \text{ cubic, } r \text{ given by } (*) \end{cases}$$

For δ near 0.9, the optimal $\bar{p}, \bar{q}, \bar{r}, \bar{\alpha}$ have persistent structure:

$$\bar{q} \text{ scalar}, \quad \bar{r}(w) = k_\delta w^5, \quad \bar{\alpha} < 0 \quad (\textbf{chocolate!}).$$

Once observed computationally, we can verify this structure analytically. (More in Burke's talk. . .)

10. OPTIMAL ROOTS FOR BELGIAN CHOCOLATE

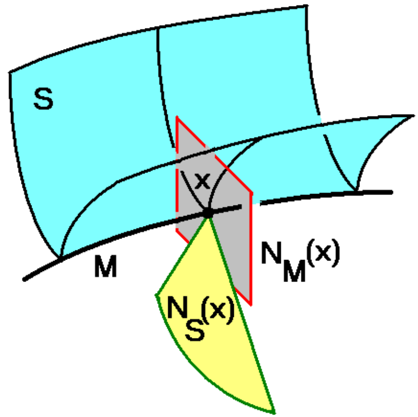


Roots of optimal polynomials $\bar{p}$ ($+$) and $\bar{r}$ ($\times$) for various δ.
Stable if $\delta < \frac{1}{2}\sqrt{2+\sqrt{2}} = 0.92\ldots$ Why does structure persist?

11. PARTLY SMOOTH SETS

Nonsmoothness is common but usually **structured**.

$S \subset \mathbf{R}^n$ is **partly smooth** relative to a manifold $M \subset S$ if



- S is **regular** throughout M
- M is a **ridge** of S: $N_S(x)$ spans $N_M(x)$ for $x \in M$.
- $N_S(\cdot)$ is **continuous** on M.

Examples

- polyhedra, relative to their faces
- $\{x : \text{smooth } g_i(x) \leq 0\}$, relative to $\{x : \textbf{active } g_i(x) = 0\}$
- $\mathbf{S}^n_+$, relative to manifolds of **fixed rank** (Oustry 2000)

12. **SENSITIVITY**

Partial smoothness unifies **active set** ideas in optimization. If

- $S \subset \mathbf{R}^n$ is partly smooth relative to M
- **strict complementarity** for $\boxed{\min_S \langle \bar{y}, \cdot \rangle}$: $y \in \text{ri}\, N_S(\bar{x})$
- $\langle \bar{y}, \cdot \rangle$ **grows quadratically** on M near $\bar{x}$,

then the perturbed problem $\boxed{\min_S \langle y, \cdot \rangle}$ (for y near $\bar{y}$)

- has a unique critical point $x(y) \in M$ near $\bar{x}$
- $x(y)$ depends smoothly on y
- common algorithms **identify** M finitely (Wright 1993)
- $x(y)$ is a local minimizer if S is prox-regular throughout M.

Hence **persistent structure** in optimization: basis persistence in linear programs, rank persistence in semidefinite programs... and the Belgian chocolate problem...

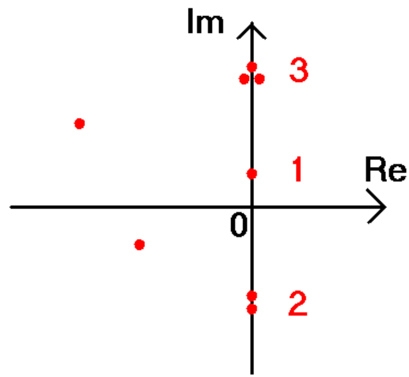
13. THE STABLE SET BOUNDARY

Consider a stable monic p with some purely imaginary roots.

List imaginary root multiplicities down the imaginary axis, and let

$$M = \{\text{monics with same list}\}.$$

Then M is a manifold.



Theorem $\Delta = \{\text{stable monics}\}$ is partly smooth relative to M.

Corollary $\{\text{stable matrices}\}$ partly smooth at any nonderogatory X relative to the manifold of matrices near X with the same imaginary eigenvalue multiplicity list.

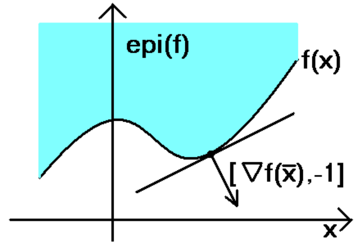
Hence **optimally stable** matrices are typically nonderogatory with multiple eigenvalues (Burke/Lewis/Overton 2000).

14. NONSMOOTH ANALYSIS

For $f : \mathbf{R}^n \to [-\infty, \infty]$,

$\partial f(\bar{x}) = \{y : (y, -1) \in N_{\mathrm{epi}(f)}(\bar{x}, f(\bar{x}))$

(by analogy with the smooth case).



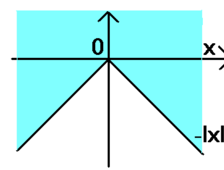
f is (**prox-**)**regular** at $\bar{x}$ if epi(f) is (prox-)regular at $(\bar{x}, f(\bar{x}))$.

For Lipschitz f, $\partial f(\bar{x}) = \mathrm{conv}\,\{\lim \nabla f(x_r) : x_r \to \bar{x}\}$,
and regularity means the directional derivative satisfies

$$f'(\bar{x}; d) = \limsup_{x \to \bar{x}} \langle \nabla f(x), d \rangle \quad \text{for all } d.$$

With regularity, the optimality condition $0 \in \partial f(\bar{x})$ strengthens to

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



By contrast, $0 \in \partial(-|\cdot|)(0)$.

15. DISTANCE TO INSTABILITY

Subsmooth functions $f : \mathbf{R}^n \to \mathbf{R}$ are common in applications: locally, they have the form

$$f(x) = \max_{t \in T} f_t(x) :$$

T is compact, each f_t is $C^{(2)}$, $(x,t) \mapsto \nabla f_t(x)$ is continuous.

Subsmooth functions are prox-regular, so in particular, regular.

Example If $X \in \mathbf{M}^n$ is strictly stable, its **stability radius** (distance from the unstable matrices) is (locally)

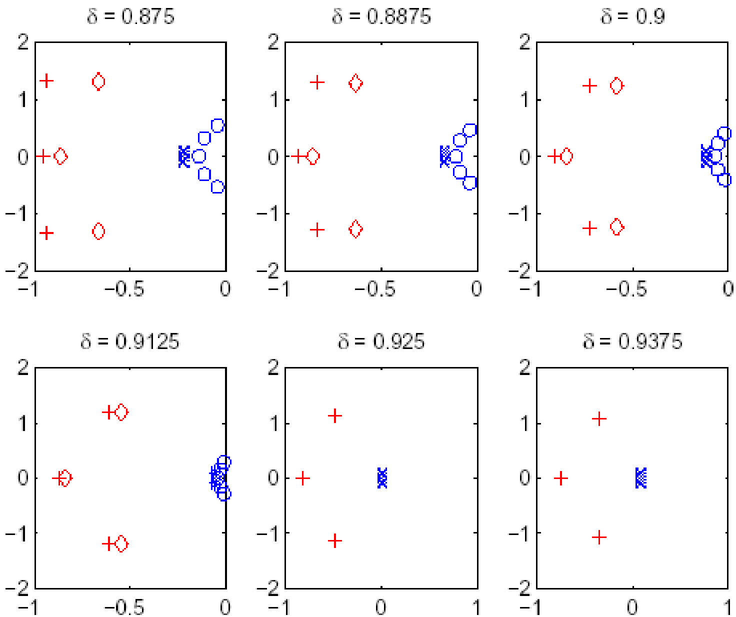
$$\beta(X) \quad = \quad \min\{\|Xu - zu\| : \|u\| = 1,\ \mathrm{Re}\, z \geq 0,\ |z| \leq k\}$$

for large k.

Hence $-\beta$ is regular, so maximizing β is well-behaved.

A similar argument applies to the **$\mathbf{H}^\infty$-norm** in robust control.

16. OPTIMIZING THE STABILITY RADIUS



Roots of polynomials $\bar{p}$ ($\diamondsuit$) and $\bar{r}$ ($\circ$) maximizing $\min\{\beta(p), \beta(r)\}$ (over monic cubic p and scalar q).

17. GRADIENT SAMPLING METHOD

These numerical results were obtained by **gradient sampling**: a simple, intuitive, reliable method for minimizing nonsmooth nonconvex $f : \mathbf{R}^n \to \mathbf{R}$.

If f is Lipschitz and regular at x, **steepest descent** direction is

$$\bar{d} \;=\; -\lim_{\epsilon \downarrow 0} \text{argmin}\{\|d\| : d \in \mathrm{cl\,conv}\, \nabla f(x + \epsilon B)\}$$

(where B is the unit ball). Approximate $\bar{d}$ by **random**

$$\widehat{d} \;=\; -\,\text{argmin}\{\|d\| : d \in \mathrm{conv}\,\{\nabla f(Y_j) : j = 1, \ldots, m\}\}$$

for independent uniform $Y_j \in x + \epsilon B$. ($\epsilon > 0$ and $m > n$ fixed). (Aside: the expected value of $\widehat{d}$ depends continuously on x.) Now do a simple **linesearch** to choose stepsize

$$\bar{t} \;\approx\; \text{argmin}_{t \geq 0} f(x + t\widehat{d}),$$

update $x \leftarrow x + \bar{t}\widehat{d}$, and repeat. **Stop** when $\widehat{d}$ is small.

18. LIDSKII VIA NONSMOOTH CALCULUS

Theorem $\lambda(X) - \lambda(Y) \in \mathrm{conv}(\mathbf{P}^n \lambda(X - Y))$ for $X, Y \in \mathbf{S}^n$.

Proof

- Via a separating hyperplane, we need, for any $w \in \mathbf{R}^n$

  $$w^T(\lambda(Y) - \lambda(X)) \;\leq\; [w]^T \lambda(Y - X),$$

  where $w \mapsto [w]$ maps components into decreasing order.

- Consider the (nonconvex) spectral function

  $$F(X) = w^T \lambda(X).$$

  A **nonsmooth mean value theorem** shows

  $$F(Y) - F(X) = \langle Z, Y - X \rangle$$

  for some $Z \in \partial F(W)$ where $W \in [X, Y]$.

- Now apply the spectral normal cone formula. □

19. SUMMARY

- Eigenvalues of symmetric matrices:
  - Many primal properties extend to hyperbolic polynomials via the Lax conjecture.
  - A rich duality theory leads to normal cone and subgradient formulae.
- Nonsmoothness is often structured: Clarke regularity and partial smoothness.
- Stability optimization is an example: solutions have multiple eigenvalues.
- Gradient sampling is an effective basic numerical tool.