1

Eigenvalues and nonsmooth optimization

Adrian Lewis

Cornell University

Abstract

Variational analysis concerns the geometry and calculus of nonsmooth sets and functions, often viewed from an optimization perspective. Over several decades, variational analysis has matured into a powerful and elegant theory. One rich source of concrete examples involves the eigenvalues of symmetric and nonsymmetric matrices, sometimes deriving from dynamical or feedback control questions. This essay presents some central ideas of variational analysis, developed from first principles, including convexity and duality, generalized gradients, sensitivity, Clarke regularity, and numerical nonsmooth optimization. Illustrative examples from eigenvalue optimization, many from joint work with J.V. Burke and M.L. Overton, include semidefinite programming, asymptotic stability, simultaneous plant stabilization, and the distance to instability.

1.1 Introduction

The eigenvalues of a matrix vary nonsmoothly as we perturb the matrix. For example, as the real parameter τ decreases through zero, the eigenvalues of the matrix

$$\left[\begin{array}{cc} 0 & \tau \\ 1 & 0 \end{array}\right]$$

coalesce at zero from opposite sides of the real axis and then split along the imaginary axis. This inherent nonsmoothness constrains standard developments of eigenvalue perturbation theory, such as Kato (1982), Bhatia (1997). The traditional theory, albeit a powerful tool in many applications, primarily focuses either on precise sensitivity results with

respect to a single parameter, or on broader bounds for more general perturbations.

The modern theory of variational analysis offers an elegant attack on this dilemma. Growing originally out of the calculus of variations, and driven by the broad successes of the systematic approach to convexity popularized by Rockafellar's *Convex Analysis* (1970), the nonconvex theory pioneered by Clarke (1973, 1983) has now blossomed into a comprehensive and powerful framework for optimization and variational problems beyond the realm of classical calculus. The monographs Clarke (1998) and Rockafellar and Wets (1998) give excellent overviews of variational analysis; Borwein and Lewis (2000) is a broad introduction.

This essay sketches the symbiotic relationship between variational analysis and eigenvalue perturbation theory. I illustrate the main themes with examples chosen heavily from my own recent work on symmetric matrices and my collaboration with Jim Burke and Michael Overton on nonsymmetric matrices. On the one hand, the language and tools of variational analysis and nonsmooth optimization crystallize spectral properties of matrices beyond the usual reach of eigenvalue perturbation theory. On the other hand, classical mathematical knowledge about matrix spectra, and their broad applicability, ensure that nonsmooth spectral analysis serves as a significant testing ground for nonsmooth optimization theory.

1.2 Convexity, hyperbolic polynomials, and Lidskii's theorem

Modern variational analysis grew originally from a systematic study of convexity, so it is with convexity that we begin. Eigenvalues of real symmetric matrices exhibit remarkable convexity properties, underlying an explosion of interest throughout the optimization community over the last decade in a far-reaching generalization of linear programming known as *semidefinite programming*: see Ben-Tal and Nemirovski (2001) and Todd (2001).

Denote by $\mathbf{S}^n$ the Euclidean space of n-by-n real symmetric matrices, equipped with the inner product $\langle X, Y \rangle = \text{trace}(XY)$. Within this space, the positive semidefinite matrices $\mathbf{S}^n_+$ constitute a closed convex cone. Semidefinite programming is the study of linear optimization over intersections of $\mathbf{S}^n_+$ with affine subspaces.

An illuminating and strikingly general framework in which to consider the most basic convexity properties of symmetric matrix eigenvalues is that of *hyperbolic polynomials*, a notion originally associated with par-

tial differential equations— see Gårding (1951). The determinant is a hyperbolic polynomial on $\mathbf{S}^n$ relative to the identity matrix I: in other words, it is homogeneous (of degree n), and for any $X \in \mathbf{S}^n$, the polynomial $\lambda \mapsto \det(X - \lambda I)$ has all real roots, namely the eigenvalues $\lambda_1(X) \geq \cdots \geq \lambda_n(X)$. With this notation, we can consider the *characteristic map* $\lambda : \mathbf{S}^n \to \mathbf{R}^n$. A *spectral set* in $\mathbf{S}^n$ is an inverse image of the form

$$\lambda^{-1}(S) = \{X \in \mathbf{S}^n : \lambda(X) \in S\}$$

for any set $S \subset \mathbf{R}^n$.

The core perturbation property of the eigenvalues of symmetric matrices is the following result (which forms a central theme of Bhatia (1997), for example). We denote the group of n-by-n permutation matrices by $\mathbf{P}^n$. For a vector $x \in \mathbf{R}^n$, we denote by $\mathbf{P}^n x$ the set $\{Px : P \in \mathbf{P}^n\}$. Analogously, for a set $S \subset \mathbf{R}^n$, we denote by $\mathbf{P}^n S$ the set $\cup_{x \in S} \mathbf{P}^n x$, and we call S *symmetric* if $\mathbf{P}^n S = S$. We denote the convex hull operation by conv, the standard Euclidean norm on $\mathbf{R}^n$ by $\|\cdot\|$, and the positive orthant and its interior by $\mathbf{R}^n_+$ and $\mathbf{R}^n_{++}$ respectively.

Theorem 1.2.1 (Lidskii, 1950) *Any matrices $X, Y \in \mathbf{S}^n$ satisfy*

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

Immediate corollaries include many important classical properties of eigenvalues of symmetric matrices, some of which are collected below: see Horn and Johnson (1985), and Stewart and Sun (1990). Lidskii's theorem is not the easiest avenue to any of these results, but it does provide a unifying perspective: see Bhatia (1997).

Corollary 1.2.1 (characteristic map behavior) *The characteristic map $\lambda : \mathbf{S}^n \to \mathbf{R}^n$ has the following properties.*

Monotonicity *The map λ is monotone relative to the orderings induced by the cones $\mathbf{S}^n_+$ and $\mathbf{R}^n_+$: any matrices $X, Y \in \mathbf{S}^n$ satisfy*

$$X - Y \in \mathbf{S}^n_+ \quad \Rightarrow \quad \lambda(X) - \lambda(Y) \in \mathbf{R}^n_+.$$

Convexity *If the set $C \subset \mathbf{R}^n$ is symmetric and convex, then the spectral set $\lambda^{-1}(C)$ is convex. In particular, the **hyperbolicity cone** $\lambda^{-1}(\mathbf{R}^n_{++})$ is convex.*

Nonexpansivity *The map λ is nonexpansive: $\|\lambda(X) - \lambda(Y)\| \leq \|X - Y\|$. The same inequality also holds using the infinity norm on $\mathbf{R}^n$ and the spectral norm on $\mathbf{S}^n$.*

Hyperbolic polynomials are strikingly simple to define, and form a broad, rich class: see Bauschke et al. (2001). Nonetheless, hyperbolic polynomials in three or fewer variables have a very specific structure. In one or two variables, this observation is easy and uninteresting; in three variables, it is neither. The following result, conjectured in Lax (1958), was observed in Lewis et al. (2005) to be equivalent to a recent result of Helton and Vinnikov (2002).

Theorem 1.2.2 ("Lax conjecture") *A polynomial p on $\mathbf{R}^3$ is homogeneous of degree 3, hyperbolic relative to the direction $e = (1, 0, 0)$, and satisfies $p(e) = 1$, if and only if it has the form*

$$p(x) = \det(x_1 I + x_2 A + x_3 B)$$

for some matrices $A, B \in \mathbf{S}^n$.

At first sight, the Lax conjecture looks rather narrow in it applicability. However, as the next corollary due to Gurvits (2004) exemplifies, it is a much more general tool than first appearances suggest.

Corollary 1.2.2 *Lidskii's theorem holds for any hyperbolic polynomial.*

Proof Suppose the degree-n polynomial p is hyperbolic on $\mathbf{R}^k$ relative to the direction d. By normalizing, we can suppose $p(d) = 1$. For any vectors x, y, we want to prove $\lambda(x) - \lambda(y) \in \operatorname{conv}(\mathbf{P}^n \lambda(x - y))$. Apply the Lax conjecture to the polynomial on $\mathbf{R}^k$ defined by $w \in \mathbf{R}^3 \mapsto p(w_1 d + w_2 x + w_3 y)$, which is itself hyperbolic relative to e. The result now follows by appealing to Lidskii's theorem on $\mathbf{S}^n$. $\qquad\square$

As an immediate consequence of this result, or alternatively, by directly applying the same proof technique, each part of Corollary 1.2.1 also holds for any hyperbolic polynomial. Each of these results has a more direct proof. The monotonicity result appeared in Gårding (1959), which also contains a short proof of the central fact that the hyperbolicity cone is convex. The more general convexity result appears in Bauschke et al. (2001), along with the nonexpansive property, for which we need to make the nondegeneracy assumption $\lambda(x) = 0 \ \Rightarrow \ x = 0$ and define $\|x\| = \|\lambda(x)\|$.

1.3 Duality and normal cones

A characteristic feature of convex analysis and optimization is the heavy use of duality arguments, featuring separating hyperplanes in various guises: see Rockafellar (1970). The most basic form of this idea is duality for cones. The *dual cone* of a set $S \subset \mathbf{R}^n$ is the closed convex cone

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

(interpreting $\emptyset^* = \mathbf{R}^n$). The set S is itself a closed convex cone if and only $S = S^{**}$.

In addition to the "primal" properties of the characteristic map $\lambda : \mathbf{S}^n \to \mathbf{R}^n$ listed in Corollary 1.2.1, λ also behaves well under duality operations. The following basic theorem is one of several analogous results concerning polar sets and Fenchel conjugate functions in Lewis (1996c).

Theorem 1.3.1 (dual spectral cones) *For symmetric sets $S \subset \mathbf{R}^n$,*

$$(\lambda^{-1}(S))^* = \lambda^{-1}(S^*).$$

This result is reminiscent of von Neumann's 1937 characterization of unitarily invariant matrix norms on the Euclidean space of n-by-n complex matrices $\mathbf{M}^n$ (equipped with the Frobenius norm). Part of von Neumann's development is the formula

$$(\sigma^{-1}(G))^D = \sigma^{-1}(G^D),$$

where $\sigma : \mathbf{M}^n \to \mathbf{R}^n$ maps any matrix to a vector with components its singular values (in decreasing order), G is any symmetric norm-unit-ball satisfying $x \in G \Leftrightarrow |x| \in G$ (the absolute value applied componentwise), and G^D denotes the dual unit ball. Semisimple Lie theory provides one algebraic framework for exploring the parallels between von Neumann's duality formula and Theorem 1.3.1 (dual spectral cones) Lewis (2000). Other authors have investigated results like Theorem 1.3.1 for Euclidean Jordan algebras, a popular setting in which to study interior-point optimization algorithms: see Baes (2004) and Sun and Sun (2004).
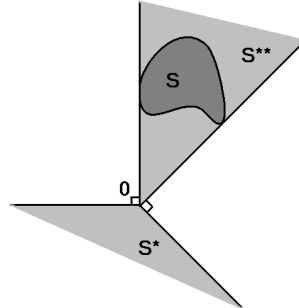
A principal application of the dual cone idea is in the development of optimality conditions for constrained optimization problems. Given a convex set $C \subset \mathbf{R}^n$, the *normal cone* to C at a point $\bar{x} \in C$ is

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

Using this notation, we have the *best approximation condition*:

$$\bar{x} \in \operatorname{argmin}\{\|z - x\| : x \in C\} \quad \Rightarrow \quad z - \bar{x} \in N_C(\bar{x}), \qquad (1.1)$$

(and in fact the converse also holds). Theorem 1.3.1 (dual spectral cones) is a special case of the following characterization of normal cones to spectral sets: see Lewis (1996a).

Theorem 1.3.2 (spectral normal cones) *If the set $C \subset \mathbf{R}^n$ is symmetric and convex, then the spectral set $\lambda^{-1}(C)$ is convex, and matrices $X, Y \in \mathbf{S}^n$ satisfy $Y \in N_{\lambda^{-1}(C)}(X)$ if and only if there exists vectors $x, y \in \mathbf{R}^n$ and a real n-byn matrix U satisfying*

$$
\begin{array}{rcll}
X & = & U^T(\operatorname{Diag} x)U, & U^T U = I \qquad (1.2) \\
Y & = & U^T(\operatorname{Diag} y)U, & y \in N_C(x). \qquad (1.3)
\end{array}
$$

In other words, if we can recognize normals to the symmetric convex set C, then we can recognize normals to the convex spectral set $\lambda^{-1}(C)$ via simultaneous spectral decompositions.

### 1.4 Normals to nonconvex sets and Clarke regularity

The normal cone to a convex set $C \subset \mathbf{R}^n$ has the following key elementary properties, which may be found in Rockafellar (1970), for example.

(i) $N_C(x)$ is a **convex cone** for any point $x \in C$.

(ii) The **best approximation** condition (1.1) holds.

(iii) The set-value mapping $x \in C \mapsto N_C(x)$ has **closed graph**: if $(x_r, y_r) \to (x, y)$ in $\mathbf{R}^n \times \mathbf{R}^n$ and $y_r \in N_C(x_r)$, then $y \in N_C(x)$.

This latter property guarantees some robustness for the normal cone, in theory and algorithmic practice.
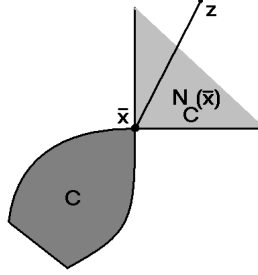
To broaden the context of variational analysis to nonconvex closed sets $S \subset \mathbf{R}^n$ (such as smooth manifolds), we define the *Clarke normal cone mapping* $N_S : S \to \mathbf{R}^n$ to be the set-valued mapping satisfying properties (i), (ii), (iii) with minimal graph: see Clarke (1973) and Clarke et al. (1998). Thus the normal cone at a point $\bar{x} \in S$ consists of all convex combinations of limits of directions from points near $\bar{x}$ to their projections on S.
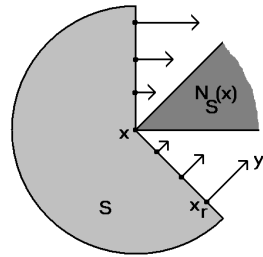
Variational analysis can also be developed in a parallel fashion without the assumption of convexity in property (i): see Mordukhovich (1976) and Rockafellar and Wets (1998). However, the Clarke cone suffices for this essay.

The Clarke normal cone is a useful tool for describing necessary optimality conditions for variational problems. For example, the best approximation condition (1.1) generalizes as follows: see Clarke (1983).

Theorem 1.4.1 (necessary optimality condition) *If the point $\bar{x}$ minimizes the smooth function $f : \mathbf{R}^n \to \mathbf{R}$ on the closed set $S \subset \mathbf{R}^n$, then $-\nabla f(\bar{x}) \in N_S(\bar{x})$.*

We call a closed set S *Clarke regular* at a point $x \in S$ if any tangent direction to S at x lies in $N_S(x)^*$: see Clarke (1983). Geometrically, for any sequences of points $w_r \in S$ and $z_r \in \mathbf{R}^n$ approaching x, if z_r has a nearest point x_r in S, and the angle between $z_r - x_r$ and $w_r - x$ converges to θ, then θ is obtuse. Clarke regularity is in fact independent of the inner product. Convex sets and manifolds are regular at every point x, in fact having the stronger property of *prox-regularity*: every point near x has a unique nearest point in S (see Poliquin et al. (2000)).
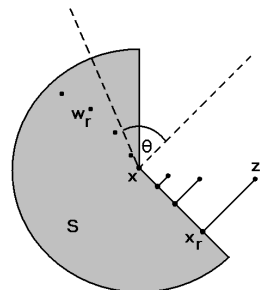
Clarke regularity is a recurrent theme in this essay, and plays a central role both in theoretical variational analysis and in algorithmic matters. The following result, from Clarke (1983), is an example of the kind of calculus that Clarke regularity expedites.

Theorem 1.4.2 (chain rule) *Suppose that the map $\Phi : \mathbf{R}^m \to \mathbf{R}^n$ is smooth around the point $y \in \mathbf{R}^m$, and that the closed set $S \subset \mathbf{R}^n$ is Clarke regular at the point $\Phi(y)$. If*

$$N_S(\Phi(y)) \cap \ker(\nabla\Phi(y))^* = \{0\},$$

then the inverse image $\Phi^{-1}(S)$ is Clarke regular at y, with Clarke normal cone given by

$$N_{\Phi^{-1}(S)}(y) = (\nabla\Phi(y))^* N_S(\Phi(y)).$$

We return to the implications of Clarke regularity for nonsmooth optimization algorithms towards the end of this essay.

Remarkably, the characteristic map behaves just as well with respect to the Clarke normal cone as it does for normal cones to convex sets: see Lewis (1996b). Furthermore, Clarke regularity "lifts" from $\mathbf{R}^n$ to $\mathbf{S}^n$.

Theorem 1.4.3 (spectral Clarke normal cones) *If the set $S \subset \mathbf{R}^n$ is symmetric and closed, then matrices $X, Y \in \mathbf{S}^n$ satisfy $Y \in N_{\lambda^{-1}(S)}(X)$ if and only if equations (1.2) and (1.3) hold. Furthermore, the spectral set $\lambda^{-1}(S)$ is Clarke regular at X if and only if S is Clarke regular at the point $\lambda(X)$.*

This result even remains unchanged for the nonconvex normal cone: see Lewis (1999b).

As an example, consider the optimization problem

$$\sup\{\langle X, Y \rangle : X \in \mathbf{S}^n, \ \lambda(X) = x\},$$

for a given vector $x \in \mathbf{R}^n$ with nonincreasing components and a given matrix $Y \in \mathbf{S}^n$. The characteristic map λ is nonexpansive, by Corollary 1.2.1 (characteristic map behavior), so in particular continuous, and $\|\lambda(X)\| = \|X\|$ for all $X \in \mathbf{S}^n$. Hence continuity and compactness ensure this problem has an optimal solution X_0. Applying Theorem 1.4.1 (necessary optimality condition) shows $Y \in N_\Omega(X_0)$, where Ω is the spectral set $\lambda^{-1}(x) = \lambda^{-1}(\mathbf{P}^n x)$, so Theorem 1.4.3 (spectral Clarke normal cones) shows that the matrices X_0 and Y have a simultaneous spectral decomposition. An elementary argument then shows $\langle X_0, Y \rangle = x^T \lambda(Y)$, so we deduce the well-known inequality (essentially due to von Neumann (1937)).

$$\langle X, Y \rangle \leq \lambda(X)^T \lambda(Y), \quad \text{for any } X, Y \in \mathbf{S}^n. \tag{1.4}$$

1.5 Stability and the Belgian chocolate problem

We turn next to eigenvalues of nonsymmetric matrices. Our primary focus is on the set of *stable* matrices $\mathbf{M}^n_{\mathrm{st}}$, which consists of those matrices in $\mathbf{M}^n$ having all eigenvalues in the closed left halfplane. The stability of a matrix $A \in \mathbf{M}^n$ is closely related to the asymptotic behavior of the dynamical system $\dot{x} = Ax$: specifically, as time t increases, e^{At} decays like $e^{\alpha t}$ if and only if $A - \alpha I$ is stable.

Analogously, a polynomial $p(z)$ is *stable* if all its roots lie in the closed left halfplane: if in fact they lie in the *open* halfplane, we call the polynomial *strictly* stable. Thus a matrix is stable exactly when its characteristic polynomial is stable. The set of stable monic polynomials

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

has the following beautiful variational property: see Burke and Overton (2001b).

Theorem 1.5.1 (regularity of stable polynomials) *The set of stable monic polynomials Δ_n is everywhere Clarke regular.*

The corresponding property for the stable matrices $\mathbf{M}^n_{\mathrm{st}}$ elegantly illustrates the power of nonsmooth calculus. We consider the map $\Phi : \mathbf{M}^n \to \mathbf{C}^n$ taking a matrix $X \in \mathbf{M}^n$ to its characteristic polynomial:

$$\det(X - zI) = z^n + \sum_{j=0}^{n-1} \Phi(X)_j z^j.$$

With this notation we have $\mathbf{M}^n_{\mathrm{st}} = \Phi^{-1}(\Delta_n)$. Even if X has a multiple eigenvalue (as a root of its characteristic polynomial), the *nonderogatory* case where each eigenspace is one-dimensional is "typical" (from the perspective of Arnold's stratification of $\mathbf{M}^n$ into manifolds with fixed Jordan structure—see Arnold (1971)). In this case, the derivative $\nabla \Phi(X)$ is onto, so we can calculate the Clarke normal cone to $\mathbf{M}^n_{\mathrm{st}}$ at X easily using the chain rule (Theorem 1.4.2), thereby recapturing the central result of Burke and Overton (2001a).

Corollary 1.5.1 (regularity of stable matrices) *The set of stable matrices $\mathbf{M}^n_{st}$ is Clarke regular at any stable nonderogatory matrix $X \in \mathbf{M}^n$, with Clarke normal cone*

$$N_{\mathbf{M}^n_{st}}(X) = \nabla \Phi(X)^* N_{\Delta_n}(\Phi(X)).$$

An instructive two-part problem involving sets of stable polynomials was proposed by Blondel (1994), as a challenge to illustrate the difficulty of simultaneous plant stabilization in control. This problem illustrates the interplay between modelling (in this case, control-theoretic), computational experiments, and nonsmooth optimization theory.

Problem   Given a real parameter δ, consider the problem of finding real stable polynomials p, q, r satisfying

$$r(z) = (z^2 - 2\delta z + 1)p(z) + (z^2 - 1)q(z). \qquad (1.5)$$

(Notice the problem admits no solution if $\delta = 1$.) Solve this problem when $\delta = 0.9$, and calculate how far δ can increase before the problem is unsolvable.

Blondel offered a prize of one kilogram of Belgian chocolate for each part of this problem. The first part was solved by a randomized search in Patel et al. (2002). The second part remains open, although, following work surveyed in Patel et al. (2002), the answer is known to be strictly less than one.

Consider the following variational approach. We vary polynomials p and q of fixed degree in order to move the roots of p, q, and r as far to the left in the complex plane as possible. After normalizing so that the product pqr is monic, we arrive at the following numerical optimization problem.

$$(\text{Bl}_\delta) \quad \begin{cases} \text{minimize} & \alpha \\ \text{subject to} & p(z+\alpha)q(z+\alpha)r(z+\alpha) \text{ stable monic} \\ & p, q \text{ cubic}, \ r \text{ given by equation } (1.5) \end{cases}$$

In Section 1.8 we describe a simple, general-purpose, "gradient sampling" method for numerical nonsmooth optimization. Computational experiments with this technique suggest that, for all values of the parameter δ near 0.9, the optimal solution $\bar{p}, \bar{q}, \bar{r}, \bar{\alpha}$ of the problem (Bl_δ) has a persistent structure:

- the polynomial $\bar{q}$ is scalar;
- the polynomial $z \mapsto \bar{r}(z + \bar{\alpha})$ is a multiple of z^5;
- the objective value satisfies $\bar{\alpha} < 0$ **(solving Blondel's problem)**;
- the polynomial $z \mapsto \bar{p}(z + \bar{\alpha})$ is strictly stable.

The figure below (from Burke et al. (2005a)) shows the roots of optimal polynomials $\bar{p}$ (+) and $\bar{r}$ (×) for various values of δ.
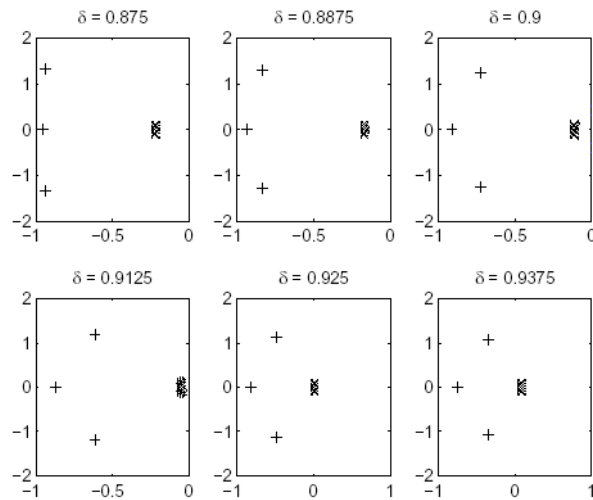
Having observed this structure computationally, some simple algebra shows that for any value of δ near

$$\bar{\delta} = \frac{1}{2}\sqrt{2 + \sqrt{2}} = 0.92\ldots,$$

the problem (Bl_δ) has a unique feasible solution with this structure, solving Blondel's problem for $\delta \leq \bar{\delta}$. Furthermore, a little nonsmooth calculus using Theorem 1.5.1 (regularity of stable polynomials) shows, at least with the extra restriction that q is scalar, that this solution is a strict local minimizer for (Bl_δ): see Burke et al. (2005a).

## 1.6 Partly smooth sets and sensitivity

The persistent structure of optimal solutions for Blondel's problem in the previous section exemplifies a widespread phenomenon in optimization. Assuming appropriate nondegeneracy conditions, optimal solutions for linear, nonlinear, semidefinite and semi-infinite programs all have structures that persist under small perturbations to the problem: in linear programs, the optimal basis is fixed, in nonlinear programs, the active set stays unchanged, and the rank of the optimal matrix in a semidefinite program is constant. Variational analysis offers a unifying perspective on this phenomenon.

Nonsmoothness abounds in optimization, but is usually **structured**. The following definition from Lewis (2003) captures a key structural

idea for the sensitivity analysis of smooth and nonsmooth optimization problems.

We call a closed set $S \subset \mathbf{R}^n$ *partly smooth* relative to a smooth manifold $M \subset S$ if the following properties hold.

- S is **Clarke regular** throughout M.
- M is a **"ridge"** of S: that is, $N_S(x)$ spans $N_M(x)$ for all points $x \in M$.
- The set-valued mapping $x \in M \mapsto N_S(x)$ is **continuous**.
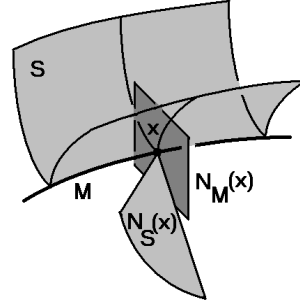
For example, feasible regions of linear programs are polyhedral: any polyhedron is partly smooth relative to the relative interior of any of its faces. Nonlinear programming considers more general feasible regions of the form

$$S = \{x \in \mathbf{R}^n : g_i(x) \le 0 \text{ for } i = 1, 2, \dots, m\},$$

for smooth functions $g_i : \mathbf{R}^n \to \mathbf{R}$. Suppose the point $\bar{x} \in S$ satisfies the *linear independence constraint qualification*:

$\{\nabla g_i(\bar{x}) : i \in I\}$ is linearly independent, where $I = \{i : g_i(\bar{x}) = 0\}$.

In this case the set defined by the active constraints

$$M = \left\{x : g_i(x) = 0 \text{ for } i \in I, \ \|x - \bar{x}\| < \epsilon\right\}$$

is a manifold for small $\epsilon > 0$, relative to which the set S is partly smooth.

As a final example, consider the semidefinite cone $\mathbf{S}_+^n$. In the space $\mathbf{S}^n$, for any integer $r = 0, 1, \dots, n$, the set of matrices in $\mathbf{S}^n$ of rank r constitute a manifold, relative to which $\mathbf{S}_+^n$ is partly smooth. Feasible regions of semidefinite programs are inverse images of $\mathbf{S}_+^n$ under affine maps. We can see that such sets are also partly smooth, using a chain rule analogous to Theorem 1.4.2.

The notion of partial smoothness unifies a variety of **active set** ideas in optimization. Typical sensitivity analysis for variational problems shows that smooth perturbation of the parameters defining a problem leads to a solution that varies smoothly while retaining a constant underlying structure, often reflecting a persistent set of binding or "active" constraints. Partial smoothness abstracts this general observation, generalizing earlier work on convex optimization in Burke and Moré (1988) and Wright (1993).

Consider for example a feasible region $S \subset \mathbf{R}^n$ and an optimization problem

$$(\mathrm{P}_y) \qquad \inf\{\langle y, x \rangle : x \in S\},$$

depending on the parameter $y \in \mathbf{R}^n$. By Theorem 1.4.1 (necessary optimality condition), any optimal solution x for (P_y) must satisfy

$$(\mathrm{OC}_y) \qquad -y \in N_S(x).$$

Suppose the instance $(\mathrm{P}_{\bar{y}})$ (for some particular vector $\bar{y} \in \mathbf{R}^n$) has an optimal solution $\bar{x}$ lying on a manifold $M \subset S$ relative to which S is partly smooth. Let us make two further assumptions, typical in sensitivity analysis:

   (i) the Clarke normal cone $N_S(\bar{x})$ contains the vector $-\bar{y}$ in its relative interior (that is, relative to its span);
   (ii) perturbing the point $\bar{x}$ on M leads to quadratic growth of the linear function $\langle \bar{y}, \cdot \rangle$.

Condition (i) is a strengthening of condition $(\mathrm{OC}_{\bar{y}})$ typically known as a *strict complementarity condition*. Condition (ii) is a *second-order sufficient condition*. With these assumptions, for any y near $\bar{y}$, the optimality condition (OC_y) has a unique solution $x(y) \in M$ near $\bar{x}$, depending smoothly on y. If we assume that S is in fact prox-regular (rather than simply Clarke regular) throughout M, then $x(y)$ must be a local minimizer for the instance (P_y). Furthermore, in this case, a variety of common conceptual algorithms applied to (P_y) "identify" the manifold M finitely: the algorithm generates iterates eventually lying in this manifold— see Hare and Lewis (2004).

Partial smoothness offers a simple unifying language to illuminate the persistent structure of the optimal solutions of perturbed linear, nonlinear, and semidefinite programs. We next apply this idea to the Belgian chocolate problem.

If a polynomial lies on the boundary of the set of stable monics, then it has some purely imaginary roots $iy_1, iy_2, \ldots, iy_r$ (where we assume $y_1 > y_2 > \ldots > y_r$). If each such root iy_j has multiplicity m_j, we call the sequence $m_1, m_2, \ldots, m_r$ the *imaginary multiplicity list*. In the example to the right, the multiplicity list is $3, 1, 2$.
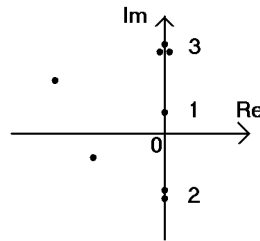
The root cause for the persistent structure in the solutions to the Belgian chocolate problem is then the following result.

Theorem 1.6.1 (partial smoothness of stable polynomials) *Consider a polynomial lying in the set of stable monics Δ_n. The set of nearby monics with the same imaginary multiplicity list constitute a manifold, with respect to which Δ_n is partly smooth.*

Applying a suitable chain rule using the characteristic polynomial map, just as we derived Corollary 1.5.1 (regularity of stable matrices) from Theorem 1.5.1 (regularity of stable polynomials), we deduce the analogous matrix version below: see Lewis (2003).

Corollary 1.6.1 (partial smoothness of stable matrices) *Consider a nonderogatory matrix lying in the stable set $\mathbf{M}_{st}^n$. The set of nearby matrices with the same imaginary eigenvalue multiplicity list constitute a manifold, with respect to which $\mathbf{M}_{st}^n$ is partly smooth.*

In practice, varying a parametrized matrix in order to move its eigenvalues as far as possible into the left halfplane typically leads to nonderogatory optimal solutions with multiple eigenvalues: see Burke et al. (2002b, 2005b). The above result crystallizes the underlying theoretical cause of this phenomenon: see Burke et al. (2000, 2001).

### 1.7   Nonsmooth analysis and the distance to instability

So far in this essay we have taken a geometric approach to variational analysis and nonsmooth optimization, emphasizing the role of the Clarke normal cone. Conceptually, however, the theory is much broader, encompassing powerful generalizations of the derivative and of classical calculus: see Clarke et al. (1998) and Rockafellar et al. (1998). We next briefly sketch the beginnings of this development, building on the geometric ideas we have already introduced.

Consider a function $f : \mathbf{R}^n \to [-\infty, \infty]$ with closed *epigraph*

$$\mathrm{epi}(f) = \{(x, r) \in \mathbf{R}^n \times \mathbf{R} : r \geq f(x)\}.$$

By analogy with the smooth case, we define the *Clarke generalized derivative* by

$$\partial f(\bar{x}) \;=\; \{y : (y, -1) \in N_{\mathrm{epi}(f)}(\bar{x}, f(\bar{x}))\}.$$
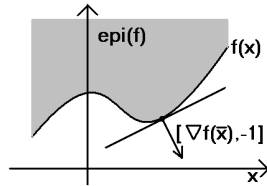
Theorem 1.4.1 (necessary optimality condition) implies the following central role for the generalized derivative in optimization:

$$\bar{x} \text{ minimizes } f \quad \Rightarrow \quad 0 \in \partial f(\bar{x}). \tag{1.6}$$

We call f *Clarke regular* at $\bar{x}$ if epi(f) is Clarke regular at $(\bar{x}, f(\bar{x}))$, and make the analogous definition for prox-regularity: see Poliquin and Rockafellar (1996). For example, any smooth function f is Clarke regular, with generalized derivative $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. Any convex function is also Clarke regular, with generalized derivative agreeing with the classical convex subdifferential: see Rockafellar (1970).

Our approach to the generalized derivative above is appealing in its theoretical economy, but is conceptually opaque. The definition makes little obvious connection with classical differentiation. The following result from Clarke (1973) relates the generalized derivative of a Lipschitz function to the local behavior of its derivative, which is defined almost everywhere by virtue of Rademacher's theorem.

Theorem 1.7.1 (generalized derivatives of Lipschitz functions)
The Clarke generalized derivative of a Lipschitz function $f : \mathbf{R}^n \to \mathbf{R}$ at a point $\bar{x} \in \mathbf{R}^n$ is given by

$$\partial f(\bar{x}) = \text{conv} \{\lim \nabla f(x_r) : x_r \to \bar{x}\}.$$

The function f is Clarke regular at $\bar{x}$ if and only if its directional derivative satisfies

$$f'(\bar{x}; d) = \limsup_{x \to \bar{x}} \langle \nabla f(x), d \rangle$$

for every direction $d \in \mathbf{R}^n$

Without Clarke regularity, the optimality condition (1.6) may be weak. For example, zero *maximizes* minus the absolute value function, yet $0 \in \partial(-|\cdot|)(0)$. With regularity however, (1.6) strengthens to the more intuitive condition $f'(\bar{x}; d) \geq 0$ for all directions d.
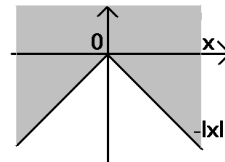
A class of functions very common in applications are those $f : \mathbf{R}^n \to \mathbf{R}$ that can be written locally in the form

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

where the parameter set T is compact, each function $f_t : \mathbf{R}^n \to \mathbf{R}$ is $C^{(2)}$, and the map $(x, t) \mapsto \nabla f_t(x)$ is continuous. Such functions

are called *lower-$C^{(2)}$*: they are prox-regular, so in particular, Clarke regular—see Rockafellar et al. (1998).

A typical example of a lower-$C^{(2)}$ function, arising in robust control systems design, is called the *distance to instability* in Byers (1988), and is also known as the *complex stability radius*—see Hinrichson and Pritchard (1986). It is the distance from a matrix in $\mathbf{M}^n$ to the set of unstable matrices. An easy argument shows that, for any matrix $X_0 \in \mathbf{M}^n$, this function $\beta : \mathbf{M}^n \to \mathbf{R}$ can be written in the form

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

for all matrices $X \in \mathbf{M}^n$ near X_0, where the constant k depends on X_0. If X_0 is strictly stable, then the quantity $\|Xu - zu\|$ is bounded away from zero for all X near X_0, unit vectors $u \in \mathbf{C}^n$, and complex z with $\mathrm{Re}\, z \geq 0$. Consequently, the function $-\beta$ is lower-$C^{(2)}$ on the strictly stable matrices. For the $\mathbf{H}^\infty$-norm in robust control (see Zhou et al. (1996)), a similar analysis applies.

The figure below, from Burke et al. (2005a), shows the results of maximizing the minimum of the two distances to instability of the companion matrices corresponding to the polynomials p and r in the chocolate problem. We restrict p to be a monic cubic, plotting its roots as $\diamondsuit$, and q to be a scalar; we plot the roots of r as $\circ$. To compare, we leave the old optimally stable roots in the plot. Notice how maximizing the stability radius causes the root of order five to split, moving the roots closer to the imaginary axis but nonetheless increasing the distance to instability.
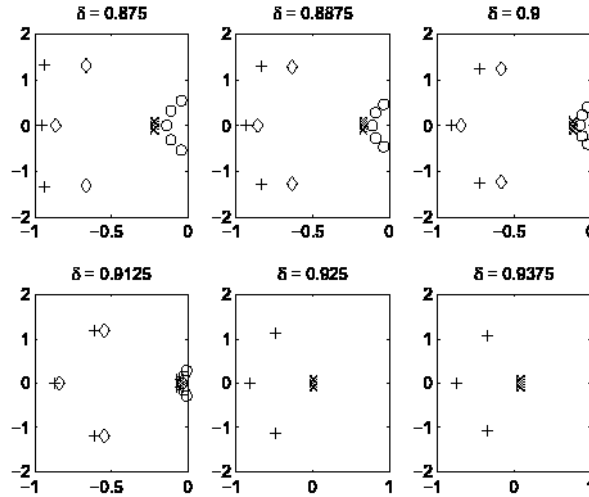
1.8 The gradient sampling method

Despite half a century of advances in computational optimization, and several decades of development in the theory of nonsmooth optimization, numerical minimization of nonsmooth nonconvex functions remains challenging: dependable publicly-available code is scarce. The results described above for the Belgian chocolate problem were obtained in Burke et al. (2005a) using a simple intuitive numerical method based on *gradient sampling*: see Burke et al. (2002b, 2005b).

To motivate this method, consider a Lipschitz function $f : \mathbf{R}^n \to \mathbf{R}$ that is Clarke regular at the point $x \in \mathbf{R}^n$. The *direction of steepest descent* is the unit vector $u \in \mathbf{R}^n$ minimizing the directional derivative $f'(x; u)$. Theorem 1.7.1 (generalized derivatives of Lipschitz functions) shows that this vector lies in the direction of the vector

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

(where B is the unit ball in $\mathbf{R}^n$). For example, if f is smooth at x, then $\overline{d} = -\nabla f(x)$.

To approximate the direction $\overline{d}$, we fix some small radius $\epsilon > 0$, and sample some number $m > n$ random, independent, uniformly distributed points $Y_j \in x + \epsilon B$. Almost surely, f is differentiable at each Y_j: we assume, as is often the case in practice, that the gradients $\nabla f(Y_j)$ are readily available. We then define an approximate steepest descent direction by

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

In practice, we choose $Y_1 = x$, to guarantee $f'(x; \widehat{d}) < 0$.

Finally, we imitate the classical steepest descent method for smooth minimization. We perform a simple *linesearch* to choose a *stepsize*

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

(in practice often simply requiring the descent condition $f(x + \overline{t}\widehat{d}) < f(x)$). We then update $x \leftarrow x + \overline{t}\widehat{d}$, and repeat the whole process. The loop terminates when the vector $\widehat{d}$ is small, at which point we may stop, or restart with a smaller radius ϵ.

For nonsmooth nonconvex minimization problems with cheap function and gradient evaluations and involving not too many variables, computational experience suggests that the gradient sampling method is a robust and reliable tool: see Burke et al. (2005b). The random sampling approach to approximating the Clarke generalized derivative is motivated

theoretically in Burke et al. (2002a). Under reasonable conditions and suitably implemented, the minimization method converges almost surely to a point whose generalized derivative contains zero (see Burke et al. (2005b)): as discussed above, assuming Clarke regularity, this condition guarantees that there are no descent directions. Random sampling helps the method avoid a common difficulty in designing nonsmooth optimization algorithms: the expected value of the random search direction $\widehat{d}$ depends continuously on the current point x (see Lewis (2005)), so the algorithm does not "jam".

1.9 Lidskii's theorem again

The Clarke normal cone and generalized derivative are powerful and versatile tools. Our discussion in the previous section indicates their use in understanding algorithms for nonsmooth optimization. Our analysis of the Belgian chocolate problem and subsequent sketch of the idea of partial smoothness suggests the utility of nonsmooth language for optimality conditions and sensitivity analysis. To bring this essay full circle, we end with a purely analytic application of nonsmooth optimization ideas, using variational analysis to prove Lidskii's eigenvalue perturbation result (Theorem 1.2.1). The argument follows Lewis (1999a).

Given two matrices $X, Y \in \mathbf{S}^n$, we wish to show

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

If this inclusion fails, the separating hyperplane theorem implies the existence of a vector $w \in \mathbf{R}^n$ satisfying the inequality

$$w^T(\lambda(X) - \lambda(Y)) \;>\; \max_{P \in \mathbf{P}^n} w^T P \lambda(X - Y).$$

An elementary argument identifies the right hand side as $[w]^T \lambda(X - Y)$, where the vector $[w] \in \mathbf{R}^n$ has the same components as w rearranged into nonincreasing order.

Turning our attention to the left hand side of the above inequality, we consider the (nonconvex) spectral function $F = w^T \lambda$. A suitable nonsmooth version of the mean value theorem (see Clarke (1983)) applied to this function shows that, for some matrix V on the line segment between X and Y and some matrix $Z \in \partial F(V)$, we have

$$F(X) - F(Y) = \langle Z, X - Y \rangle \le \lambda(Z)^T \lambda(X - Y),$$

using von Neumann's inequality (1.4). The analogous result to Theorem

1.4.3 (spectral Clarke normal cones) for generalized derivatives shows there exists vectors $v, z \in \mathbf{R}^n$ and a real n-by-n matrix U satisfying

$$
\begin{aligned}
V &= U^T(\operatorname{Diag} v)U, \quad U^TU = I \\
Z &= U^T(\operatorname{Diag} z)U, \quad z \in \partial f(v),
\end{aligned}
$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is the function defined by $f(v) = w^T[v]$. Theorem 1.7.1 (generalized derivatives of Lipschitz functions) implies $\partial f(v) \subset \operatorname{conv}(\mathbf{P}^n w)$, so

$$
\lambda(Z) = [z] \in \operatorname{conv}(\mathbf{P}^n[w]).
$$

We quickly deduce the contradiction $\lambda(Z)^T\lambda(X-Y) \leq [w]^T\lambda(X-Y)$, completing the proof.

1.10 Final thoughts

Modern variational analysis deserves a wider mathematical audience than it has so far reached. This essay aims to make converts, by illustrating the elegant interplay between eigenvalues and nonsmooth optimization.

Convexity is a ubiquitous mathematical idea, and its significance for spectral properties of symmetric matrices is well known: Lidskii's theorem is a central example. Hyperbolic polynomials provide a starkly simple setting in which to view many of these classical properties. Conversely, the truth of the Lax conjecture extends a wealth of symmetric matrix results and methods, including Lidskii's theorem, to hyperbolic polynomials.

Over several decades, convex analysis has developed into a widely-used language for diverse areas of mathematics beyond just optimization, neatly unifying the geometry, duality, and calculus of convex sets and functions, and their normal cones and generalized derivatives. Building on this success, the nonconvex theory pioneered by Clarke has matured into a versatile toolkit. The spectral behavior of symmetric matrices provides a powerful illustration of this theory in action.

Far from being pathological or rare, nonsmoothness is fundamental to our understanding of optimization. Its occurrence in concrete problems is typically structured. In particular, Clarke regularity is often a reasonable assumption, with far-reaching implications both in theory and computational practice. Many nonsmooth optimization problems exhibit partial smoothness, an advantageous mix of smooth and non-

smooth behavior helping our understanding of sensitivity analysis and algorithm convergence.

Optimizing the stability of polynomials or matrices exemplifies partially smooth behavior: active roots or eigenvalues cluster at optimal solutions. The Belgian chocolate problem is a typical illustration, both of the theory and of the robust effectiveness of Gradient Sampling, a simple and intuitive computational approach to nonsmooth optimization.

Variational analysis, in its full generality, is less familiar and more challenging than in the convex case. However, the nonsmooth nature of eigenvalues makes it a natural ingredient for perturbation theory in particular, and matrix analysis in general. It seems likely that the interplay between eigenvalues and nonsmooth optimization, briefly sketched in this essay, will continue to flourish.

References

Arnold, V.I. (1971). On matrices depending on parameters. *Russian Mathematical Surveys*, 26:29–43.

Baes, M. (2004). Spectral functions on Jordan algebras: differentiability and convexity properties. Technical report, CORE, Louvain, Belgium.

Bauschke, H.H., Güler, O., Lewis, A.S., and Sendov, H.S. (2001). Hyperbolic polynomials and convex analysis. *Canadian Journal of Mathematics*, 53:470–488.

Ben-Tal, A. and Nemirovski, A. (2001). *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, Philadelphia.

Bhatia, R. (1997). *Matrix Analysis*. Springer, New York.

Blondel, V.D. (1994). *Simultaneous stabilization of linear systems*. Springer, Berlin.

Borwein, J.M. and Lewis, A.S. (2000). *Convex Analysis and Nonlinear Optimization*. Springer, New York.

Burke, J.V., Henrion, D., Lewis, A.S., and Overton, M.L. (2005a). Analysis of a Belgian chocolate stabilization problem. *IEEE Transactions on Automatic Control*. To appear.

Burke, J.V., Lewis, A.S., and Overton, M.L. (2000). Optimizing matrix stability. *Proceedings of the American Mathematical Society*, 129:1635–1642.

Burke, J.V., Lewis, A.S., and Overton, M.L. (2001). Optimal stability and eigenvalue multiplicity. *Foundations of Computational Mathematics*, 1:205–225.

Burke, J.V., Lewis, A.S., and Overton, M.L. (2002a). Approximating subdifferentials by random sampling of gradients. *Mathematics of Operations Research*, 27:567–584.

Burke, J.V., Lewis, A.S., and Overton, M.L. (2002b). Two numerical methods for optimizing matrix stability. *Linear Algebra and its Applications*, 351/2:117–145.

Burke, J.V., Lewis, A.S., and Overton, M.L. (2003). A nonsmooth, nonconvex optimization approach to robust stabilization by static output feedback and low-order controllers. Technical report, Courant Institute. To appear, Proceedings of the 4th IFAC Symposium on Robust Control Design, RO-COND'03.

Burke, J.V., Lewis, A.S., and Overton, M.L. (2005b). A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM Journal on Optimization*, 15:751–779.

Burke, J.V. and Moré, J.J. (1988). On the identification of active constraints. *SIAM Journal on Numerical Analysis*, 25:1197–1211.

Burke, J.V. and Overton, M.L. (2001a). Variational analysis of non-Lipschitz spectral functions. *Mathematical Programming*, 90:317–352.

Burke, J.V. and Overton, M.L. (2001b). Variational analysis of the abscissa mapping for polynomials. *SIAM Journal on Control and Optimization*, 39:1651–1676.

Byers, R. (1988). A bisection method for computing the distance of a stable matrix to the unstable matrices. *SIAM Journal on Scientific and Statistical Computing*, 9:875–881.

Clarke, F.H. (1973). *Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations*. PhD thesis, University of Washington, Seattle.

Clarke, F.H. (1983). *Optimization and Nonsmooth Analysis*. Wiley, New York. Republished as Vol. 5, Classics in Applied Mathematics, SIAM, 1990.

Clarke, F.H., Ledyaev, Yu.S., Stern, R.J., and Wolenski, P.R. (1998). *Nonsmooth Analysis and Control Theory*. Springer-Verlag, New York.

Gårding, L. (1951). Linear hyperbolic differential equations with constant coefficients. *Acta Mathematica*, 85:2–62.

Gårding, L. (1959). An inequality for hyperbolic polynomials. *Journal of Mathematics and Mechanics*, 8:957–965.

Gurvits, L. (2004). Combinatorics hidden in hyperbolic polynomials and related topics. Technical report, Los Alamos National Laboratory.

Hare, W.L. and Lewis, A.S. (2004). Identifying active constraints via partial smoothness and prox-regularity. *Journal of Convex Analysis*, 11:251–266.

Helton, J.W. and Vinnikov, V. (2002). Linear matrix inequality representation of sets. Technical report, Department of Mathematics, University of California San Diego. Available as `arXiv:math.OC/0306180v1 11 Jun 2003`.

Hinrichsen, D. and Pritchard, A.J. (1986). Stability radii of linear systems. *Systems and Control Letters*, 7:1–10.

Horn, R.A. and Johnson, C. (1985). *Matrix Analysis*. Cambridge University Press, Cambridge, U.K.

Kato, T. (1982). *A Short Introduction to Perturbation Theory for Linear Operators*. Springer-Verlag, New York.

Lax, P.D. (1958). Differential equations, difference equations and matrix theory. *Communications on Pure and Applied Mathematics*, 6:175–194.

Lewis, A.S. (1996a). Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6:164–177.

Lewis, A.S. (1996b). Derivatives of spectral functions. *Mathematics of Operations Research*, 6:576–588.

Lewis, A.S. (1996c). Group invariance and convex matrix analysis. *SIAM Journal on Matrix Analysis and Applications*, 17:927–949.

Lewis, A.S. (1999a). Lidskii's theorem via nonsmooth analysis. *SIAM Journal on Matrix Analysis and Applications*, 21:379–381.

Lewis, A.S. (1999b). Nonsmooth analysis of eigenvalues. *Mathematical Programming*, 84:1–24.

Lewis, A.S. (2000). Convex analysis on Cartan subspaces. *Nonlinear Analysis, Theory, Methods and Applications*, 42:813–820.

Lewis, A.S. (2003). Active sets, nonsmoothness and sensitivity. *SIAM Journal on Optimization*, 13:702–725.

Lewis, A.S. (2005). Local structure and algorithms in nonsmooth optimization. In Jarre, F., Lemaréchal, C., and Zowe, J., editors, *Oberwolfach Proceedings: Workshop on Optimization*.

Lewis, A.S., Parrilo, P.A., and Ramana, M.V. (2005). The Lax conjecture is true. *Proceedings of the American Mathematical Society*, 133:2495–2499.

Mordukhovich, B.S. (1976). Maximum principle in the problem of time optimal response with nonsmooth constraints. *Journal of Applied Mathematics and Mechanics*, 40:960–969.

von Neumann, J. (1937). Some matrix inequalities and metrization of matricspace. *Tomsk University Review*, 1:286–300. In: *Collected Works*, Pergamon, Oxford, 1962, Volume IV, 205-218.

Patel, V.V., Doehare, G., and Viswanath, T. (2002). Some applications of a randomized algorithms for control system design. *Automatica*, 28:2085–2092.

Poliquin, R.A. and Rockafellar, R.T. (1996). Prox-regular functions in variational analysis. *Transactions of the American Mathematical Society*, 348:1805–1838.

Poliquin, R.A., Rockafellar, R.T., and Thibault, L. (2000). Local differentiability of distance functions. *Transactions of the American Mathematical Society*, 352:5231–5249.

Rockafellar, R.T. (1970). *Convex Analysis*. Princeton University Press, Princeton, N.J.

Rockafellar, R.T. and Wets, R.J.-B. (1998). *Variational Analysis*. Springer, Berlin.

Stewart, G.W. and Sun, J.G. (1990). *Matrix Perturbation Theory*. Academic Press, Boston.

Sun, D. and Sun, J. (2004). Löwner's operator and spectral functions in Euclidean Jordan algebras. Technical report, National University of Singapore.

Todd, M.J. (2001). Semidefinite optimization. *Acta Numerica*, 10:515–560.

Wright, S.J. (1993). Identifiable surfaces in constrained optimization. *SIAM Journal on Control and Optimization*, 31:1063–1079.

Zhou, K., Doyle, J.C., and Glover, K. (1996). *Robust and Optimal Control*. Prentice-Hall, Upper Saddle River, NJ.