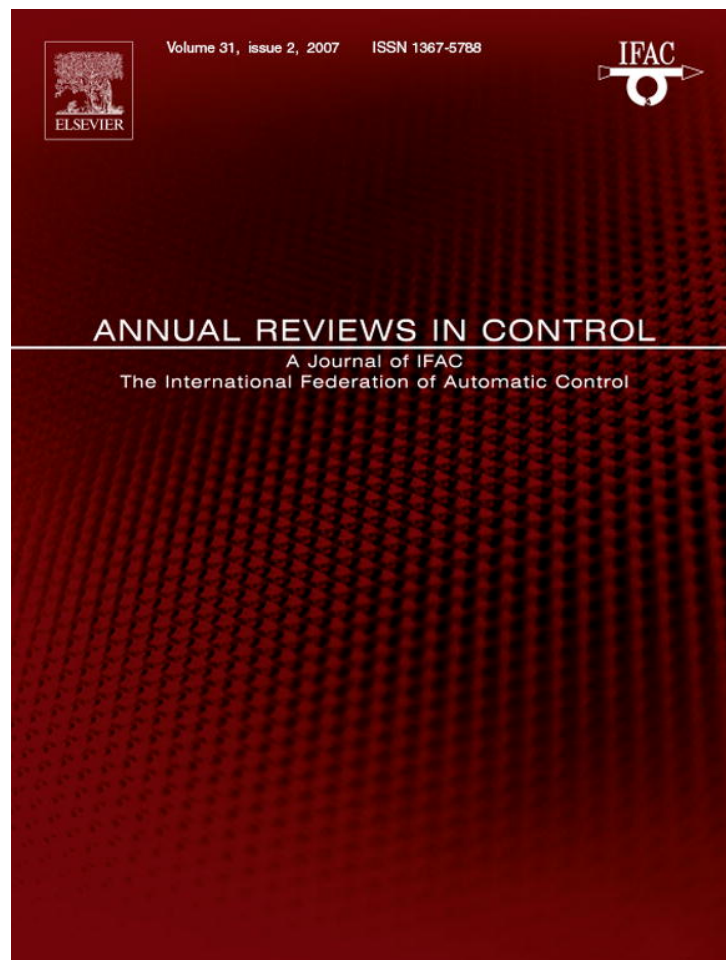


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Nonsmooth optimization and robust control[☆]

A.S. Lewis

School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853, USA

Received 11 January 2007; accepted 5 September 2007

Available online 7 November 2007

Abstract

Many questions of robust control analysis and synthesis fundamentally involve nonsmooth sets and functions, and their variational properties. Central examples include distances to instability and uncontrollability, the H^∞ norm, and pseudospectra. This plenary presentation at the Fifth IFAC Symposium on Robust Control Design (2006), surveys what current ideas from nonsmooth analysis say about the structure and conditioning of such functions and sets, and their numerical optimization. The presentation focuses on notions of nonsmooth derivatives and regularity and on structural tools such as partial smoothness and semi-algebraic techniques, illustrating how each idea helps in analysis and algorithm design. This survey relies heavily on joint work of the author with J.V. Burke (University of Washington) and M.L. Overton (Courant Institute).

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Nonsmooth optimization; Robust control; Variational analysis; Semi-algebraic; Stability radius; Distance to uncontrollability; H -Infinity; Algorithm; Sensitivity analysis; Partly smooth; Pseudospectra; Transient dynamics; Pole placement

1. Outline

The mathematical formulation of robust control analysis and synthesis problems inevitably introduces sets and functions not amenable to classical calculus and optimization due to inherent nonsmoothness. The theory of variational analysis and nonsmooth optimization has matured over the last several decades into an elegant and powerful toolkit for just such analysis (Clarke, 1983; Clarke, Ledyaev, Stern, & Wolenski, 1998; Mordukhovich, 2006; Rockafellar & Wets, 1998). This paper aims to illustrate potential roles for modern variational analysis in the contemporary theory and computational practice of robust control. We make no claim to a comprehensive survey. Instead, in the form of a variety of case studies, and keeping our mathematical language as informal as possible, we simply extend an invitation to variational analysis for robust control specialists.

We begin by considering a list of typical nonsmooth sets and functions appearing in robust control models. Taking as a first example the distance to uncontrollability, we illustrate the use of semi-algebraic ideas in its analysis.

Robust control synthesis often results in optimization models whose solutions occur at points where the objective function is nonsmooth. We illustrate using Blondel's "chocolate problem", a well-known illustration of the difficulty of simultaneous plant stabilization (Blondel, 1994). Numerical algorithms for optimizing the kind of nonsmooth functions appearing in such problems (or in H^∞ synthesis, to take another example) are not widely available. We discuss one conceptually simple and easily implementable method known as "gradient sampling".

Nonsmooth sets and functions appearing in robust control models are free from many of the pathologies of concern in the general mathematical theory of variational analysis. Instead they are highly structured: in particular, while nonsmooth points are fundamental to the control model, in practice these points are confined to certain smooth manifolds. We discuss one particularly important structure ("partial smoothness") capitalizing on this behavior, useful for analyzing the sensitivity of optimal designs and the behavior of solution algorithms.

As a final example, we consider "pseudospectra", the sets of all eigenvalues of matrices in a neighborhood of given matrix. Pseudospectra are highly effective tools for studying many properties of nonnormal matrices, and in particular for understanding the transient dynamics of associated systems (Trefethen & Embree, 2005). Until recently, their application has been primarily confined to analysis, using the powerful computational tool EIGTOOL (Embree & Trefethen).

[☆] Research supported in part by National Science Foundation Grant DMS-0504032.

E-mail address: aslewis@orie.cornell.edu.

URL: <http://www.orie.cornell.edu/~aslewis>.

However, studying the variational analysis of the pseudospectrum as a set-valued mapping opens up a range of possibilities for robust pole placement and control synthesis.

Modern robust control analysis and synthesis has been revolutionized by “linear matrix inequality” techniques (Boyd, El Ghaoui, Feron, & Balakrishnan, 1994). Such techniques elegantly transform a rich variety of control problems into convex optimization models that are in principle computationally tractable. However, the resulting models involve subsidiary semidefinite matrix variables that can be very large in practice, making the resulting semidefinite programs often challenging or effectively impossible to solve computationally. The techniques we survey in this work take a completely different tack, aiming to avoid this challenge by focusing on the original design variables, and handling the resultant nonsmoothness directly.

2. Examples of nonsmooth matrix functions

We begin by considering a variety of matrix functions, commonly appearing in robust control models, and inherently nonsmooth. We denote the space of complex column n -vectors by \mathbf{C}^n , and on this space $\|\cdot\|$ denotes the usual 2-norm. We denote the space of n -by- n complex matrices by \mathbf{M}^n , and on this space (and more generally on the space of m -by- n matrices) $\|\cdot\|$ denotes the associated operator 2-norm (the *spectral norm*):

$$\|A\| = \max \{ \|Au\| : u \in \mathbf{C}^n, \|u\| = 1 \} \quad \text{for } A \in \mathbf{M}^n.$$

This quantity is just the largest singular value of A .

Notice that we are immediately confronted with nonsmoothness, even in the matrix norm $\|\cdot\|$. If the largest singular value of A is not simple (or in other words if the matrix A^*A has a multiple largest eigenvalue), then $\|\cdot\|$ is not smooth at A . However, being convex, the norm $\|\cdot\|$ is straightforward to understand using classical convex analysis as developed in Rockafellar (1970) (see for example Watson, 1992; Zietak, 1993), and to optimize using interior point techniques (Nesterov & Nemirovskii, 1994). By contrast, our interest in this work is in *nonconvex* functions.

A simple but representative example of the kind of nonsmooth nonconvex function ubiquitous in robust control is the *smallest* singular value:

$$\sigma_{\min}(A) = \min \{ \|Au\| : u \in \mathbf{C}^n, \|u\| = 1 \} \quad \text{for } A \in \mathbf{M}^n.$$

The smallest singular value measures the distance to the nearest singular matrix, or more generally, for rectangular A , to the nearest rank-deficient matrix.

Designing a parametrized matrix in order to maximize its distance to singularity is perhaps a rather artificial problem. Of more genuine interest (Van & Loan, 1985) is the distance to the nearest *unstable* matrix, by which we mean a matrix with an eigenvalue having an eigenvalue with nonnegative real part. An easy and well-known result expresses this function, also known as the *complex stability radius*, as

$$r_{\mathbf{C}}(A) = \min \{ \sigma_{\min}(zI - A) : z \in \mathbf{C}, \operatorname{Re}z > 0 \} \quad \text{for } A \in \mathbf{M}^n.$$

Like the smallest singular value, this function is nonsmooth and nonconvex.

The complex stability radius is a special case of the H^∞ norm, a key notion in robust control (Zhou, Doyle, & Glover, 1996). The H^∞ norm of a transfer function

$$H(z) = C(zI - A)^{-1}B + D \quad (z \in \mathbf{C})$$

(for appropriately sized matrices A, B, C, D) is the quantity

$$\|H\|_\infty = \max \{ \|H(z)\| : z \in \mathbf{C}, \operatorname{Re}z > 0 \}.$$

In terms of the variables A, B, C, D , this function is again nonsmooth and nonconvex. (Setting B and C to be identity matrices and $D = 0$ gives the reciprocal of the complex stability radius.)

Given two appropriately sized matrices A and B , classical theory (Zhou et al., 1996) shows that the control system $\dot{x} = Ax + Bu$ is uncontrollable exactly when there exists a point $z \in \mathbf{C}$ for which the matrix $[A - zI \quad B]$ is rank-deficient. An elementary argument (Eising, 1984) then shows that the *distance to uncontrollability* for a matrix pair (A, B) is the quantity

$$\begin{aligned} \min \{ \|(A', B') - (A, B)\| : \dot{x} = A'x + B'u \text{ uncontrollable} \} \\ = \min \{ \sigma_{\min}[A - zI \quad B] : z \in \mathbf{C} \}, \end{aligned}$$

which once again depends on the variables A, B in a nonsmooth nonconvex fashion.

If we denote the spectrum of a matrix $A \in \mathbf{M}^n$ by $\Lambda(A)$, then the asymptotic rate of decay of the dynamical system $\dot{x} = Ax$ is determined by the *spectral abscissa*

$$\alpha(A) = \max \{ \operatorname{Re}z : z \in \Lambda(A) \}.$$

In particular, the matrix A is stable exactly when $\alpha(A) < 0$. Like the functions above, the spectral abscissa is continuous, nonsmooth and nonconvex, but unlike the previous examples it is not even Lipschitz.

The last example we consider in this section is a set-valued mapping rather than a real-valued function. A general tool for studying robust properties of the spectrum of a matrix $A \in \mathbf{M}^n$ is the ϵ -*pseudospectrum*

$$\Lambda_\epsilon(A) = \bigcup_{\|X-A\| \leq \epsilon} \Lambda(X)$$

(for given $\epsilon \geq 0$). When A is not normal (meaning $A^*A \neq AA^*$), pseudospectra are often more informative for modeling purposes than the spectrum, in particular providing better information about the transient behavior of the dynamical system $\dot{x} = Ax$. Often more convenient is the equivalent definition

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

Variational analysis is ideally suited to the study of such set-valued mappings.

3. Nonsmooth design and LMIs

Simply put, design synthesis is numerically challenging because computational minimization of nonsmooth, nonconvex functions is hard. To take one artificially simple but illustrative example, consider the problem of choosing a small real parameter t with the aim of maximizing the distance to singularity function

$$\sigma_{\min} \begin{bmatrix} 1 & t+d & 0 & 0 \\ -t-d & 0 & 0 & 0 \\ 0 & 0 & 1 & t-d \\ 0 & 0 & -t+d & 0 \end{bmatrix},$$

for a given small constant $d > 0$. It is not hard to see that the optimal solution is $t = 0$. However, we can see from the graph in Fig. 1 that the function we seek to maximize is not concave, and is nonsmooth at the optimal solution, making traditional analysis and computational approaches impossible.

In nonsmooth optimization in general and robust control synthesis in particular, we confront a variety of challenging questions:

- How many local optima does the problem have?
- How can we recognize local optima?
- How can we approximate local optima computationally?
- How sensitive are local optima to small changes in the model data?

As we mentioned in the introduction, many robust control problems are “semidefinite representable” (Ben-Tal & Nemirovski, 2001; Boyd et al., 1994), so computationally tractable, at least in principle. To take just one example, as a consequence of the “bounded real lemma”, the transfer function $z \mapsto C(zI - A)^{-1}B + D$ (where the matrices A, B, C, D are real) has H^∞ norm strictly less than γ if and only if there exists a positive-definite symmetric matrix Y with

$$\begin{bmatrix} A^T Y + YA & YB & C^T \\ B^T Y & -\gamma I & D^T \\ C & D & -\gamma I \end{bmatrix}$$

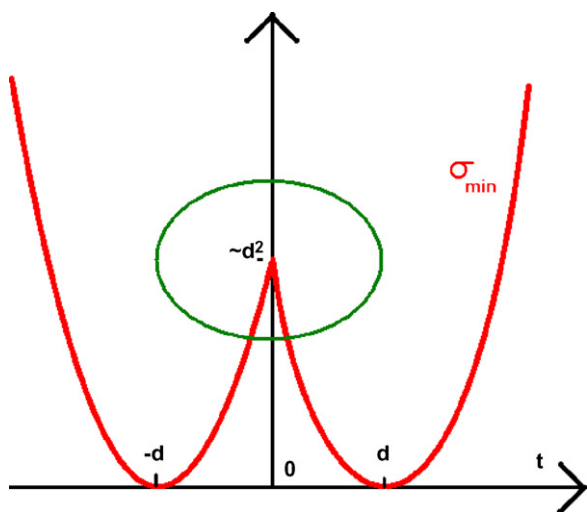


Fig. 1. The distance to singularity.

negative-definite. Suppose the matrices A, B, C, D depend linearly on design variables that we wish to choose in order to minimize the H^∞ norm. We can approach this problem by simultaneously choosing the subsidiary variables Y and γ in order to minimize γ . The optimal solution of the resulting semidefinite program can be approximated in polynomial time by interior point techniques. A variety of more general problems in robust control can be attacked analogously, using techniques from smooth optimization (see for example Apkarian, Noll, Thevenet, & Tuan, 2004; Apkarian, Noll, & Tuan, 2003; Leibfritz, 2001; Leibfritz & Mustafa, 2003; Leibfritz & Volkwein, 2006). However, if the dimension of the state space is large, then the matrix Y is correspondingly large, making the semidefinite programs computationally challenging.

Suppose instead we consider the H^∞ norm simply as a function of the design variables, avoiding the introduction of subsidiary variables. This function has the substantial disadvantage of being neither convex nor smooth. On the other hand, the design variables may be few. Attacking robust control synthesis problems directly via their nonsmooth models has gained increasing attention due to the challenge of large-scale semidefinite programming (Apkarian and Noll, 2006a, 2006b, 2006c; Burke, Lewis, & Overton, 2004a; Burke, Henrion, Lewis, & Overton, 2006b). How analytically and computationally challenging is this direct approach?

4. Semi-algebraic techniques

In optimization theory, even smooth functions can behave badly. For example, given a real parameter d , consider the problem of minimizing the real function

$$x \mapsto x^2 \left(2 + \sin \left(\frac{1}{x} \right) \right) + dx$$

As Fig. 2 illustrates, this problem has many local minima, the optimal solution depends discontinuously on d , and classical optimization algorithms will fail.

However, the kind of oscillatory pathology driving this example (and many like it throughout real and variational analysis) is artificial. To be specific, all the functions we considered in Section 2 are *semi-algebraic*. In other words, the graph of each function is a *semi-algebraic set*, meaning it can be decomposed into a finite union of sets, each of which is defined by finitely many polynomial inequalities. This rich class of sets is often easy to recognize, and is stable under many common operations, including, most importantly, projection—

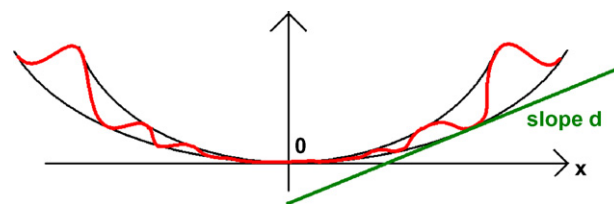


Fig. 2. A nonisolated local minimum.

a consequence of the Tarski–Seidenberg Theorem (see for example [Benedetti & Risler, 1990](#)).

The oscillations causing difficulties in the example above cannot occur in semi-algebraic sets and functions. More generally, a variety of important “regularity” properties that may fail in general variational analysis nonetheless hold in a semi-algebraic context. We illustrate with a couple of recent examples.

Sard’s Theorem in classical analysis shows that if a function $F : \mathbf{R}^n \rightarrow \mathbf{R}^m$ is sufficiently smooth then the set of solutions x to the equation $F(x) = y$ “typically” depends on the right-hand side y in a “stable” way. (To be precise, the set of values $F(x)$ for which the Jacobian $\nabla F(x)$ is rank-deficient has Lebesgue measure zero.) See [Bates \(1993\)](#) for the most precise version known for Sard’s Theorem. Versions for general semi-algebraic real-valued nonsmooth functions exist (see for example [Bolte, Daniilidis, Lewis, & Shiota, 2005](#); [Bolte, Daniilidis, & Lewis, 2006](#)), but very recently, Sard’s Theorem was radically and beautifully extended to semi-algebraic *set-valued* mappings ([Ioffe, 2006](#)), opening up the possibility of considering systems of nonsmooth equations and inequalities.

As a second example, we consider Newton’s method for solving an equation $F(x) = 0$. Traditionally, we assume that the function $F : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is twice continuously differentiable, in which case the method converges locally quadratically under a nondegeneracy condition. However, versions of Newton’s method can work under much less stringent smoothness assumptions on F , opening up applications to problems with complementarity or equilibrium conditions: a good survey is ([Facchinei & Pang, 2003](#)). Superlinear convergence of such nonsmooth Newton methods depends on *semismoothness*: the directional derivative of F should satisfy the condition

$$F'(x + y; y) - F'(x; y) = o(y) \quad \text{for small } y.$$

This condition holds *whenever* F is semi-algebraic ([Bolte, Daniilidis, & Lewis, 2007](#)).

Semi-algebraic techniques can be very helpful in a variety of robust control models. Consider for example the problem discussed above of the distance to uncontrollability for the system $\dot{x} = Ax + Bu$. To compute this quantity (a problem solved only rather recently [Burke, Lewis, & Overton, 2004b](#); [Gu, 2000](#); [Gu, Mengi, Overton, Xia, & Zhu, 2006](#)), we must minimize the function

$$z \mapsto \sigma_{\min}[A - zI \quad B],$$

over the complex plane. We could ask how many local minimizers this functions may have ([Wright & Trefethen, 2002](#)). Relatedly, how many components might the level set

$$\{z \in \mathbf{C} : \sigma_{\min}[A - zI \quad B] \leq \epsilon\}$$

have? Using classical techniques from real algebraic geometry, [Burke et al. \(2004b\)](#) shows that the number of components cannot exceed $2m^2 - m + 1$, where m is the number of rows in the matrix A .

5. Nonsmoothness and Blondel’s problem

Why exactly does nonsmoothness cause analytic and computational difficulties in optimization problems in general and robust control models in particular?

To understand this issue, recall first how we use smoothness in basic optimization problems. Gradients allow us to recognize minimizers of smooth functions (see [Fig. 3](#)). Furthermore, Lipschitz functions are differentiable almost everywhere (by Rademacher’s Theorem—see for example [Borwein & Lewis, 2006](#)), as are semi-algebraic functions. Hence, in practice, one might imagine that nonsmoothness should rarely cause issues.

However, nonsmooth points, although in some sense rare, are crucial in variational problems, because optimizing a function value tends to push variables to nonsmooth points. To see this phenomenon in a control context, we consider a well-known challenge problem due to [Blondel \(1994\)](#). Part of this problem asks (in return for a prize of a kilogram of Belgian chocolate) for *stable* polynomials p and q (that is, having all roots with negative real parts) such that the polynomial

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

is also stable, and where the number δ is 0.9.

One variational approach to this problem is to fix the degrees of p and q , and then to vary these polynomials in order to minimize the *abscissa* of the polynomial pqr (meaning the largest real part of a root). If we succeed in forcing the abscissa below zero, we have solved our problem.

To be concrete, suppose we restrict the polynomial p to be cubic and the polynomial q to be constant, and try to minimize the function

$$\alpha(p(z)[(z^2 - 2\delta z + 1)p(z) + (z^2 - 1)q]),$$

where α denotes the abscissa function. The function α is nonsmooth, nonconvex, and even nonlipschitz around polynomials with a multiple root having maximum real part ([Burke & Overton, 2001b](#)). Polynomials with multiple roots are, in some sense, rare. However, numerical computation suggests that, at the

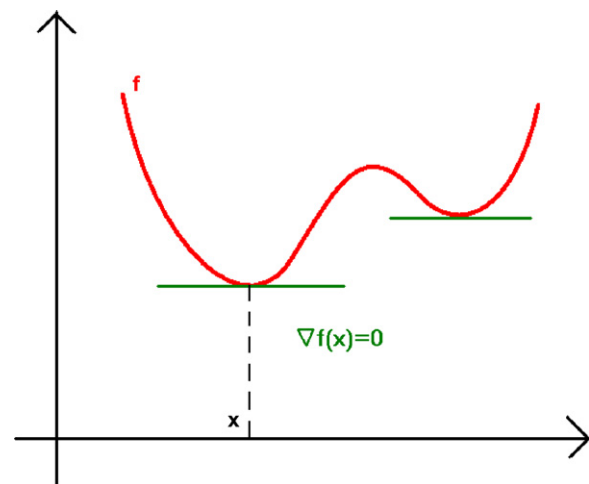


Fig. 3. Critical points.

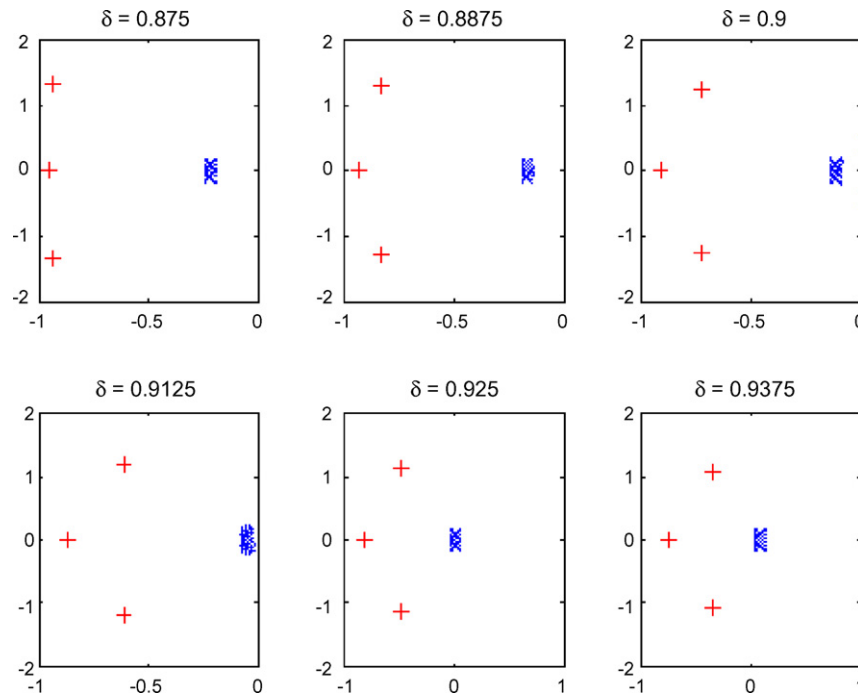


Fig. 4. Roots for Blondel's problem.

optimal solution, the quintic polynomial pqr has a single root of multiplicity five, making the abscissa highly nonsmooth. Fig. 4 (from Burke et al., 2006b) shows the approximate locations of the roots at the optimal solution, obtained numerically, for various values of the constant δ . Roots of the optimal polynomial p are marked with a +; the remaining five, marked \times , are those of the polynomial r . Local optimality can in fact be verified rigorously via nonsmooth calculus (Burke et al., 2006b).

6. Nonsmooth analysis and algorithms

Traditional calculus is inadequate for many of the functions appearing in robust control models. To make progress, we need nonsmooth tools.

The central idea of nonsmooth analysis, as pioneered in Clarke (1973, 1983), is to replace the gradient $\nabla f(x)$ of a Lipschitz function f by the “generalized gradient” or *Clarke subdifferential* $\partial f(x)$, which consists of all *subgradients*: convex combinations of limits of gradients at nearby points (see Fig. 5). Nonsmooth analysis has evolved rapidly (Clarke et al., 1998; Mordukhovich, 2006; Rockafellar & Wets, 1998), but Clarke’s basic idea will suffice for our development here.

For computational minimization of nonsmooth convex functions, *bundle methods* (Lemaréchal, 1975) have proven highly effective. The bundle idea has been extended to nonconvex functions (Kiwiel, 1985; Schramm & Zowe, 1992; Vlček & Lukšan, 2001), although implementation is very delicate, and publicly available code is scarce. A less sophisticated idea, both conceptually and for implementation, grows directly from the definition of the Clarke subdifferential. This “gradient sampling” approach, suggested in Burke, Lewis, & Overton (2002), and developed further in Burke et al. (2004a) and Burke, Lewis, and Overton (2005), approximates

the subdifferential at the current point x_{old} by the convex hull of the random set

$$\{\nabla f(X_i) : X_1, X_2, \dots, X_k \text{ random near } x_{old}\},$$

where the sample size k might be twice the dimension of x , for example. We then compute the shortest element g in this approximation (via a quadratic program), and update:

$$x_{new} = x_{old} - tg,$$

where the step size t is found by an inexact line search. Under reasonable conditions (Kiwiel, 2007), iterates converge to a *Clarke-critical point* \bar{x} : that is, $0 \in \partial f(\bar{x})$.

Gradient sampling code is freely available (MATLAB code), and forms part of the HANSO package for general nonsmooth

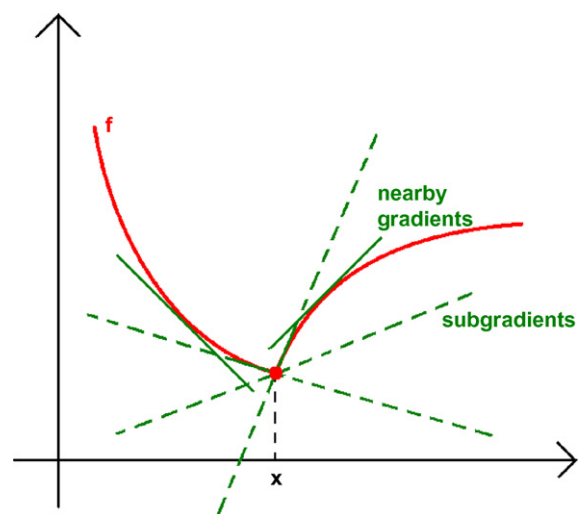


Fig. 5. The Clarke subdifferential.

optimization, and the HIFOO package for H^∞ synthesis (Burke, Henrion, Lewis, & Overton, 2006a). The method involves many gradient computations. However, in many applications, and in robust control in particular, objective function gradients are cheap to compute, often involving the eigenvectors associated with an eigenvalue computation needed to evaluate the objective: typical examples are the complex stability radius, the H^∞ norm, the distance to uncontrollability, and the spectral abscissa. The optimization problems associated with Blondel's chocolate problem above were solved by gradient sampling, for example, and Burke et al. (2004b) describes a gradient-sampling approach for maximizing the distance to uncontrollability.

7. Recognizing minimizers: regularity

The Clarke subdifferential reduces simply to the gradient for smooth functions, and furthermore, for continuous convex functions, it coincides with the classical subdifferential of convex analysis (Rockafellar, 1970). Thus Clarke's notion is an elegant unification of two important classical ideas.

In convex optimization, the critical point condition $0 \in \partial f(\bar{x})$ holds exactly when the point \bar{x} minimizes the function f . For nonconvex f , minimizers must be critical, but even in the smooth case the converse obviously fails, and in the nonsmooth case it can fail badly, as Fig. 6 illustrates.

In the example above, clearly zero is an average of gradients at points arbitrarily near the point \bar{x} , so $0 \in \partial f(\bar{x})$, and yet \bar{x} is actually a “sharp” local maximizer of the function f . One way to describe the bad behavior in this example is to observe that the directional derivative of f jumps down as we pass through the point \bar{x} .

A crucial theoretical idea in nonsmooth analysis is *subdifferential regularity*. A Lipschitz function f is *regular* at a point \bar{x} if, for all directions d , its directional derivative $f'(x; d)$ is upper semicontinuous as the point x varies near \bar{x} (precisely the property that fails in our example). In this case, if \bar{x} is a critical point, then it is a local minimizer, at least to first order:

$$f(x) \geq f(\bar{x}) + o(\|x - \bar{x}\|) \quad \text{for } x \text{ near } \bar{x}.$$

Subdifferential regularity has many ramifications in variational analysis. We restrict ourselves to the comment that many powerful calculus rules for computing subdiffer-

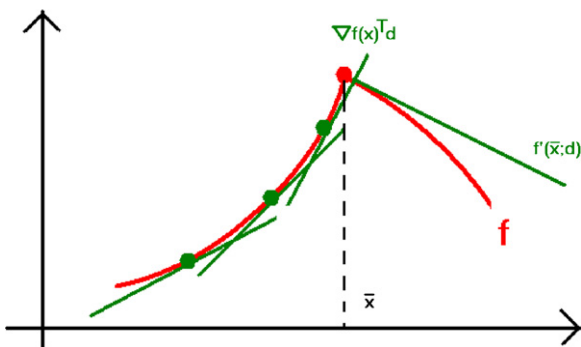


Fig. 6. A nonregular critical point.

entials depend on regularity (Rockafellar & Wets, 1998). The notion extends to nonlipschitz functions: the following result is a remarkable example, explaining in some part the success of the gradient sampling method on Blondel's chocolate problem.

Theorem 7.1 (Burke and Overton (2001b)).

The polynomial abscissa is regular throughout the space of monic polynomials. More precisely, on the space of complex polynomials p of degree n , the function

$$p \mapsto \alpha(z^{n+1} + p(z))$$

is everywhere regular.

The spectral abscissa of a matrix (discussed in Section 2) is just the abscissa of the characteristic polynomial. As a consequence of standard nonsmooth calculus rules, the spectral abscissa is therefore regular at any matrix for which all eigenspaces corresponding to the abscissa are one-dimensional (Burke & Overton, 2001a; Lewis, 2003).

Many of the functions ubiquitous in robust control applications enjoy the important property of subdifferential regularity. In engineering models, often the easiest way to recognize when a function f is regular is to write it in the form

$$f(x) = \max_{s \in S} f_s(x)$$

for some compact set S and where the function

$$(x, s) \mapsto (f_s(x), \nabla f_s(x), \nabla^2 f_s(x))$$

is continuous. (A typical example in the robust control context is the approach of Bompert, Noll, & Apkarian (2006).) Such *subsmooth* functions are everywhere Lipschitz and regular (Rockafellar & Wets, 1998).

Consider the H^∞ norm, for example. Given a transfer function

$$\bar{C}(zI - \bar{A})^{-1}\bar{B} + \bar{D},$$

providing the matrix \bar{A} has no purely imaginary eigenvalues, we can fix a constant k such that the H^∞ norm of any nearby transfer function

$$C(zI - A)^{-1}B + D$$

is given by

$$\begin{aligned} \max \{ \operatorname{Re}(u^*(C(sI - A)^{-1}B + D)v) \\ : \|u\| = 1 = \|v\|, s \in [-k, k] \}. \end{aligned}$$

Thus the H^∞ norm is subsmooth, and hence, in particular, everywhere regular.

Subsmooth functions f in fact have the even stronger property (to which we return a little later) of *prox-regularity*: given any point in the *epigraph*

$$\{(x, r) : r \geq f(x)\},$$

any nearby point has a unique nearest point in the epigraph. Prox-regularity of the abscissa function, by contrast, seems to be an open question.

8. Sensitivity and partial smoothness

Robust control synthesis problems share a fundamental aspect with most practical variational problems: we typically need to know something of how the model solution varies with the nominal data of the problem. In optimization theory, such questions go under the name “sensitivity analysis”.

To take a concrete example, we return to Blondel’s problem from Section 5. Recall that, using our optimization approach, we seek to minimize the objective function

$$f_\delta(p, q) = \alpha(pqr) \quad \text{where}$$

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

over cubic polynomials p and scalars q . We consider the dependence of the solution on the parameter δ .

A reasonable conjecture from the numerical evidence in Section 5 is that, for all values of the parameter δ near 0.9, the optimal solution results in a polynomial r having a quintuple root. Having guessed this structure, it turns out to be a straightforward calculation to find the corresponding solution exactly (Burke et al., 2006b). Furthermore, notwithstanding the nonlipschitz nature of the problem, this solution depends smoothly on δ .

Many optimization models display this kind of behavior: the optimal solution displays some kind of structure, invariant under small changes to the data (in this case, the quintuple root); having identified this structure, the problem is greatly simplified; furthermore, this structured solution varies smoothly with the problem data. In linear or nonlinear programming, for example, under reasonable conditions, small changes to a problem leave unchanged the “active set” (the set of inequality constraints holding with equality at the solution).

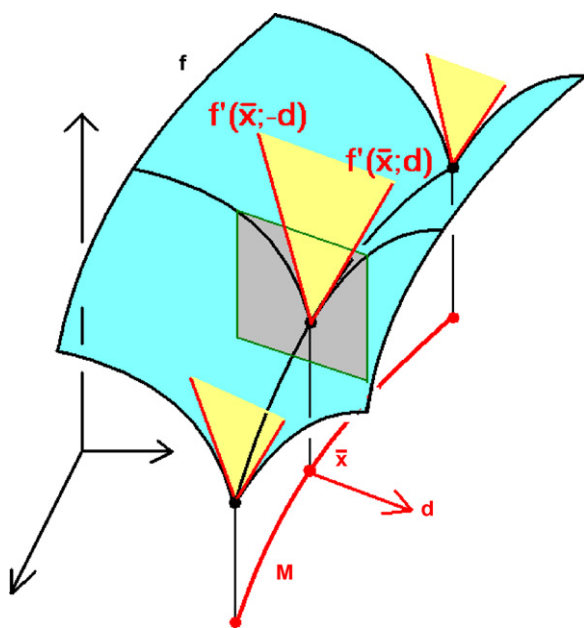


Fig. 7. Partial smoothness.

In semidefinite programming, the rank of the optimal semidefinite matrix typically remains unchanged under small problem perturbations. Understanding such structure is important not only for sensitivity analysis, but for the related convergence theory of solution algorithms.

One broad approach to formalizing the “structure” of an optimal solution is the *partial smoothness* property introduced in Lewis (2003). Fig. 7 illustrates the idea. The idea of “structure” is captured in the definition by membership in a particular *manifold* M (by which we mean the solution set of a smooth system of equations with surjective Jacobian).

A Lipschitz function $f : R^n \rightarrow R$ is *partly smooth* relative to M if its restriction to M is smooth and it is regular throughout M , and furthermore its directional derivative $f'(x; d)$ in any given direction d satisfies the “sharpness” condition

$$f'(x; -d) > -f'(x; d),$$

and behaves continuously as the point x varies on M . (An analogous definition applies to nonlipschitz functions.)

This idea is well suited to sensitivity analysis. Corresponding to a manifold M , consider a partly smooth function $f_\delta : R^n \rightarrow R$, depending in some suitable fashion (see Lewis, 2003) on the parameter δ . Suppose at the point \bar{x} , that the function f satisfies appropriate optimality conditions:

- Strict complementarity: in directions orthogonal to M , the function f_δ grows at least linearly.
- Second-order sufficiency: restricted to M , the function f_δ grows at least quadratically.

Under these conditions, near \bar{x} the function f_δ has a Clarke-critical point x_δ on the manifold M , and this point varies smoothly with the parameter δ . If the function f_δ is also prox-regular (as discussed in the previous section), then the point x_δ is actually a strict local minimizer (Hare & Lewis, 2004).

Partial smoothness may be a useful tool not just for sensitivity analysis but also for understanding the speed of computational nonsmooth optimization algorithms. The packages HANSO (for general nonsmooth optimization) and HIFOO (for H^∞ synthesis) Burke et al. (2006a) use a hybrid approach, involving both the gradient-sampling technique outlined above and a *smooth* quasi-Newton method. Suitably implemented, the quasi-Newton method is surprisingly effective on nonsmooth optimization problems, often converging at an apparent linear rate (Overton, 2006). Preliminary and ongoing experiments suggest that the rate of convergence on a partly smooth function f relative to a manifold M depends on factors such as the curvature of M , the Lipschitz modulus of f , and the conditioning of the restricted Hessian $\nabla^2 f|_M$.

9. Pseudospectra

The last example we introduced in Section 2 is a *set-valued* mapping: the pseudospectrum of a matrix $A \in M^n$:

$$A \mapsto \Lambda_\epsilon(A) = \bigcup_{\|X-A\| \leq \epsilon} \Lambda(X) = \{z : \sigma_{\min}(A - zI) \leq \epsilon\},$$

for some fixed constant $\epsilon > 0$. Pseudospectra capture robust properties of eigenvalues, by their very definition. For example, the matrix A is “robustly stable”, meaning it stays stable under arbitrary perturbations of size no larger than ϵ , exactly when the pseudospectrum $\Lambda_\epsilon(A)$ lies in the left halfplane. Relatedly, the complex stability radius of a stable matrix A is the smallest ϵ for which $\Lambda_\epsilon(A)$ intersects the imaginary axis.

As argued at length in Trefethen and Embree (2005), the spectrum of a nonnormal matrix is often deceptive as a modeling tool, whereas pseudospectra may be much more informative. The advent of efficient computational techniques (Wright) has made analysis using pseudospectral plots very appealing. Studying the variational analysis of pseudospectra opens up the possibility of extending their use beyond analysis to design problems.

To focus on robust control applications, let us consider the stability of the system $\dot{x} = Ax$. We know that the asymptotic behavior of this system depends on the spectrum of the matrix A : in particular, when A is stable, trajectories decay exponentially to the origin. Nonetheless, these trajectories may exhibit large “transient peaks” in the course of this decay, of possibly devastating significance for a practical model. The size of these peaks is related precisely to the pseudospectra of A , and in turn to the corresponding Lyapunov system, via the famous “Kreiss Matrix Theorem”, which we state informally below.

Theorem 9.1 (Kreiss (1962)).

The following properties are equivalent.

- Trajectories for $\dot{x} = Ax$ may have large transient peaks.
- The pseudospectrum $\Lambda_\epsilon(A)$ grows quickly into the right halfplane as ϵ increases away from zero.
- Any positive-definite matrix Y with $A^T Y + YA$ negative-definite must be ill-conditioned.

The Kreiss Theorem is more precise than the statement here (see the discussion in Trefethen & Embree, 2005): a complete statement concerns families of matrices A , and quantifies and relates the notions of “large”, “quickly”, and “ill-conditioned” in the theorem.

As an illustration, consider the following example of Demmel (1987):

$$A = - \begin{bmatrix} 1 & 5 & 5^2 & 5^3 & 5^4 \\ 0 & 1 & 5 & 5^2 & 5^3 \\ 0 & 0 & 1 & 5 & 5^2 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Fig. 8 plots the norm of a trajectory for the system $\dot{x} = Ax$, and clearly shows a transient peak roughly a hundred times the size of the initial point.

The EIGTOOL plot in Fig. 9 explains the large transient peak. The matrix A has a single eigenvalue at -1 . The left-hand picture shows the boundaries of various pseudospectra corresponding to choices of ϵ whose logarithms (to base 10) are shown in the right-

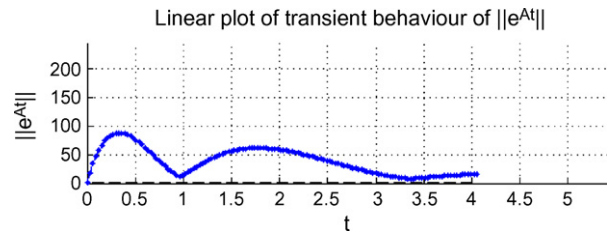


Fig. 8. Trajectory norm for Demmel’s example.

hand color-code. In particular, notice that the pseudospectrum $\Lambda_{10^{-2}}(A)$ extends into the right halfplane, telling us that some unstable matrix X satisfies $\|X - A\| \leq 10^{-2}$. The right-hand picture plots the function $-\log_{10} \sigma_{\min}(A - zI)$: as a curiosity, notice the local minimizer at zero.

The Kreiss Matrix Theorem suggests that a useful tool for studying transient peaks is the *pseudospectral abscissa*

$$\alpha_\epsilon(A) = \max \{ \text{Re} \lambda : \lambda \in \Lambda_\epsilon(A) \}.$$

Fig. 10 illustrates the steps of a “criss-cross” algorithm (Burke, Lewis, & Overton, 2003b) to compute α_ϵ , available in EIGTOOL. The method relies on two key properties of any pseudospectrum.

- Each component contains an eigenvalue (by the maximum modulus principle).
- The intersections of the pseudospectral boundary with any line are easy to compute by a Hamiltonian eigensolver.

As Fig. 10 illustrates, the algorithm proceeds by starting with a rightmost eigenvalue, and then constructing a sequence of vertical lines interspersed with horizontal lines passing through midpoints of vertical segments of intersection with the pseudospectrum. The method is robust, globally and locally quadratically convergent (under reasonable conditions), and seems extremely reliable in practice. Furthermore, as is so often the case for functions associated with spectra properties of matrices, the algorithm returns the gradient $\nabla \alpha_\epsilon$ (when it exists) at essentially no extra cost, allowing the use of the gradient sampling method outlined above to optimize the pseudospectral abscissa (Burke, Lewis, & Overton, 2003a). An analogous algorithm computes the pseudospectral radius (Mengi & Overton, 2005).

From the perspective of robust control specialists, pseudospectra may be seen as an overly conservative tool. The ϵ -pseudospectrum of a real matrix A consists of all eigenvalues of matrices resulting from an *unstructured, complex* perturbation of size no larger than ϵ . Why should such general perturbations be interesting, practically? In measuring the distance to instability, for example, is it not more reasonable to consider the *real stability radius*, where we allow only *real* perturbations (Qiu et al., 1995)?

The Kreiss Matrix Theorem gives one answer to this question. Since the three statements in the result are equivalent, the size of transient peaks associated with a real matrix are determined by the pseudospectra, and hence by *general complex* perturbations, rather than by the real stability radius. Thus, at least in the absence of information about the initial

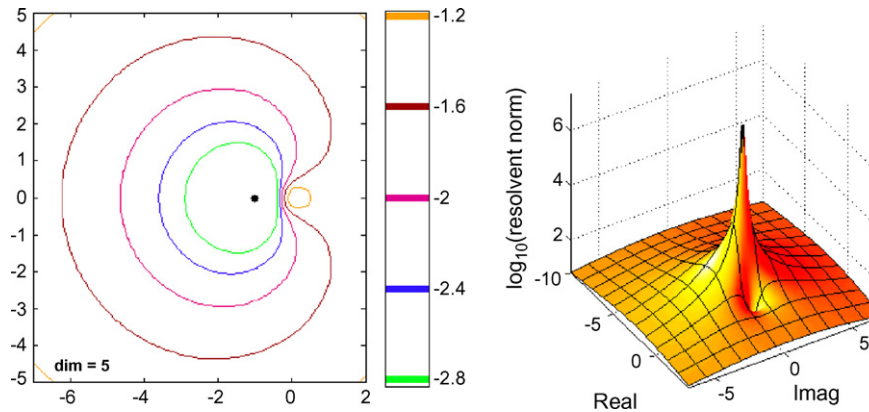


Fig. 9. Pseudospectral plots for Demmel's example.

state, no matter what the structure of the matrix A , pseudospectra provide a correct picture of the transient behavior of the system.

Multiple eigenvalues of nondiagonalizable matrices A are very sensitive to perturbation. To be more precise, the spectrum $X \mapsto \Lambda(X)$, regarded as a set-valued mapping, is not *Lipschitz* around A : there exists no constant k such that

$$d(\Lambda(X), \Lambda(Y)) \leq k \|X - Y\| \quad \text{for all } X, Y \text{ near } A.$$

Here, d denotes the usual *Hausdorff distance* between sets:

$$d(U, V) = \max \left\{ \sup_{u \in U} \inf_{v \in V} \|u - v\|, \sup_{v \in V} \inf_{u \in U} \|v - u\| \right\}.$$

In addition to its advantages as a modeling tool and as a vehicle for robust spectral properties, one may hope that, in computations, the pseudospectrum may behave in a more numerically stable fashion than the spectrum. However, as Fig. 11 illustrates, the pseudospectral mapping $X \mapsto \Lambda_\epsilon(X)$ can also be nonlipschitz. For the given A , Fig. 11 plots several pseudos-

pectral boundaries. The shaded region is the pseudospectrum $\Lambda_\epsilon(A)$ with $\epsilon = (\sqrt{5} - 1)/2$. As we increase ϵ , the boundary moves up the imaginary axis in a nonlipschitz fashion. Nonlipschitz behavior of Λ_ϵ around A follows.

Fortunately, the nonlipschitz behavior illustrated above is not typical: in many settings, we can be confident that the pseudospectrum and associated functions like the pseudospectral abscissa will be Lipschitz, and hence better behaved in numerical applications than the spectrum and the spectral abscissa (Burke et al., 2003a; Burke, Lewis, & Overton, 2007a).

The full power of modern variational analysis gives a concise and elegant way to understand the Lipschitz behavior of the pseudospectrum Λ_ϵ (Lewis & Pang, 2006). A careful analysis shows that Lipschitz behavior fails for a precise reason: the presence on the pseudospectral boundary of a *resolvent-critical point*, by which we mean a Clarke-critical point of the function

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

In the illustration above, for example, the Lipschitz property fails due to the behavior of the pseudospectral mapping near zero, and zero is indeed resolvent-critical. The function above is semi-algebraic, and such functions have a Sard-like property similar to that discussed in Section 4: they can have at most

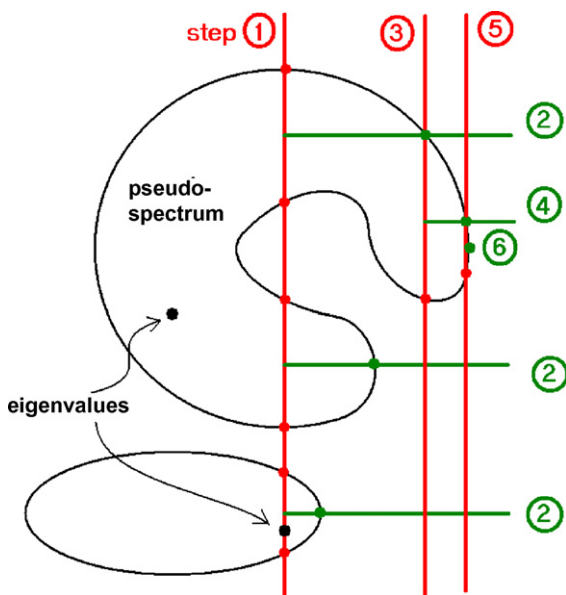


Fig. 10. The criss-cross algorithm for the pseudospectral abscissa.

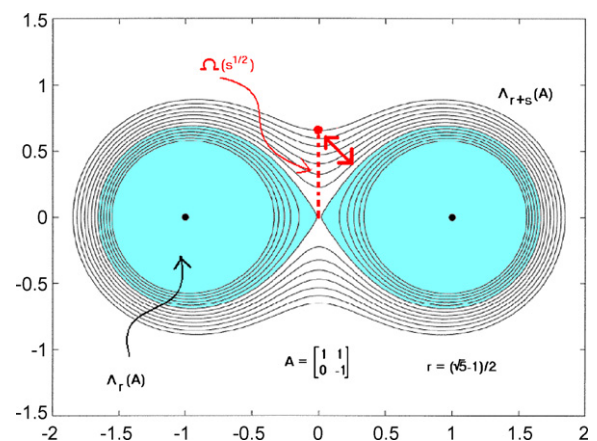


Fig. 11. Nonlipschitz pseudospectral behavior.

finitely many Clarke-critical values (Bolte et al., 2005). Hence, given any matrix A , the pseudospectrum Λ_ϵ must be Lipschitz around A for all but finitely many values of ϵ , and so in particular for all small $\epsilon > 0$.

Resolvent-critical points are significant in a variety of contexts. Consider the basic problem of computing the spectrum of a matrix A . The conditioning of this problem is related to the distance from A to the nearest matrix with a multiple eigenvalue (Demmel, 1990). As shown by Alam and Bora (2005) (see also Burke, Lewis, & Overton, 2007b), this distance is the smallest ϵ for which two distinct components of the pseudospectrum $\Lambda_\epsilon(A)$ coalesce, and the point of coalescence, which is also the required multiple eigenvalue, must be resolvent-critical.

10. Summary

The aim of this survey has been to give specialists in robust control a taste of the flexibility and power of current ideas in nonsmooth optimization. Modern variational analysis furnishes a broad toolkit for analyzing solutions to nonsmooth optimization problems, understanding sensitivity, and developing computational algorithms. The rich structure inherent in robust control synthesis guarantees ongoing fertile interaction with the field of nonsmooth optimization.

References

- Alam, R., & Bora, S. (2005). On sensitivity of eigenvalues and eigendecompositions of matrices. *Linear Algebra and its Applications*, *396*, 273–301.
- Apkarian, P., & Noll, D. (2006c). Controller design with nonsmooth multidirectional search. *SIAM Journal on Control and Optimization*, *44*, 1923–1949.
- Apkarian, P., & Noll, D. (2006a). IQC analysis and synthesis via nonsmooth optimisation. *System and Control Letters*, *55*, 971–981.
- Apkarian, P., & Noll, D. (2006b). Nonsmooth H^∞ synthesis. *IEEE Transactions on Automatic Control*, *51*, 71–86.
- Apkarian, P., Noll, D., Thevenet, J.-B., & Tuan, H. D. (2004). A spectral quadratic SDP method with applications to fixed-order H^2 and H^∞ synthesis. *European Journal of Control*, *10*, 527–538.
- Apkarian, P., Noll, D., & Tuan, H. D. (2003). Fixed-order H^∞ control design via a partially augmented lagrangian method. *International Journal of Robust and Nonlinear Control*, *13*, 1137–1148.
- Bates, S. M. (1993). Towards a precise smoothness hypothesis in Sard's Theorem. *Proceedings of the American Mathematical Society*, *117*, 279–283.
- Ben-Tal, A., & Nemirovski, A. (2001). *Lectures on modern convex optimization: Analysis, algorithms, and engineering applications*. Philadelphia: SIAM.
- Benedetti, R., & Risler, J.-J. (1990). *Real algebraic and semi-algebraic sets*. Paris: Hermann.
- Blondel, V. (1994). *Simultaneous stabilization of linear systems*. Berlin: Springer.
- Bolte, J., Daniilidis, A., & Lewis, A. S. (2006). The Morse–Sard Theorem for non-differentiable subanalytic functions. *Journal of Mathematical Analysis and Applications*, *321*, 729–740.
- Bolte, J., Daniilidis, A., & Lewis, A. S. (in press). Tame mappings are semismooth. *Mathematical Programming, Series B*.
- Bolte, J., Daniilidis, A., Lewis, A. S., & Shiota, M. (2005). Clarke critical values of subanalytic Lipschitz continuous functions. *Annales Polonici Mathematici*, *87*, 13–25.
- Bompert, V., Noll, D., & Apkarian, P. (2006). Second order nonsmooth optimization for H^∞ synthesis. In *Proceedings of ROCOND 2006*.
- Borwein, J. M., & Lewis, A. S. (2006). *Convex analysis and nonlinear optimization* (2nd ed.). New York: Springer.
- Boyd, S., El Ghaoui, L., Feron, E., & Balakrishnan, V. (1994). *Linear matrix inequalities in system and control theory*. Philadelphia: SIAM.
- Burke, J. V., Henrion, D., Lewis, A. S., & Overton, M. L. (2006a). HIFOO—A MATLAB package for fixed-order controller design and H -infinity optimization. In *Proceedings of ROCOND 2006*.
- Burke, J. V., Henrion, D., Lewis, A. S., & Overton, M. L. (2006b). Stabilization via nonsmooth, nonconvex optimization. *IEEE Transactions on Automatic Control*, *51*, 1760–1769.
- Burke, J. V., Lewis, A. S., & Overton, M. L. (2002). Two numerical methods for optimizing matrix stability. *Linear Algebra and its Applications*, *351/2*, 117–145.
- Burke, J. V., Lewis, A. S., & Overton, M. L. (2003). Optimization and pseudospectra, with applications to robust stability. *SIAM Journal on Matrix Analysis and Applications*, *25*, 80–104. Corrigendum: http://www.cs.nyu.edu/cs/faculty/overton/papers/pseudo_corrigenum.html.
- Burke, J. V., Lewis, A. S., & Overton, M. L. (2003b). Robust stability and a criss-cross algorithm for pseudospectra. *IMA Journal of Numerical Analysis*, *23*, 359–375.
- Burke, J. V., Lewis, A. S., & Overton, M. L. (2004a). A nonsmooth, nonconvex optimization approach to robust stabilization by static output feedback and low-order controllers. In S. Bittanti & P. Colaneri (Eds.), In *Proceedings of fourth IFAC symposium on robust control design, Milan, June 2003* (pp. 175–181). Elsevier.
- Burke, J. V., Lewis, A. S., & Overton, M. L. (2004b). Pseudospectral components and the distance to uncontrollability. *SIAM Journal on Matrix Analysis and its Applications*, *26*, 350–361.
- Burke, J. V., Lewis, A. S., & Overton, M. L. (2005). A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM Journal on Optimization*, *15*, 751–779.
- Burke, J. V., Lewis, A. S., & Overton, M. L. (2007a). Convexity and Lipschitz behavior of small pseudospectra. *SIAM Journal on Matrix Analysis and Applications*, *29*, 586–595.
- Burke, J. V., Lewis, A. S., & Overton, M. L. (2007b). Spectral conditioning and pseudospectral growth. *Numerische Mathematik*, *107*, 27–37.
- Burke, J. V., & Overton, M. L. (2001a). Variational analysis of non-Lipschitz spectral functions. *Mathematical Programming*, *90*, 317–352.
- Burke, J. V., & Overton, M. L. (2001b). Variational analysis of the abscissa mapping for polynomials. *SIAM Journal on Control and Optimization*, *39*, 1651–1676.
- 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, 1983. Republished as Vol. 5, *Classics in Applied Mathematics*, SIAM, 1990.
- Clarke, F. H., Ledyaev, Yu. S., Stern, R. J., & Wolenski, P. R. (1998). *Nonsmooth analysis and control theory*. New York: Springer-Verlag.
- Demmel, J. W. (1987). A counterexample for two conjectures about stability. *IEEE Transactions on Automatic Control*, *32*, 340–343.
- Demmel, J. W. (1990). Nearest defective matrices and the geometry of ill-conditioning. In M. G. Cox & S. Hammerling (Eds.), *Reliable numerical computation* (pp. 35–55). Oxford: Oxford University Press.
- Eising, R. (1984). Between controllable and uncontrollable. *Systems Control Letters*, *4*, 263–264.
- Embree, M., & Trefethen, L. N. Pseudospectra gateway. <http://web.comlab.ox.ac.uk/pseudospectra>.
- Facchinei, F., & Pang, J.-S. (2003). *Finite-dimensional variational inequalities and complementarity problems, Vol. 2*. New York: Springer.
- Gu, M. (2000). New methods for estimating the distance to uncontrollability. *SIAM Journal on Matrix Analysis and Applications*, *21*, 989–1003.
- Gu, M., Mengi, E., Overton, M. L., Xia, J., & Zhu, J. (2006). Fast methods for estimating the distance to uncontrollability. *SIAM Journal on Matrix Analysis and Applications*, *28*, 477–502.
- Hare, W. L., & Lewis, A. S. (2004). Identifying active constraints via partial smoothness and prox-regularity. *Journal of Convex Analysis*, *11*, 251–266.
- Ioffe, A. D. (2006). *A Sard theorem for tame set-valued mappings*. Technion preprint.

- Kiwiel, K. C. (1985). *Minimization methods for non-differentiable functions*, Vol. 1133 of *Lecture notes in mathematics*. Berlin: Springer-Verlag.
- Kiwiel, K. C. (2007). Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization. *SIAM Journal on Optimization*, 18, 379–388.
- Kreiss, H. O. (1962). Über die Stabilitätsdefinition für Differenzengleichungen die partielle Differenzialgleichungen approximieren. *BIT*, 2, 153–181.
- Leibfritz, F. (2001). A LMI-based algorithm for designing suboptimal static H^2/H^∞ output feedback controllers. *SIAM Journal on Control and Optimization*, 39, 1711–1735.
- Leibfritz, F., & Mustafa, E. M. E. (2003). Trust region methods for solving the optimal output feedback design problem. *International Journal of Control*, 76, 501–519.
- Leibfritz, F., & Volkwein, S. (2006). Reduced order output feedback control design for pde systems using proper orthogonal decomposition and nonlinear semidefinite programming. *Linear Algebra and its Applications*, 415, 542–575.
- Lemaréchal, C. (1975). An extension of Davidon's methods to nondifferentiable optimisation. *Mathematical Programming Study*, 3, 95–109.
- Lewis, A. S. (2003). Active sets, nonsmoothness and sensitivity. *SIAM Journal on Optimization*, 13, 702–725.
- Lewis, A. S. (2003). The mathematics of eigenvalue optimisation. *Mathematical Programming*, 97, 155–176.
- Lewis, A. S., & Pang, C. H. J. (2006). *Variational analysis of pseudospectra*. Technical report. Cornell University.
- MATLAB code. Nonsmooth optimization packages: Gradient Sampling, HANSO, HIFOO. <http://cs.nyu.edu/cs/faculty/overton/software/index.html>.
- Mengi, E., & Overton, M. L. (2005). Algorithms for the computation of the pseudospectral radius and the numerical radius of a matrix. *IMA Journal of Numerical Analysis*, 25, 648–669.
- Mordukhovich, B. S. (2006). *Variational analysis and generalized differentiation. I. Basic Theory; II. Applications*. New York: Springer.
- Nesterov, Y. E., & Nemirovskii, A. S. (1994). *Interior-point polynomial algorithms in convex programming*. Philadelphia: SIAM.
- Overton, M. L. (2006). Variable metric methods for nonsmooth optimization, 2006. Presented at the international conference on numerical analysis and optimization dedicated to M.J.D. Powell's 70th Birthday.
- Qiu, L., Bernhardsson, B., Rantzer, A., Davison, E. J., Young, P. M., & Doyle, J. C. (1995). A formula for computation of the real stability radius. *Automatica*, 31, 879–890.
- Rockafellar, R. T. (1970). *Convex analysis*. Princeton, NJ: Princeton University Press.
- Rockafellar, R. T., & Wets, R.J.-B. (1998). *Variational analysis*. Berlin: Springer.
- Schramm, H., & Zowe, J. (1992). A version of the bundle idea for minimizing a nonsmooth function: conceptual idea, convergence analysis, numerical results. *SIAM Journal on Optimization*, 2, 121–152.
- Trefethen, L. N., & Embree, M. (2005). *Spectra and pseudospectra: The behavior of nonnormal matrices and operators*. Princeton, NJ: Princeton University Press.
- Van Loan, C. F. (1985). How near is a stable matrix to an unstable matrix? *Contemporary Mathematics*, 47, 465–477.
- Vlček, J., & Lukšan, L. (2001). Globally convergent variable metric method for nonconvex nondifferentiable unconstrained minimization. *Journal of Optimization Theory and Applications*, 111, 407–430.
- Watson, G. A. (1992). Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170, 33–45.
- Wright, T. EigTool. <http://www.web.comlab.ox.ac.uk/pseudospectra>.
- Wright, T. G., & Trefethen, L. N. (2002). Pseudospectra of rectangular matrices. *IMA Journal of Numerical Analysis*, 22, 501–519.
- Zhou, K., Doyle, J. C., & Glover, K. (1996). *Robust and optimal control*. Upper Saddle River, NJ: Prentice-Hall.
- Zietak, K. (1993). Subdifferentials, faces and dual matrices. *Linear Algebra and its Applications*, 185, 125–141.

Adrian S. Lewis was born in England in 1962. He is a Professor at Cornell University in the School of Operations Research and Information Engineering. Following his B.A., M.A., and Ph.D. degrees from Cambridge, and Research Fellowships at Queens' College, Cambridge and Dalhousie University, Canada, he worked in Canada at the University of Waterloo (1989–2001) and Simon Fraser University (2001–2004). He is an Associate Editor of the *SIAM Journals on Optimization* and on *Matrix Analysis and Applications*, *Mathematics of Operations Research*, and the *SIAM/MPS Book Series on Optimization*, and is a Co-Editor for *Mathematical Programming*. He received the 1995 Aisenstadt Prize, from the Canadian Centre de recherches mathématiques, the 2003 Lagrange Prize for Continuous Optimization from SIAM and the Mathematical Programming Society, and an Outstanding Paper Award from SIAM in 2005. He co-authored *Convex Analysis and Nonlinear Optimization* with J.M. Borwein. His research concerns variational analysis and nonsmooth optimization, with a particular interest in optimization problems involving eigenvalues.