Mathematical Programming manuscript No.
(will be inserted by the editor)

A.S. Lewis

The mathematics of eigenvalue optimization

Received: date / Revised version: date

Abstract. Optimization problems involving the eigenvalues of symmetric and nonsymmetric matrices present a fascinating mathematical challenge. Such problems arise often in theory and practice, particularly in engineering design, and are amenable to a rich blend of classical mathematical techniques and contemporary optimization theory. This essay presents a personal choice of some central mathematical ideas, outlined for the broad optimization community. I discuss the convex analysis of spectral functions and invariant matrix norms, touching briefly on semidefinite representability, and then outlining two broader algebraic viewpoints based on hyperbolic polynomials and Lie algebra. Analogous nonconvex notions lead into eigenvalue perturbation theory. The last third of the article concerns stability, for polynomials, matrices, and associated dynamical systems, ending with a section on robustness. The powerful and elegant language of nonsmooth analysis appears throughout, as a unifying narrative thread.

Key words. Eigenvalue optimization – convexity – nonsmooth analysis – duality – semidefinite program – subdifferential – Clarke regular – chain rule – sensitivity – eigenvalue perturbation – partly smooth – spectral function – unitarily invariant norm – hyperbolic polynomial – stability – robust control – pseudospectrum – $\mathbf{H}_\infty$ norm

---

PART I: INTRODUCTION

1. Von Neumann and invariant matrix norms

Before outlining this survey, I would like to suggest its flavour with a celebrated classical result. This result, von Neumann's characterization of unitarily invariant matrix norms, serves both as a historical jumping-off point and as an elegant juxtaposition of the central ingredients of this article.

Von Neumann [64] was interested in *unitarily invariant* norms $\|\cdot\|$ on the vector space $\mathbf{M}^n$ of n-by-n complex matrices:

$$\|UXV\| = \|X\| \ \text{ for all } U, V \in \mathbf{U}^n,\ X \in \mathbf{M}^n,$$

where $\mathbf{U}^n$ denotes the group of unitary matrices. The singular value decomposition shows that the invariants of a matrix X under unitary transformations of the form $X \mapsto UXV$ are given by the singular values $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_n(X)$, the eigenvalues of the matrix $\sqrt{X^*X}$. Hence any invariant norm $\|\cdot\|$ must be a

---

A.S. Lewis: Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. e-mail: `aslewis@sfu.ca`. Website: `www.cecm.sfu.ca/~aslewis`. Research supported in part by NSERC.

Mathematics Subject Classification (1991): 90C30, 15A42, 65F15, 49K40

function of the vector $\sigma(X)$, so we can write $\|X\| = g(\sigma(X))$ for some function $g : \mathbf{R}^n \to \mathbf{R}$. We ensure this last equation if we define $g(x) = \|\mathrm{Diag}\, x\|$ for vectors $x \in \mathbf{R}^n$, and in that case g is a *symmetric gauge function*: that is, g is a norm on $\mathbf{R}^n$ whose value is invariant under permutations and sign changes of the components. Von Neumann's beautiful insight was that this simple necessary condition is in fact also sufficient.

Theorem 1.1 (von Neumann, 1937). *The unitarily invariant matrix norms are exactly the symmetric gauge functions of the singular values.*

Von Neumann's proof rivals the result in elegance. He proceeds by calculating the dual norm of a matrix $Y \in \mathbf{M}^n$,

$$\|Y\|_* = \max\{\langle X, Y\rangle : \|X\| = 1\}$$

where the real inner product $\langle X, Y\rangle$ in $\mathbf{M}^n$ is the real part of the trace of X^*Y. Specifically, he shows that any symmetric gauge function g satisfies the simple duality relationship

$$(g_* \circ \sigma)_* = g \circ \sigma, \tag{1}$$

where g_* is the norm on $\mathbf{R}^n$ dual to g. From this, the hard part of his result follows: $g \circ \sigma$ is indeed a norm, being the dual of some positively homogeneous function.

The proof of formula (1) is instructive. Von Neumann uses two key ingredients: the fact that any norm g is its own bidual g_{**}, and the fundamental inequality

$$\langle X, Y\rangle \le \langle \sigma(X), \sigma(Y)\rangle. \tag{2}$$

He proves this inequality by considering the optimality conditions for the problem of maximizing $\langle Y, Z\rangle$ as the variable Z varies over the matrix manifold

$$\{UXV : U, V \in \mathbf{U}^n\} = \{Z \in \mathbf{M}^n : \sigma(Z) = \sigma(X)\}. \tag{3}$$

The ideas driving this lovely argument illustrate many of the threads we pursue in this survey: groups of matrix transformations (such as unitary multiplication) and associated canonical forms (such as the singular value decomposition); matrix functions invariant under these transformations (such as unitarily invariant norms) and associated symmetric functions (such as symmetric gauges); convexity and duality arguments (such as the biduality relationship for norms); variational arguments over matrix manifolds (like that defined in equation (3)).

One last important ingredient for us is missing from von Neumann's original development, namely sensitivity analysis or perturbation theory. The crucial tool here is the *subdifferential* [56] of a convex function $f : \mathbf{E} \to \overline{\mathbf{R}}$, where $\mathbf{E}$ is an arbitrary Euclidean space and $\overline{\mathbf{R}} = [-\infty, +\infty]$. The subdifferential at a point $x \in \mathbf{E}$ where $f(x)$ is finite is the set

$$\partial f(x) = \{y \in \mathbf{E} : \langle y, z - x\rangle \le f(z) - f(x) \text{ for all } z \in \mathbf{E}\}.$$

The subdifferential gives sensitivity information via directional derivatives:

$$f'(x; w) = \sup\{\langle y, w\rangle : y \in \partial f(x)\}.$$

A more careful study of the key inequality (2) shows the attractive formula

$$\partial(g \circ \sigma)(X) = \{U(\operatorname{Diag} y)V : U, V \in \mathbf{U}^n,\ y \in \partial g(x),\ X = U(\operatorname{Diag} x)V\} \quad (4)$$

for any symmetric gauge function g [65,67,35].

2. An outline

This article presents a personal view of some of the beautiful mathematics underlying eigenvalue optimization. As an expository mathematical essay, it contains no proofs, algorithms, applications, or computational results. A survey of the broader picture may be found in [43]. Although making frequent use of the language of convex analysis [56,29,6], the writing is, as far as possible, self-contained.

After this outline, which concludes Part I, the survey breaks into three further parts: Part II on convex optimization involving the eigenvalues of symmetric matrices; Part III on applications of variational analysis to eigenvalue perturbation theory and to the stability of polynomials; Part IV on the stability and robust stability of nonsymmetric matrices. A portion of Part III and all of Part IV relies heavily on an extended and ongoing joint project with Jim Burke and Michael Overton. Many of the ideas I sketch in those sections are shared with them; all the errors are mine.

Each of the three main parts further breaks down into sections. I have tried to organize each section around both a theme in optimization and a fundamental, often classical, mathematical result underlying the theme. Most sections include an open question or two that I find interesting.

Part II begins with Davis's characterization of those convex functions of symmetric matrices that are invariant under orthogonal similarity transformations, a close analogue of von Neumann's result. In the search for unifying framework in which to pursue this analogy, we explore two algebraic settings. In the first, we consider the eigenvalues of a symmetric matrix as a special case of the characteristic roots of a hyperbolic polynomial, and we consider the implications of Gårding's theorem on the convexity of the associated hyperbolicity cone. The second framework, better suited for questions about duality, invokes classical semisimple Lie theory. In this setting, we consider the Kostant convexity theorem, a far-reaching generalization of Horn's characterization of the diagonals of symmetric matrices with given eigenvalues. We also follow various analogues of the Chevalley restriction theorem concerning the smoothness of invariant functions of eigenvalues.

Part III begins with a rapid sketch of some of the central ideas we need from contemporary nonsmooth optimization, as developed by Clarke, Ioffe, Mordukhovich, Rockafellar and others. We explore Lidskii's theorem, one of the central results of eigenvalue perturbation theory, from a variational analysis perspective. Turning to polynomials, we relate a fundamental result of Burke and Overton—that the set of stable monics (all of whose zeros have negative real part) is everywhere Clarke regular—to the famous Gauss-Lucas theorem. Nonsmooth calculus then gives an analogous result for "power-stable" polynomials, all of whose zeros lie in the unit disk.

Part IV concerns the stability of nonsymmetric matrices. Either by appealing to classical results of Arnold on manifolds of matrices with fixed Jordan structure, or by applying a nonsmooth chain rule to the previous regularity result for stable polynomials, we arrive at Burke and Overton's characterization of the points of regularity of the cone of stable matrices (all of whose eigenvalues have negative real part). The Jordan structure of a solution to an optimization problem over this cone typically persists under small perturbations to the problem. We end with a discussion of robust stability. We follow two approaches, one via a classical semidefinite characterization of Lyapunov, and the other via pseudospectra: the two approaches can be reconciled by a celebrated result from numerical analysis, the Kreiss matrix theorem. The pseudospectral approach shows considerable computational promise.

PART II: SYMMETRIC MATRICES

3. Convex spectral functions

Twenty years after the appearance of von Neumann's characterization of unitarily invariant matrix norms, Davis [20] proved an exactly analogous result for convex *spectral functions* on the vector space $\mathbf{S}^n$ of n-by-n symmetric matrices, that is, convex functions $F : \mathbf{S}^n \to \overline{\mathbf{R}}$ satisfying

$$F(U^T XU) = F(X) \quad \text{for all } U \in \mathbf{O}^n,\ X \in \mathbf{S}^n,$$

where $\mathbf{O}^n$ denotes the group of orthogonal matrices. The spectral decomposition shows that we can write any such function in the form $F(X) = f(\lambda(X))$, where $\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X)$ are the eigenvalues of X (whence the name "spectral" function). Following von Neumann's argument, we could define $f(x) = F(\operatorname{Diag} x)$ for vectors $x \in \mathbf{R}^n$, and in that case f is a symmetric convex function. Davis observed that this necessary condition is also sufficient.

Theorem 3.1 (Davis, 1957). *The convex spectral functions are exactly the symmetric convex functions of the eigenvalues.*

Indeed, one can parallel von Neumann's duality-based argument closely to prove this result [36], but we defer further discussion of duality approaches to Section 5. A nice summary is [31].

The interior point revolution and Nesterov and Nemirovski's seminal 1994 book [52] shifted focus in convex optimization towards questions of tractability. A particularly compelling approach [52,3,66] considers the *semidefinite-representable* sets, namely those of the form

$$\{x \in \mathbf{R}^m : \mathcal{A}x \in B + L + \mathbf{S}^n_+\}$$

for some positive integer n, linear map $\mathcal{A} : \mathbf{R}^m \to \mathbf{S}^n$, matrix $B \in \mathbf{S}^n$, and linear subspace $L \subset \mathbf{S}^n$, where $\mathbf{S}^n_+$ denotes the cone of positive semidefinite matrices in $\mathbf{S}^n$. (We use the term somewhat loosely: for tractability, we would want a "concrete" representation, and one where the dimension n does not grow too quickly as the dimension m grows. Furthermore, for optimization over a full-dimensional convex set, a semidefinite representation of any convex set with the same boundary usually suffices.) A *semidefinite representable function* $f : \mathbf{R}^m \to \overline{\mathbf{R}}$ is one whose *epigraph* $\{(x,r) : f(x) \le r\}$ is semidefinite-representable. Davis's theorem has a nice analogue in this framework, essentially a rephrasing of [3, Prop 4.2.1].

Theorem 3.2 (Ben-Tal/Nemirovski, 2001). *The semidefinite-representable spectral functions are exactly the symmetric semidefinite-representable functions of the eigenvalues.*

Furthermore, von Neumann's characterization (Theorem 1.1) has a corresponding analogue: the semidefinite-representable unitarily invariant norms are exactly the semidefinite-representable symmetric gauge functions of the singular values (cf. [3, Prop 4.2.2]).

The striking parallels between von Neumann's framework of unitarily invariant norms and Davis's framework of convex spectral functions demand a unified analysis. In the next two sections we consider two perspectives, the first a remarkably elementary but purely primal approach, and the second a more sophisticated, duality-based approach.

4. Hyperbolic polynomials

Any symmetric matrix X has all real eigenvalues. In other words, the zeros of the characteristic polynomial $t \mapsto \det(X - tI)$ are all real. Remarkably, this very simple property drives many convexity results about eigenvalues.

To see this, consider a homogeneous polynomial p of degree n on $\mathbf{R}^m$. We call p *hyperbolic* with respect to a vector $d \in \mathbf{R}^m$ if $p(d) \neq 0$ and the polynomial $t \mapsto p(x - td)$ has all real zeros for any vector $x \in \mathbf{R}^m$ [22]. We call these zeros the *characteristic roots* of x, and denote them $\lambda_1(x) \ge \lambda_2(x) \ge \cdots \ge \lambda_n(x)$. The *hyperbolicity cone* is the set $C_p = \{x : \lambda_n(x) > 0\}$. With this terminology, the determinant is a hyperbolic polynomial on the vector space of symmetric matrices with respect to the identity matrix, the characteristic roots are just the eigenvalues, and the hyperbolicity cone consists of the positive definite matrices.

The hyperbolic polynomials form a rich class. For example (see [2, Fact 2.10]), for any hyperbolic polynomial p as above, the polynomials $p_r(\cdot) = E_r(\lambda(\cdot))$ are all hyperbolic with respect to d, where E_r denotes the rth elementary symmetric function on $\mathbf{R}^n$, namely the sum over all r-fold products of the variables.

Fundamental to questions about convexity is the following wonderfully concise result. For an elementary (although not transparent) proof, see [23].

Theorem 4.1 (Gårding, 1951). *Hyperbolicity cones are convex.*

Building on Gårding's observation, Güler [25] proved (with the above notation) that the function $-\ln(p(\cdot))$ is a self-concordant barrier for the hyperbolicity cone C_p, with barrier parameter n. (See [52,55] for the definition and algorithmic implications of self-concordance.) Güler's result suggests efficient primal algorithms for linearly-constrained optimization problems posed over the hyperbolicity cone.

Güler's barrier is an example of a convex symmetric function of the characteristic roots, namely $-\sum_j \ln \lambda_j(\cdot)$. In fact, by bootstrapping up from Gårding's theorem we obtain a complete generalization of the interesting direction in Davis's characterization (Theorem 3.1). The following result appeared in [2].

Theorem 4.2 (convex functions of characteristic roots). *Symmetric convex functions of the characteristic roots of a hyperbolic polynomial are convex.*

The success of the idea of semidefinite representability now suggests the tempting question of the modeling power of hyperbolic polynomials. In particular, does the characterization of Ben-Tal and Nemirovski (Theorem 3.2) have any extension in this framework? Specifically, *which symmetric functions of* $\lambda(x)$ *have tractable representations?* One potentially useful tool is the representability of *hyperbolic means* [45]: in the above notation, these are functions of the form $p^{1/n}$ on the domain C_p. Sets of the form

$$\{(x,t) \in C_p \times \mathbf{R}_+ : t^n \leq p(x)\}$$

have an easily computable self-concordant barrier with barrier parameter $O(n^2)$. By applying this fact to the hyperbolic polynomials $p_r(\cdot) = E_r(\lambda(\cdot))$, we deduce the representability of constraints on elementary symmetric functions of the characteristic roots of the form $E_r(\lambda(x)) \geq k$.

Can hyperbolic polynomials model more powerfully than semidefinite programming? More precisely, *are all hyperbolicity cones semidefinite representable*? Forms of this question go back to Lax [34], who made the following conjecture.

Conjecture 4.3 (Lax, 1958). If p is a polynomial on $\mathbf{R}^3$, then p is hyperbolic of degree d with respect to the vector $(1,0,0)$ and $p(1,0,0) = 1$ if and only if there exist matrices $B, C \in \mathbf{S}^d$ such that p is given by $p(x,y,z) = \det(xI + yB + zC)$.

This conjecture is indeed true, as follows from recent work of Helton and Vinnikov [27,63]: see [44]. This result implies that any three-dimensional hyperbolicity cone is a *semidefinite slice*: it is isomorphic to an intersection of a positive definite cone with a subspace (and hence in particular is semidefinite representable).

Homogeneous cones (open convex pointed cones for which any element of the cone can be mapped to any other by a linear automorphism of the cone) are a special case of hyperbolicity cones [25]. In this case the representability question was recently answered affirmatively in [17] (see also [21]): any homogeneous cone is a semidefinite slice.

5. Duality, smoothness, and Lie theory

The technical hurdle in extending Ben-Tal and Nemirovski's characterization of semidefinite-representable spectral functions (Theorem 3.2) to the very broad framework of hyperbolic polynomials lies in the lack of any obvious duality theory. At the opposite extreme, the algebraic framework unifying the most successful primal-dual interior-point methods for conic convex programs, namely that of Euclidean Jordan algebras, is rather narrow in scope [26]. In this section we consider an intermediate algebraic framework, covering an interesting range of examples but with a duality theory rich enough to subsume the development in Section 1 around von Neumann's characterization of unitarily invariant norms.

This section (and this section alone) is Lie-theoretic in essence. Rather than assume any familiarity with Lie theory, we simply retrace our steps through the familiar case of symmetric matrices, using the key results as illustrations of the general case. For this beautiful classical development we follow the outline provided by [53, Chap. 5, Sec. 3,4]. More details appear in [40].

We begin with our familiar decomposition of the vector space of n-by-n real matrices with trace zero, $\mathfrak{sl}(n, \mathbf{R})$, into a sum of the subspace of skew-symmetric matrices, $\mathfrak{so}(n)$, and the symmetric matrices with trace zero, $\mathfrak{p}_n$, and note this sum is direct and orthogonal with respect to the bilinear form $(X, Z) = \mathrm{tr}(XZ)$: we write $\mathfrak{sl}(n, \mathbf{R}) = \mathfrak{so}(n) \oplus \mathfrak{p}_n$. This is a special case of a *Cartan decomposition* $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, where $\mathfrak{g}$ is a real semisimple Lie algebra, $\mathfrak{k}$ is a subalgebra of $\mathfrak{g}$ tangent to a maximal compact subgroup K of any connected semisimple Lie group G with Lie algebra $\mathfrak{g}$ (for example, the adjoint group for $\mathfrak{g}$) and a complementary subspace $\mathfrak{p}$, direct with respect to the *Killing form* $(\cdot, \cdot)$. In our example, G is the special linear group $SL(n, \mathbf{R})$ of matrices with determinant one (and $\mathfrak{sl}(n, \mathbf{R})$ is the tangent space to G at the identity), and K is the subgroup $SO(n)$ of orthogonal matrices.

We say that a subalgebra $\mathfrak{a}$ of $\mathfrak{g}$ is $\mathbf{R}$-*diagonalizable* if $\mathfrak{g}$ has a basis with respect to which each element of $\mathfrak{a}$ (regarded as an operator on $\mathfrak{p}$ via the natural adjoint representation) is represented as a diagonal matrix. In our example, the diagonal matrices with trace zero, $\mathfrak{a}_n$, form a *maximal* $\mathbf{R}$-diagonalizable subalgebra.

The group K acts on the subspace $\mathfrak{p}$ via the adjoint representation: in our example, this action is $X \mapsto U \cdot X = UXU^T$ for $X \in \mathfrak{p}_n$ and $U \in SO(n)$. Now consider the normalizer and centralizer corresponding to any fixed maximal $\mathbf{R}$-diagonalizable subalgebra $\mathfrak{a}$ of $\mathfrak{g}$, defined respectively by

$$N = \{k \in K : k \cdot \mathfrak{a} = \mathfrak{a}\}$$
$$Z = \{k \in K : k \cdot x = x \text{ for all } x \in \mathfrak{a}\}.$$

The *Weyl group* is the quotient group $W = N/Z$. We can think of W, loosely, as a group of linear transformations on $\mathfrak{a}$ (which is a Euclidean space with inner product the Killing form). At this point in the development, a miracle occurs (fundamental to the classification of semisimple Lie algebras): the Weyl group *W is always a finite reflection group.* In our example, elementary calculations show that N is the group of matrices with determinant one and having exactly one nonzero entry of ± 1 in each row and column, and Z is the subgroup of diagonal matrices in N, so we can consider W as acting on $\mathfrak{a}_n$ by permuting diagonal entries: in other words, the Weyl group is isomorphic to the group of n-by-n permutation matrices $\mathbf{P}^n$.

We are now ready for the driving result of this section [32].

Theorem 5.1 (Kostant, 1973). *Suppose $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ is a Cartan decomposition of a real semisimple Lie algebra corresponding to a maximal compact subgroup K, and suppose $\mathfrak{a} \subset \mathfrak{p}$ is a maximal $\mathbf{R}$-diagonalizable subalgebra. Then the orthogonal projection (under the Killing form) onto $\mathfrak{a}$ of any orbit of K acting on $\mathfrak{p}$ is the convex hull of an orbit of the Weyl group.*

In our example, the orthogonal projection is the map that sets the off-diagonal entries of a trace-zero symmetric matrix to zero, so we deduce the following result [30]. (The trace-zero assumption below is easily dispensable.)

Corollary 5.2 (Horn, 1954). *The diagonals of all symmetric matrices similar to a given trace-zero symmetric matrix X form a polytope. This polytope is the convex hull of all vectors with entries permutations of the eigenvalues of X.*

What is the connection with the characterizations of von Neumann and Davis (Theorems 1.1 and 3.1)? One appealing way to restate the Davis characterization is that *a spectral function is convex if and only if its restriction to the subspace of diagonal matrices is convex.* In this form, it is not hard to deduce the characterization from Horn's result. Von Neumann's characterization has an analogous restatement: *a unitarily invariant function on $\mathbf{M}^n$ is a norm if and only if its restriction to the Euclidean subspace of real diagonal matrices, $\mathbf{D}^n$, is a norm.* Using Kostant's theorem, we obtain the following broad unification [40].

Corollary 5.3 (invariant convex functions). *In the framework of Kostant's theorem, a K-invariant function $h : \mathfrak{p} \to \overline{\mathbf{R}}$ is convex if and only if its restriction to the subspace $\mathfrak{a}$ is convex.*

Von Neumann's characterization follows by considering the standard Cartan decomposition of the Lie algebra $\mathfrak{su}(p,q)$ instead of $\mathfrak{sl}(n,\mathbf{R})$.

Given the results of Ben-Tal and Nemirovski on semidefinite representability (Theorem 3.2 and the subsequent comment), the following conjecture seems promising.

Conjecture 5.4 (semidefinite representability of invariant functions). With the framework of Kostant's theorem, a K-invariant function $h : \mathfrak{p} \to \overline{\mathbf{R}}$ is semidefinite-representable if and only if $h|_{\mathfrak{a}}$ is semidefinite-representable.

So far in this section, the rather sophisticated algebraic framework has succeeded in paralleling the characterizations of von Neumann and Davis, results also captured by the much more spartan framework of hyperbolic polynomials sketched in the previous section. In apparent contrast with the latter framework, however, semisimple Lie theory is rich in duality theory, allowing us to imitate von Neumann's entire development. An analogous development for Euclidean Jordan algebras appeared recently in [48].

Consider von Neumann's beautiful duality formula (1), for example. An equivalent statement is that any unitarily invariant norm $\|\cdot\|$ on $\mathbf{M}^n$ satisfies

$$\Big(\|\cdot\|\Big|_{\mathbf{D}^n}\Big)_* = \Big(\|\cdot\|_*\Big)\Big|_{\mathbf{D}^n},$$

or in words, "the dual of the restriction is the restriction of the dual". For inhomogeneous convex functions f on a Euclidean space, the *Fenchel conjugate* function

$$f^*(y) = \sup_x \{\langle x, y\rangle - f(x)\}$$

replaces the dual norm as the key duality tool [56]. An analogous result to von Neumann's holds for the Fenchel dual of a spectral function [36], and the two can be broadly unified by the following result [40], using Kostant's theory.

Theorem 5.5 (conjugates of invariant functions). *In the framework of Kostant's theorem, any K-invariant function $h : \mathfrak{p} \to \overline{\mathbf{R}}$ satisfies $(h|_{\mathfrak{a}})^* = h^*|_{\mathfrak{a}}$.*

Our algebraic framework also nicely covers subdifferential formulas like (4) and its analogue for spectral functions [36]:

$$\partial(f\circ\lambda)(X) = \{U(\operatorname{Diag} y)U^T : U \in \mathbf{O}^n,\ y \in \partial f(x),\ X = U(\operatorname{Diag} x)U^T\} \quad (5)$$

for any symmetric convex function $f : \mathbf{R}^n \to \overline{\mathbf{R}}$. The following result [40] again unifies the two threads.

Theorem 5.6 (subdifferentials of invariant functions). *In the framework of Kostant's theorem, a K-invariant convex function $h : \mathfrak{p} \to \overline{\mathbf{R}}$ at a point $x \in \mathfrak{p}$ has subdifferential*

$$\partial h(x) = \{k\cdot v : k \in K,\ u, v \in \mathfrak{a},\ x = k\cdot u,\ v \in \partial h|_{\mathfrak{a}}(u)\}.$$

Summarizing the general direction of these results, we see that the interesting convex-analytic properties of an invariant function on the "large" underlying space $\mathfrak{p}$ lift precisely from the corresponding properties on the "small" subspace $\mathfrak{a}$. These kind of "lifting" results are reminiscent of another famous piece of classical Lie theory, the Chevalley restriction theorem.

Theorem 5.7 (Chevalley). *With the framework of Kostant's theorem, any K-invariant function $h : \mathfrak{p} \to \overline{\mathbf{R}}$ is a polynomial of degree k if and only if the restriction $h|_{\mathfrak{a}}$ is a polynomial of degree k.*

For example, the spectral polynomials of degree k are exactly the degree-k symmetric polynomials of the eigenvalues. (In fact, the Chevalley result states more: the restriction map $h \mapsto h|_{\mathfrak{a}}$ is an isomorphism of algebras.)

The Chevalley restriction theorem is an example of how smoothness lifts to the space $\mathfrak{p}$ from the subspace $\mathfrak{a}$. Analogues exist for $C^{(\infty)}$ and analytic functions—see [19]. At the opposite extreme, at least for spectral functions, we have the following result (see [37,46]).

Theorem 5.8 (lifting smoothness). *Consider an open symmetric set $\Omega \subset \mathbf{R}^n$ and a symmetric function $f : \Omega \to \mathbf{R}$. Then, on the space $\mathbf{S}^n$, the function $f \circ \lambda$ is $C^{(1)}$ (or respectively $C^{(2)}$) if and only if f is $C^{(1)}$ (or respectively $C^{(2)}$).*

(A set is *symmetric* if it is invariant under coordinate permutations.) We can even write down straightforward formulas for the first and second derivatives of $f \circ \lambda$ in terms of those for f. This pattern suggests several natural conjectures. Does the analogous result hold for $C^{(k)}$ functions with $k \geq 3$? Do these results hold in our Lie-theoretic framework above? What about for symmetric functions of the characteristic roots of a hyperbolic polynomial? (It is not too hard to show that differentiability does lift in the hyperbolic polynomial setting.)

One reason for studying $C^{(3)}$ functions in optimization is as a helpful framework for investigating self-concordance. For example, the symmetric function $f : \mathbf{R}^n_{++} \to \mathbf{R}$ defined by $f(x) = -\sum_i \ln(x_i)$ is a self-concordant barrier with parameter n, and so is $f \circ \lambda \equiv -\ln\det$. Does self-concordance always lift in this manner [62]? A simple example supporting such a conjecture (and specializing a somewhat more general result) is the following [51]: if $f : (-1, 1) \to \mathbf{R}$ is a self-concordant barrier with parameter k, then so is the function $f(\|\cdot\|_2)$.

PART III: SOME VARIATIONAL ANALYSIS

6. Nonsmooth optimization

We have seen how classical optimization theory—the variational analysis of smooth or convex functions—applies well to spectral functions of symmetric matrices. In particular, derivatives or subdifferentials are easy to calculate in terms of the underlying symmetric function. Over the last thirty years, great progress has been made in our understanding of optimization problems that are neither smooth nor convex, driven by Clarke, Ioffe, Mordukhovich, Rockafellar and others. In this part, we apply this powerful theory to eigenvalue perturbation analysis and to the stability of polynomials.

We begin in this section with a quick summary of the relevant variational analysis. The two texts [18] and [57] are excellent general sources.

In this section we fix a Euclidean space $\mathbf{E}$. We call a vector $z \in E$ a *regular subgradient* of a function $f : \mathbf{E} \to \overline{\mathbf{R}}$ at a vector $x \in \mathbf{E}$ if $f(x)$ is finite and

$$f(y) - f(x) \geq \langle z, y - x \rangle + o(|y - x|) \quad \text{as } y \to x;$$

the *regular subdifferential* $\hat{\partial} f(x)$ is the set of such z. This notion is a one-sided version of the classical gradient, as well as correlating with the classical subdifferential of a convex function. We next define "limiting" subgradients by "closing" the regular subdifferential mapping $\hat{\partial} f$. Specifically, we call z a *subgradient* of f at x if there exists a sequence of vectors $x_r \to x$ and a sequence of regular subgradients $y_r \in \hat{\partial} f(x_r)$ such that $f(x_r) \to f(x)$ and $y_r \to y$; the *subdifferential* $\partial f(x)$ is the set of such z.

An attractive feature of variational analysis is that we can study a set $S \subset \mathbf{E}$ via its *indicator function* δ_S, which takes the value 0 on S and $+\infty$ outside S. The *regular normal cone* $\hat{N}_S(x)$ is just the regular subdifferential $\hat{\partial}\delta_S(x)$, and the normal cone $N_S(x)$ is the subdifferential $\partial\delta_S(x)$. We call S *(Clarke) regular* at $x \in S$ if S is locally closed at x and the normal and regular normal cones agree. Two examples are worth keeping in mind. First, a closed convex set S is regular at any element $x \in S$, and $N_S(x)$ is then just the usual normal cone in the sense of convex analysis. Secondly, if the set $\mathcal{M} \subset \mathbf{E}$ is a *manifold*, which for our purposes just means $\mathcal{M}$ can be expressed locally around any element as the zero set of some $C^{(2)}$ functions with linearly independent gradients, then $N_{\mathcal{M}}(\cdot)$ is just the usual normal space in the sense of elementary differential geometry.

We call a function $f : \mathbf{E} \to \overline{\mathbf{R}}$ *(subdifferentially) regular* at $x \in \mathbf{E}$ if $f(x)$ is finite and the epigraph of f is regular at the point $(x, f(x))$: in this case, every subgradient is regular ($\partial f(x) = \hat{\partial} f(x)$). In particular, convex functions are regular at every point where they take finite value, the subdifferential agreeing with the same notion from convex analysis. The subdifferential of a $C^{(1)}$ function just consists of the gradient.

The most basic nontrivial result of differential calculus is the mean value theorem, and variational analysis provides a useful generalization for nonsmooth functions (see for example [57]).

Theorem 6.1 (mean value). *If the function $h : \mathbf{E} \to \overline{\mathbf{R}}$ is locally Lipschitz on an open convex set containing the line segment $[x, y] \subset \mathbf{E}$, then there exists a point $w \in [x, y]$ and a subgradient $z \in \partial h(w) \cup -\partial(-h)(w)$ satisfying the equation $h(y) - h(x) = \langle z, y - x \rangle$.*

Crucial to the uses of differential calculus on all but the simplest of functions is the chain rule. From several nonsmooth generalizations, we need the following one for our development.

Theorem 6.2 (chain rule). *Suppose the function $\Phi : \mathbf{R}^m \to \mathbf{E}$ is $C^{(1)}$ at the point $v \in \mathbf{R}^m$, the set $S \subset \mathbf{E}$ is regular at the point $\Phi(v)$, and the condition $N_S(\Phi(v)) \cap \ker(\nabla\Phi(v))^* = \{0\}$ holds. Then the inverse image*

$$\Phi^{-1}(S) = \{u \in \mathbf{R}^m : \Phi(u) \in S\}$$

is regular at the point v, with normal cone given by

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

We end this section with some less standard material. From a variational perspective, much of the fascinating behaviour of eigenvalues (and indeed many other concrete functions) arises from a delicate interplay between smooth and nonsmooth properties. For the purposes of sensitivity analysis, this interplay is well captured by the idea of partial smoothness [42]. A set $S \subset \mathbf{E}$ is *partly smooth* relative to an *active* manifold $\mathcal{M} \subset S$ if, as the point x varies in $\mathcal{M}$, the set S is always regular at x with normal cone $N_S(x)$ varying continuously and spanning $N_{\mathcal{M}}(x)$. The continuity of $N_S(\cdot)$ on $\mathcal{M}$ amounts to the condition that for any convergent sequence in $\mathcal{M}$, say $x_r \to x$, and any normal vector $y \in N_S(x)$, there exists a sequence of normal vectors $y_r \in N_S(x_r)$ satisfying $y_r \to y$. The "spanning" condition ensures a kind of "sharpness" of the set S around the manifold $\mathcal{M}$. The following result, essentially due to [54], provides a nice example.

Theorem 6.3 (partial smoothness of semidefinite cone). *Given any nonnegative integers $k \leq n \geq 1$, the semidefinite cone $\mathbf{S}^n_+$ is partly smooth relative to the subset of matrices of rank k.*

Partial smoothness captures much of the essence of active set ideas in optimization. Under reasonable conditions, for example, if the optimal solution to a smooth minimization problem over a partly smooth feasible region lies on the active manifold, then the solution will remain on the manifold under small perturbations to the problem. Many operations on sets preserve partial smoothness. For example, an exact analogue of the chain rule (Theorem 6.2) describes how taking inverse images preserves partial smoothness [42]. This result, in combination with the partial smoothness of the semidefinite cone, explains why the rank of an optimal solution of a semidefinite program is typically stable under small perturbations to the problem.

7. Eigenvalue perturbation theory

At the heart of the perturbation theory of symmetric matrices stands the following result [47].

Theorem 7.1 (Lidskii, 1950). *For any given matrices $X, Y \in \mathbf{S}^n$, the vector $\lambda(Y) - \lambda(X)$ lies in the convex hull of the set $\{P\lambda(Y - X) : P \in \mathbf{P}^n\}$.*

This result and its various proofs, for example, form the unifying theme of the book [5].

Via a separating hyperplane, we arrive at the equivalent version

$$w^T(\lambda(Y) - \lambda(X)) \leq \bar{w}^T \lambda(Y - X) \quad \text{for all } w \in \mathbf{R}^n,$$

where $\bar{w} \in \mathbf{R}^n$ has the same components as w, arranged in decreasing order. In this form, it becomes tempting to apply the nonsmooth mean value result (Theorem 6.1) to the function $w^T\lambda(\cdot)$, but this function is not convex, so we first need a formula like (5) for *nonconvex* spectral functions. Happily, the formula extends unchanged [39].

Theorem 7.2 (subdifferentials of spectral functions). *For any symmetric function $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ and any matrix $X \in \mathbf{S}^n$, we have*

$$\partial(f \circ \lambda)(X) = \{U(\operatorname{Diag} y)U^T : U \in \mathbf{O}^n,\ y \in \partial f(x),\ X = U(\operatorname{Diag} x)U^T\}.$$

The corresponding result for the regular subdifferential also holds. The function $f \circ \lambda$ is regular at X if and only if f is regular at $\lambda(X)$.

By applying this result to the symmetric function $f(x) = w^T \bar{x}$, Lidskii's theorem quickly follows [38].

To what extent does Lidskii's theorem extend to the frameworks of Section 4 (Hyperbolic polynomials) and Section 5 (Duality, smoothness, and Lie theory)? Theorem 5.6 (subdifferentials of invariant functions) does indeed extend at least to nonconvex, locally Lipschitz functions, and this suffices to generalize Lidskii's theorem to the Lie-theoretic framework of Section 5 [59], since the function $x \mapsto w^T \bar{x}$ is indeed locally Lipschitz. This result, originally due to Berezin and Gel'fand [4], in fact predated Lidskii's publication of his special case [5].

This leaves two outstanding questions. First, a full unification of our characterizations of subdifferentials, Theorems 7.2 and 5.6, seems an obvious conjecture. Secondly, does a version of Lidskii's theorem hold for the characteristic roots of a hyperbolic polynomial? Weaker results do hold: for example, in the notation of Section 4 (Hyperbolic polynomials), we have the following result of Lax [34]:

$$\lambda(y - x) \geq 0 \quad \Rightarrow \quad \lambda(y) - \lambda(x) \geq 0.$$

8. Stable polynomials

Eigenvalues of matrices are intimately related to zeros of polynomials, via the idea of the the characteristic polynomial. Before moving towards eigenvalue optimization for nonsymmetric matrices in the last part of this paper, we therefore pause to consider variational properties for functions of the zeros of polynomials.

First let us fix some notation. Denote the space of complex polynomials of degree no more than n by $\mathcal{P}^n$. The natural basis elements for $\mathcal{P}^n$ are the polynomials e^k defined by $e^k(z) = z^k$ for $k = 0, 1, \ldots, n$ (together with ie^k), and we can view $\mathcal{P}^n$ as a Euclidean space if we define this basis as orthonormal. Finally, define the *stable monic set* by

$$\mathcal{P}^n_{\mathrm{st}} = \{p \in \mathcal{P}^n : z^{n+1} + p(z) = 0 \;\Rightarrow\; \operatorname{Re} z \leq 0\}.$$

The stable monic set $\mathcal{P}^n_{\mathrm{st}}$ is closed, because the set of zeros of a monic polynomial behave continuously. Surprisingly, however, the good behaviour of this set is more dramatic, from a variational perspective. Fundamentally, this good behaviour is a consequence of central results in [16], encapsulated in the following result.

Theorem 8.1 (Burke-Overton, 2001). *The stable monic set $\mathcal{P}^n_{\mathrm{st}}$ is everywhere regular.*

Proving this result in the style of [16] requires calculating the normal cone at an arbitrary element of $\mathcal{P}^n_{\mathrm{st}}$. This calculation is rather involved, needing a systematic framework for breaking polynomials into factors (the "local factorization lemma" [16, Lem. 1.4]), as well as the following fundamental calculation of the normal cone to the stable monic set at zero:

$$N_{\mathcal{P}^n_{\mathrm{st}}}(0) = \Big\{ \sum_{k=0}^{n} \mu_k e^k : 0 \geq \mu_n \in \mathbf{R},\ \operatorname{Re} \mu_{n-1} \leq 0 \Big\}. \qquad (6)$$

Some of the intricacies of the original approach are eased by an inductive use of the following famous result about polynomials (for which a good reference is [50]).

Theorem 8.2 (Gauss-Lucas, 1830). *All critical points of a nonconstant polynomial lie in the convex hull of the set of zeros of the polynomial.*

(See [14] for this approach.)

Clarke regularity is in fact just the first part of a powerful structural analysis of the stable monic set. Given any list of natural numbers $\mathcal{L}$, let $\mathcal{P}^n_{\mathcal{L}}$ denote the set of polynomials $p \in \mathcal{P}^n$ for which the purely imaginary zeros of the monic polynomial $z^{n+1} + p(z)$, when listed in order of decreasing imaginary part, have multiplicities given by the list $\mathcal{L}$. One can check that $\mathcal{P}^n_{\mathcal{L}}$ is a manifold, and in fact we arrive at the following result (cf. [42, Ex. 3.7]).

Theorem 8.3 (partial smoothness of stable monics). *For any list $\mathcal{L}$, the stable monic set $\mathcal{P}^n_{\mathrm{st}}$ is partly smooth relative to the subset $\mathcal{P}^n_{\mathrm{st}} \cap \mathcal{P}^n_{\mathcal{L}}$.*

In words, the set of stable monics is partly smooth relative to any subset of those monics having imaginary zeros with some fixed pattern of multiplicities. Using this result, we could now apply the partly smooth analogue of the chain rule (Theorem 6.2) to investigate stability problems of the form

$$\inf\{c^T u : u \in \mathbf{R}^m,\ \Phi(u) \in \mathcal{P}^n_{\mathrm{st}}\},$$

for some given cost vector $c \in \mathbf{R}^m$ and smooth map $\Phi : \mathbf{R}^m \to \mathcal{P}^n$. Partial smoothness guarantees that the pattern of multiplicities of the imaginary zeros of $\Phi(u)$ for the optimal u typically persists under small perturbations to the problem.

Suppose, instead of the stable monic set, we are interested in the *power-stable* monics:

$$\mathcal{P}^n_{\mathrm{pst}} = \{q \in \mathcal{P}^n : z^{n+1} + q(z) = 0 \ \Rightarrow\ |z| \le 1\}. \tag{7}$$

A nice application of the chain rule allows us to use our existing analysis of the stable monic set to deduce corresponding results for the power-stable case.

To see how to accomplish this, we first make a small technical restriction, easy to circumvent later: we define a dense open subset of the space $\mathcal{P}^n$ by

$$\hat{\mathcal{P}}^n = \{q \in \mathcal{P}^n : 1 + q(1) \neq 0\},$$

and in place of the set of all power-stable monics, we consider instead the subset $\hat{\mathcal{P}}^n_{\mathrm{pst}} = \hat{\mathcal{P}}^n \cap \mathcal{P}^n_{\mathrm{pst}}$. Now notice that the transformation of the complex plane $w \mapsto (w+1)/(w-1)$ maps the open right halfplane (minus the point 1) onto the complement of the closed unit disk. Hence a polynomial $q \in \hat{\mathcal{P}}^n$ lies in $\mathcal{P}^n_{\mathrm{pst}}$ if and only if

$$w \neq 1 \text{ and } \left(\frac{w+1}{w-1}\right)^{n+1} + q\left(\frac{w+1}{w-1}\right) = 0 \quad \Rightarrow \quad \operatorname{Re} w \le 0.$$

To capture this property using the chain rule, define a map $\Phi : \hat{\mathcal{P}}^n \to \mathcal{P}^n$ by

$$\Phi(q)(w) = \frac{1}{1+q(1)}\left[(w+1)^{n+1} + (w-1)^{n+1} q\left(\frac{w+1}{w-1}\right)\right] - w^{n+1}, \quad \text{for } w \neq 1$$

(extended by continuity at $w = 1$). Then we have

$$\hat{\mathcal{P}}^n_{\text{pst}} \quad = \quad \{q \in \hat{\mathcal{P}}^n : w^{n+1} + \Phi(q)(w) = 0 \ \Rightarrow\ \operatorname{Re} w \leq 0\} \quad = \quad \Phi^{-1}(\mathcal{P}^n_{\text{st}}).$$

Elementary calculations show the map Φ is smooth, with everywhere invertible Jacobian, so the chain rule (Theorem 6.2) now demonstrates that the power-stable monic set $\hat{\mathcal{P}}^n_{\text{pst}}$ is everywhere Clarke regular. We could, in principle, compute the normal cone using formulas for the normal cone to the stable monic set.

The power of Burke and Overton's concise observation that the stable monic set is everywhere Clarke regular (Theorem 8.1) suggests at least two lines of questions. First, the elegance of the result is not matched by existing proofs: even with the use of the Gauss-Lucas theorem, extended elementary calculations seem necessary. Is there, perhaps, a broader framework in which the fact of regularity is more transparent? Secondly, are more powerful results true? For example, *prox-regularity* of sets is a stronger property than Clarke regularity, in particular implying the existence of single-valued projections onto the set from points nearby. Is the stable monic set prox-regular?

PART IV: STABILITY OF NONSYMMETRIC MATRICES

9. Stable matrices

Eigenvalues of nonsymmetric matrices show much more unruly behaviour than those of symmetric matrices. Lidskii's theorem (7.1) shows the eigenvalue map for symmetric matrices $\lambda : \mathbf{S}^n \to \mathbf{R}^n$ is Lipschitz, and Davis's theorem (3.1) demonstrates its elegant convexity properties. By contrast, the eigenvalues of the matrix $\begin{pmatrix} 0 & t \\ 1 & 0 \end{pmatrix}$, for example, are not Lipschitz in the variable t. Furthermore, unlike in the symmetric case, the set of complex n-by-n *stable* matrices $\mathbf{M}^n_{\text{st}}$, by which we mean (in not entirely standard fashion) those matrices having no eigenvalues in the open right halfplane, is not convex. For example, both the matrices $\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}$ are stable, whereas their midpoint $\begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix}$ is not.

Nonetheless, the set of stable matrices $\mathbf{M}^n_{\text{st}}$ has some attractive properties from a variational perspective. It is certainly closed, because the set of eigenvalues of a matrix behaves continuously. In this section we follow Burke and Overton's generalization of their result on the regularity of the stable monic set $\mathcal{P}^n_{\text{st}}$ (Theorem 8.1) to a more interesting result about $\mathbf{M}^n_{\text{st}}$. The matrix case really is more general, because the zeros of the monic polynomial $e^n - \sum_{j=1}^{n-1} a_j e^j$ are exactly the eigenvalues of the companion matrix

$$\begin{bmatrix} a_{n-1} & 0 & \cdots & 0 \\ a_{n-2} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_0 & 0 & \cdots & 0 \end{bmatrix} + J, \quad \text{for the Jordan block} \quad J = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Matrices are also the more interesting case, because of their more immediate modeling powers: $A \notin \mathbf{M}^n_{\mathrm{st}}$ exactly when trajectories of the dynamical system $\dot{x} = Ax$ have exponential growth.

For nonsymmetric matrices, the *algebraic* multiplicity of an eigenvalue (its multiplicity as a zero of the characteristic polynomial) may differ from its *geometric* multiplicity (the dimension of the corresponding eigenspace). Indeed, eigenvalues "typically" have geometric multiplicity one, in which case we call them *nonderogatory.* This observation is made precise by the following result from celebrated work of Arnold [1].

Theorem 9.1 (Arnold, 1971). *The space of square matrices $\mathbf{M}^n$ stratifies into manifolds corresponding to fixed Jordan structure. Among those manifolds corresponding to a fixed Jordan structure for all but one particular eigenvalue, the manifold corresponding to that eigenvalue being nonderogatory has strictly largest dimension.*

The following remarkable result shows that, like the stable monic set, the set of stable matrices is surprisingly well-behaved near typical elements [15].

Theorem 9.2 (Burke-Overton, 2001). *The set of stable matrices $\mathbf{M}^n_{\mathrm{st}}$ is regular at a particular matrix exactly when each imaginary eigenvalue of the matrix is nonderogatory.*

The more powerful part of this result, that nonderogatory imaginary eigenvalues implies regularity, was originally proved using further work of Arnold discussed below. Before returning to this approach, consider first an attack very much in the spirit of this article, using the results on polynomials from the previous section and the chain rule. To make life simple, we consider the result for a *nonderogatory* matrix (a matrix all of whose eigenvalues are nonderogatory).

We consider first the map $\Phi : \mathbf{M}^n \to \mathcal{P}^{n-1}$ defined via the characteristic polynomial:

$$\Phi(X)(z) = \det(zI - X) - z^n \quad (z \in \mathbf{C}). \tag{8}$$

Assuming the matrix X is nonderogatory, the local factorization lemma [16, Lem. 1.4] shows that the Jacobian map $\nabla\Phi(X)$ is surjective. But the set of stable matrices $\mathbf{M}^n_{\mathrm{st}}$ is just the inverse image $\Phi^{-1}(\mathcal{P}^{n-1}_{\mathrm{st}})$, and regularity now follows by the chain rule (Theorem 6.2).

The chain rule also gives a formula for the normal cone at the matrix X. For example, from formula (6) we can deduce the normal cone at the Jordan block J:

$$N_{\mathbf{M}^n_{\mathrm{st}}}(J) = \Big\{ \sum_{j=0}^{n} \theta_j (J^j)^* : \theta_0 \in \mathbf{R}_+, \ \operatorname{Re}\theta_1 \geq 0 \Big\}. \tag{9}$$

Burke and Overton's work [15] leads to an analogous formula for the normal cone at any stable matrix having all its imaginary eigenvalues nonderogatory. They also present some theory for the general case, although a complete result remains an open question.

In a similar vein, the partly smooth analogue of the chain rule [42] allows us to deduce from Theorem 8.3 (partial smoothness of stable monics) that $\mathbf{M}^n_{\mathrm{st}}$ is partly smooth, at least in the nonderogatory case. The complete result is as follows [42]. By analogy with Section 8 (Stable polynomials), given a list of natural numbers $\mathcal{L}$, let $\mathbf{M}^n_{\mathcal{L}}$ denote the set of matrices in $\mathbf{M}^n$ all of whose purely imaginary eigenvalues are nonderogatory, and, when listed in order of decreasing imaginary part, have multiplicities given by the list $\mathcal{L}$. Arnold's results show $\mathbf{M}^n_{\mathcal{L}}$ is a manifold.

Theorem 9.3 (partial smoothness of stable matrices). *For any list $\mathcal{L}$, the set of stable matrices $\mathbf{M}^n_{\mathrm{st}}$ is partly smooth relative to the subset $\mathbf{M}^n_{\mathrm{st}} \cap \mathbf{M}^n_{\mathcal{L}}$.*

Solutions of "semistable programming" problems of the form

$$\inf\{f(u) : u \in \mathbf{R}^m,\ \Psi(u) \in \mathbf{M}^n_{\mathrm{st}}\},$$

for some smooth cost function $f : \mathbf{R}^m \to \mathbf{R}$ and smooth constraint map $\Psi : \mathbf{R}^m \to \mathbf{M}^n$, have an interesting structural stability property. Specifically, for a typical critical point $\bar{u}$, the Jordan structure of the matrix $\Psi(\bar{u})$ persists under small perturbations to the problem [10,42]. The reason, fundamentally, is Theorem 9.3.

The original approaches to Burke and Overton's characterization of regular points in the set of stable matrices (Theorem 9.2) and to the partial smoothness of this set (Theorem 9.3) depended much more explicitly on Arnold's results [1]. A very special case suffices to illustrate the essential idea.

Consider the orbit of the Jordan block $J \in \mathbf{M}^n$ under similarity transformations:

$$\{PJP^{-1} : P \in \mathbf{M}^n \text{ is invertible}\}.$$

The core of Arnold's technique is the following result.

Lemma 9.4 (Arnold, 1971). *The orbit of the Jordan block $J \in \mathbf{M}^n$ under similarity transformations is a manifold with normal space N at J spanned by the matrices $I, J^*, (J^2)^*, \ldots, (J^{n-1})^*$. Any matrix $X \in \mathbf{M}^n$ close to J is similar to a unique matrix in $J + N$, depending smoothly on X.*

Building on this idea, Arnold was able to construct on a neighbourhood of *any* matrix, canonical forms that, unlike the Jordan form, vary *smoothly*. Applying his results in the simple special case of nonderogatory matrices gives the tools we need for this section, the form of formula (9) being an obvious example.

"Power-stable" matrices are also of great interest: if we define the set of matrices $\mathbf{M}^n_{\mathrm{pst}} \subset \mathbf{M}^n$ to consist of those matrices with all eigenvalues in the closed unit disk, then $A \notin \mathbf{M}^n_{\mathrm{pst}}$ exactly when the iteration $x_{k+1} = Ax_k$ exhibits exponential growth. By analogy with Burke and Overton's characterization of regularity for the stable matrices $\mathbf{M}^n_{\mathrm{st}}$, it turns out that the set $\mathbf{M}^n_{\mathrm{pst}}$ is regular at any element all of whose eigenvalues of modulus one are nonderogatory, and for example,

$$N_{\mathbf{M}^n_{\mathrm{pst}}}(I+J) = \Big\{ \sum_{j=0}^{n} \theta_j (J^j)^* : \theta_0 \in \mathbf{R}_+,\ \mathrm{Re}\,\theta_1 \geq -\theta_0 \Big\}.$$

Two approaches suggest themselves, again in the spirit of this article, the first applying the chain rule (Theorem 6.2) to the inverse image $\Phi^{-1}(\mathcal{P}^n_{\mathrm{pst}})$, where Φ is the characteristic polynomial map (8) and $\mathcal{P}^n_{\mathrm{pst}}$ is the set of power-stable monics (7).

The second, more direct, matrix-based approach works by analogy with Section 8 (Stable polynomials). For technical reasons, we first consider the subset

$$\mathbf{M}^n_1 = \{W \in \mathbf{M}^n : W - I \text{ invertible}\},$$

and define a map $\Phi : \mathbf{M}^n_1 \to \mathbf{M}^n$ by $\Phi(W) = (W-I)^{-1}(W+I)$. For a matrix $W \in \mathbf{M}^n_1$ it is easy to check $W \in \mathbf{M}^n_{\mathrm{pst}}$ if and only if $\Phi(W) \in \mathbf{M}^n_{\mathrm{st}}$, so again we can apply the chain rule.

10. Robust stability

In the previous section we remarked that a matrix $A \in \mathbf{M}^n$ has an eigenvalue outside the closed unit disk exactly when the iteration $x_{k+1} = Ax_k$ exhibits exponential growth. These equivalent properties can also be checked via semidefinite programming, using the following well-known result. The original, continuous-time case is due to Lyapunov [49]. As usual, $\|\cdot\|$ denotes the operator two-norm on $\mathbf{M}^n$ (which is exactly the largest singular value $\sigma_1(\cdot)$). For two matrices G and H in the space $\mathbf{H}^n$ of n-by-n Hermitian matrices, we write $H \succ G$ to mean $H - G$ is positive definite.

Theorem 10.1 (Lyapunov, 1893). *For any matrix $A \in \mathbf{M}^n$, the following properties are equivalent:*

(i) *All eigenvalues of A lie in the open unit disk;*
(ii) $\|A^k\| \to 0$ *exponentially as $k \to \infty$;*
(iii) *There exists a matrix $H \succ 0$ in $\mathbf{H}^n$ such that $H \succ A^*HA$.*

Attractively simple as it is, this result is often deceptive for two reasons, which, as we see in the next result, are related. The first difficulty is one of *robustness*: as we saw in the last section, the dependence of the eigenvalues on A is not Lipschitz. Consequently, when a stable matrix A is close to a matrix with a multiple eigenvalue, unexpectedly small perturbations to A can destroy stability. Secondly, there is the difficulty of *transient growth*: even if $\|A^k\| \to 0$ asymptotically, intermediate values of $\|A^k\|$ may be very large [60,61].

These difficulties can be quantified by the following classical result in numerical analysis [33]. We need some more notation. First, consider the *Lyapunov condition factor* for a matrix $A \in \mathbf{M}^n$, defined by

$$L(A) = \inf\{\gamma > 1 : \gamma I \succ H \succ \gamma^{-1} I,\ H \succ A^* H A,\ H \in \mathbf{H}^n\}.$$

Condition (iii) in Lyapunov's theorem above says exactly that $L(A)$ is finite. We compare this condition factor with the *power bound*

$$P(A) = \sup\{\|A^k\| : k = 1, 2, \ldots\},$$

as well as with a third quantity defined in terms of pseudospectra [24]. For real $\epsilon \geq 0$, the *ϵ-pseudospectrum* of A is the set

$$\begin{aligned} \Lambda_\epsilon(A) &= \{z \in \mathbf{C} : X - zI \text{ singular}, X \in \mathbf{M}^n, \|X - A\| \leq \epsilon\} \\ &= \{z \in \mathbf{C} : \sigma_n(A - zI) \leq \epsilon\}. \end{aligned}$$

Thus the ϵ-pseudospectrum of A consists of all eigenvalues of matrices within a distance ϵ of A, and in particular, $\Lambda_0(A)$ is just the spectrum of A. The third quantity we consider is the *Kreiss constant*

$$K(A) = \sup\Big\{\frac{|z| - 1}{\epsilon} : \epsilon > 0,\ z \in \Lambda_\epsilon(A)\Big\},$$

which is a bound on the linear rate at which the ϵ-pseudospectrum bulges outside the unit disk, as a function of ϵ. This supremum in fact occurs for bounded ϵ: for matrices with eigenvalues close to the unit circle, the growth in the ϵ-pseudospectrum for small $\epsilon > 0$ is the key phenomenon.

Theorem 10.2 (Kreiss, 1962). *Over any set of matrices in $\mathbf{M}^n$, the Kreiss constant K, the power bound P, and the Lyapunov condition factor L are either all uniformly bounded above or all unbounded.*

Results parallel to Theorems 10.1 and 10.2 hold for the continuous-time version. In that case we are interested in bounding $\|e^{tA}\|$ $(t \geq 0)$, the left halfplane takes the place of the unit disk, and the Lyapunov inequality becomes $H \succ 0$, $0 \succ A^* H + HA$. The relevant pseudospectral quantity is $\epsilon^{-1} \mathrm{Re}\, z$ $(z \in \Lambda_\epsilon(A))$.

Kreiss's original work included inequalities relating the three functions K, P and L. These inequalities have gradually improved in subsequent work, culminating in the following tight result [58]

Theorem 10.3 (Spijker, 1991). *On the space $\mathbf{M}^n$, the Kreiss constant K and the power bound P are related by the inequality $K \leq P \leq enK$.*

Let us return to the continuous-time case, where we are interested in the dynamical system $\dot{x} = Ax$. Trajectories of this system all converge to the origin exponentially if and only if the *spectral abscissa*

$$\alpha(A) = \max \mathrm{Re}\, \Lambda_0(A)$$

is strictly negative. But as we observed at the beginning of the section, this condition is not robust: even if it holds, the *pseudospectral abscissa*

$$\alpha_\epsilon(A) = \max \operatorname{Re} \Lambda_\epsilon(A)$$

may be positive for small $\epsilon > 0$. In other words, nearby matrices may not be stable, and relatedly (via the Kreiss matrix theorem (10.2)), trajectories of $\dot{x} = Ax$ may have large transient peaks.

This argument suggests the pseudospectral abscissa α_ϵ (for some small $\epsilon > 0$) is a better measure of system decay than the spectral abscissa. It is closely related to two other important quantities in robust control theory: for a stable matrix A, the *complex stability radius* [28] is

$$\beta(A) = \max\{\epsilon \geq 0 : \alpha_\epsilon(A) \leq 0\}, \tag{10}$$

which is just the distance to the set of matrices that are not stable, and the $\mathbf{H}_\infty$ *norm* of the transfer function $s \mapsto (sI - A)^{-1}$, which is $\beta(A)^{-1}$ (see for example [8]).

As well as its advantages in modeling, the pseudospectral abscissa α_ϵ has computational attractions. Regularizing the spectral abscissa in this way enhances convexity properties: in particular [12],

$$\lim_{\epsilon \to \infty} \big(\alpha_\epsilon(A) - \epsilon\big) = \frac{1}{2}\lambda_1(A + A^*), \tag{11}$$

and the right hand side is a convex function of the matrix A. The pseudospectral abscissa also has better Lipschitz properties than the spectral abscissa, as the following result shows [12,41].

Theorem 10.4 (Lipschitzness of pseudospectral abscissa). *Around any given nonderogatory matrix, the pseudospectral abscissa α_ϵ is locally Lipschitz and subdifferentially regular for all small $\epsilon > 0$.*

The pseudospectral abscissa is surprisingly easy to calculate efficiently. An algorithm building on well-known methods for the $\mathbf{H}_\infty$ norm [7,9], using a test for imaginary eigenvalues of associated $2n$-by-$2n$ Hamiltonian matrices, turns out to be globally and quadratically convergent and effective in practice [13].

Control problems often involve a parametrized matrix A, and we wish to choose the parameters in order to enhance some stability aspect of the system $\dot{x} = Ax$. By minimizing the pseudospectral abscissa α_ϵ for various choices of $\epsilon \geq 0$, we construct a range of options balancing asymptotic decay and the avoidance of transient growth: large ϵ amounts to a focus on initial growth (a consequence of equation (11)); some intermediate choice of ϵ corresponds to minimizing the $\mathbf{H}_\infty$ norm (a consequence of equation (10)); $\epsilon = 0$ focuses on asymptotic decay.

Semidefinite-programming-based techniques for optimizing robust stability, involving variants of the Lyapunov inequality $H \succ 0$ and $0 \succ A^*H + HA$, may not be practical except for matrices A of low dimension due to the introduction of many subsidiary variables (the entries of the Hermitian matrix H). By contrast, as we remarked above, the pseudospectral abscissa is relatively easy to compute (along with subdifferential information), and local optimization is promising both in theory, in results like Theorem 10.4 (Lipschitzness of pseudospectral abscissa), and in computational practice [11].

We end with one more open question. Suppose the spectral abscissa minimization problem $\inf_{\mathcal{F}} \alpha$ has optimal solution some nonderogatory matrix $\bar{X}$. If we instead consider the regularized problem $\inf_{\mathcal{F}} \alpha_\epsilon$, we generate corresponding optimal solutions X_ϵ converging to $\bar{X}$ as $\epsilon \to 0$. Is the pseudospectral abscissa α_ϵ locally Lipschitz around X_ϵ? Theorem 10.4 shows the answer is yes when the feasible region $\mathcal{F}$ is a singleton, but in general the answer seems less clear.

Acknowledgements. Many people have contributed to the viewpoints I have sketched. I would like to express my particular thanks to: Jon Borwein, for teaching me the power of convex and nonsmooth analysis; Michael Overton, for his constant encouragement on eigenvalue optimization; Jim Renegar, for his broad, level-headed enthusiasm for optimization theory; Jim Burke, Jean-Baptiste Hiriart-Urruty, Claude Lemaréchal and François Oustry, for helping convince me of the convergence of nonsmooth theory and practice; Levent Tunçel and Henry Wolkowicz, my colleagues at Waterloo, for their numerous insightful presentations on semidefinite programming; Stephen Boyd, Laurent El Ghaoui and Carsten Scherer, for encouraging me to delve into the control literature; Dragomir Djoković for his patient help with Lie theory; Nick Trefethen, for his advice on pseudospectra; Chris Godsil, for his insights on linear algebra and compact disks.

References

1. V.I. Arnold. On matrices depending on parameters. *Russian Mathematical Surveys*, 26:29–43, 1971.
2. H.H. Bauschke, O. Güler, A.S. Lewis, and H.S. Sendov. Hyperbolic polynomials and convex analysis. *Canadian Journal of Mathematics*, 53:470–488, 2001.
3. A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications.* SIAM, Philadelphia, 2001.
4. F.A. Berezin and I.M. Gel'fand. Some remarks on the theory of spherical functions on symmetric Riemannian manifolds. *Tr. Mosk. Mat. Obshch.*, 5:311–351, 1956. English translation in *AMS Translations* 21:193–238, 1962.
5. R. Bhatia. *Matrix Analysis.* Springer, New York, 1997.
6. J.M. Borwein and A.S. Lewis. *Convex Analysis and Nonlinear Optimization.* Springer-Verlag, New York, 2000.
7. S. Boyd and V. Balakrishnan. A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞-norm. *Systems and Control Letters*, 15:1–7, 1990.
8. S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* SIAM, Philadelphia, 1994.
9. N.A. Bruinsma and M. Steinbuch. A fast algorithm to compute the $\mathbf{H}_\infty$-norm of a transfer function matrix. *Systems and Control Letters*, 14:287–293, 1990.
10. J.V. Burke, A.S. Lewis, and M.L. Overton. Optimal stability and eigenvalue multiplicity. *Foundations of Computational Mathematics*, 1:205–225, 2001.
11. J.V. Burke, A.S. Lewis, and M.L. Overton. A nonsmooth, nonconvex optimization approach to robust stabilization by static output feedback and low-order controllers. Technical report, Courant Institute, 2003. To appear, IFAC Proceedings, Milan.
12. J.V. Burke, A.S. Lewis, and M.L. Overton. Optimization and pseudospectra, with applications to robust stability. *SIAM Journal on Matrix Analysis and Applications*, 2003. To appear.
13. J.V. Burke, A.S. Lewis, and M.L. Overton. Robust stability and a criss-cross algorithm for pseudospectra. *IMA Journal of Numerical Analysis*, 23:1–17, 2003.

14. J.V. Burke, A.S. Lewis, and M.L. Overton. Variational analysis of the abscissa mapping for polynomials and the Gauss-Lucas theorem. *Journal of Global Optimization*, 2003. To appear.
15. J.V. Burke and M.L. Overton. Variational analysis of non-Lipschitz spectral functions. *Mathematical Programming*, 90:317–352, 2001.
16. J.V. Burke and M.L. Overton. Variational analysis of the abscissa mapping for polynomials. *SIAM Journal on Control and Optimization*, 39:1651–1676, 2001.
17. C.B. Chua. Relating homogeneous cones and positive definite cones via T-algebras. Technical report, Cornell University, 2003. Submitted to *SIAM Journal on Optimization*.
18. F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory*. Springer-Verlag, New York, 1998.
19. J. Dadok. On the C^∞ Chevalley's theorem. *Advances in Mathematics*, 44:121–131, 1982.
20. C. Davis. All convex invariant functions of hermitian matrices. *Archiv der Mathematik*, 8:276–278, 1957.
21. L. Faybusovich. On Nesterov's approach to semi-infinite programming. *Acta Applicandae Mathematicae*, 74:195–215, 2002.
22. L. Gårding. Linear hyperbolic differential equations with constant coefficients. *Acta Mathematica*, 85:2–62, 1951.
23. L. Gårding. An inequality for hyperbolic polynomials. *Journal of Mathematics and Mechanics*, 8:957–965, 1959.
24. Pseudospectra Gateway. `http://web.comlab.ox.ac.uk/projects/pseudospectra`.
25. O. Güler. Hyperbolic polynomials and interior point methods for convex programming. *Mathematics of Operations Research*, 22(2):350–377, 1997.
26. R. Hauser and O. Güler. Self-scaled barrier functions on symmetric cones and their classification. *Foundations of Computational Mathematics*, 2:121–143, 2002.
27. J.W. Helton and V. Vinnikov. Linear matrix inequality representation of sets. Technical report, Mathematics Department, UCSD, 2002.
28. D. Hinrichsen and A.J. Pritchard. Stability radii of linear systems. *Systems and Control Letters*, 7:1–10, 1986.
29. J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Springer-Verlag, Berlin, 1993.
30. R.A. Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. *American Journal of Mathematics*, 76:620–630, 1954.
31. F. Jarre. Convex analysis on symmetric matrices. In [66], 13–28.
32. B. Kostant. On convexity, the Weyl group and the Iwasawa decomposition. *Annales scientifiques de l'Ecole Normale Supérieure*, 6:413–455, 1973.
33. H.O. Kreiss. Uber die Stabilitätsdefinition für Differenzengleichungen die partielle Differenzialgleichungen approximieren. *BIT*, 2:153–181, 1962.
34. P.D. Lax. Differential equations, difference equations and matrix theory. *Communications on Pure and Applied Mathematics*, 6:175–194, 1958.
35. A.S. Lewis. The convex analysis of unitarily invariant matrix functions. *Journal of Convex Analysis*, 2:173–183, 1995.
36. A.S. Lewis. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6:164–177, 1996.
37. A.S. Lewis. Derivatives of spectral functions. *Mathematics of Operations Research*, 6:576–588, 1996.
38. A.S. Lewis. Lidskii's theorem via nonsmooth analysis. *SIAM Journal on Matrix Analysis and Applications*, 21:379–381, 1999.
39. A.S. Lewis. Nonsmooth analysis of eigenvalues. *Mathematical Programming*, 84:1–24, 1999.
40. A.S. Lewis. Convex analysis on Cartan subspaces. *Nonlinear Analysis, Theory, Methods and Applications*, 42:813–820, 2000.
41. A.S. Lewis. Robust regularization. Technical report, Simon Fraser University, 2002. Submitted to Mathematical Programming.
42. A.S. Lewis. Active sets, nonsmoothness and sensitivity. *SIAM Journal on Optimization*, 13:702–725, 2003.
43. A.S. Lewis and M.L. Overton. Eigenvalue optimization. *Acta Numerica*, 5:149–190, 1996.
44. A.S. Lewis, P. Parrilo, and M. Ramana. The Lax conjecture is true. Technical report, Simon Fraser University, 2003. Submitted to Proceedings of the American Mathematical Society.
45. A.S. Lewis and H.S. Sendov. Self-concordant barriers for hyperbolic means. *Mathematical Programming*, 91:1–10, 2001.

46. A.S. Lewis and H.S. Sendov. Twice differentiable spectral functions. *SIAM Journal on Matrix Analysis and Applications*, 23:368–386, 2001.
47. V.I. Lidskii. On the proper values of a sum and product of symmetric matrices. *Dokl. Akad. Nauk. SSSR*, 75:769–772, 1950.
48. Y. Lim, J. Kim, and L. Faybusovich. Simultaneous diagonalization on simple Euclidean Jordan algebras and its applications. Technical report, University of Notre Dame, 2003.
49. A.M. Lyapunov. *Problème général de la stabilité du mouvement.* Number 17 in Annals of Mathematics Studies. Princeton University Press, Princeton, N.J., 1947. Reprinted from *Ann. Fac. Sci. Univ. Toulouse*, 2:27-247, 1907, a French translation of a Russian original from 1893. English translation: *The General Problem of the Stability of Motion.* Taylor and Francis, London, 1992.
50. M. Marden. *Geometry of Polynomials.* American Mathematical Society, 1966.
51. A. Nemirovski, 2000. Private communication.
52. Y.E. Nesterov and A.S. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming.* SIAM, Philadelphia, 1994.
53. A.L. Onishchik and E.B. Vinberg. *Lie Groups and Algebraic Groups.* Springer-Verlag, Berlin, 1990.
54. F. Oustry. A second-order bundle method to minimize the maximum eigenvalue function. *Mathematical Programming*, 89:1–33, 2000.
55. J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization.* SIAM, Philadelphia, 2001.
56. R.T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, N.J., 1970.
57. R.T. Rockafellar and R.J.-B. Wets. *Variational Analysis.* Springer, Berlin, 1998.
58. M.N. Spijker. On a conjecture by Levesque and Trefethen related to the Kreiss matrix theorem. *BIT*, 31:551–555, 1991.
59. T.-Y. Tam and W.C. Hill. Derivatives of orbital functions, an extension of Berezin-Gel'fand's theorem and applications. Technical report, Auburn University, 2001.
60. L.N. Trefethen. Pseudospectra of linear operators. *SIAM Review*, 39:383–406, 1997.
61. L.N. Trefethen. Computation of pseudospectra. *Acta Numerica*, 8:247–295, 1999.
62. L. Tunçel, 1995. Private communication.
63. V. Vinnikov. Self-adjoint determinantal representations of real plane curves. *Mathematische Annalen*, 296:453–479, 1993.
64. J. von Neumann. Some matrix inequalities and metrization of matric-space. *Tomsk University Review*, 1:286–300, 1937. In: *Collected Works*, Pergamon, Oxford, 1962, Volume IV, 205-218.
65. G.A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:33–45, 1992.
66. H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications.* Kluwer, Boston, 2000.
67. K. Ziętak. Subdifferentials, faces and dual matrices. *Linear Algebra and its Applications*, 185:125–141, 1993.