Stabilization via Nonsmooth, Nonconvex Optimization

James V. Burke, Didier Henrion, Adrian S. Lewis, and Michael L. Overton

***Abstract*—Nonsmooth variational analysis and related computational methods are powerful tools that can be effectively applied to identify local minimizers of nonconvex optimization problems arising in fixed-order controller design. We support this claim by applying nonsmooth analysis and methods to a challenging "Belgian chocolate" stabilization problem posed in 1994: find a stable, minimum phase, rational controller that stabilizes a specified second-order plant. Although easily stated, this particular problem remained unsolved until 2002, when a solution was found using an eleventh-order controller. Our computational methods find a stabilizing third-order controller without difficulty, suggesting explicit formulas for the controller and for the closed loop system, which has only one pole with multiplicity 5. Furthermore, our analytical techniques prove that this controller is locally optimal in the sense that there is no nearby controller with the same order for which the closed loop system has all its poles further left in the complex plane. Although the focus of the paper is stabilization, once a stabilizing controller is obtained, the same computational techniques can be used to optimize various measures of the closed loop system, including its complex stability radius or $\mathbf{H}_\infty$ performance.**

***Index Terms*—Fixed-order controller design, nonconvex optimization, nonsmooth optimization, polynomials, stability.**

I. INTRODUCTION

FIXED-ORDER control design is a challenging problem in theory and in practice, and is considered important in a broad context ranging from the complexity of control problems to real industrial practice. We believe that nonsmooth variational analysis and computational methods have great potential for applications in this field. Three of us are developers of nonsmooth analytical techniques for functions of roots of polynomials and eigenvalues of matrices [11], [8] as well as computational methods appropriate for nonsmooth, nonconvex optimization [10]. Throughout the development of this body of work, we have been partially motivated by potential applications in control, both in theory and practice. It is only recently, however, that, we have taken explicit steps in this direction [9], [7]. One of the purposes of this paper is to provide a generally accessible introduction to our techniques.

In this paper, we focus specifically on stabilization. The most basic requirement of a controlled system is that the closed loop system should be stable, yet the nonconvexity of the cone of stable polynomials makes stabilization inherently difficult. Here by a stable polynomial we mean one whose roots[1] are in the open left half of the complex plane, as is relevant for continuous-time dynamical systems, but all our techniques are equally well suited to discrete-time systems where the stability region is the open unit disk. Similarly, the cone of stable matrices is nonconvex, where by a stable matrix, we mean one whose eigenvalues are in the open left half-plane. Because of this nonconvexity, incorporating stability criteria into an optimization problem, whether as part of the objective or in the constraints, normally leads to a nonconvex, and indeed typically also nonsmooth, optimization problem. Such problems are often tackled by introducing Lyapunov matrix variables, leading to a new optimization problem with bilinear matrix inequality constraints that may, or may not, be easier to solve than the original problem. In contrast, in our work we tackle nonsmooth, nonconvex optimization problems arising from stabilization objectives directly.

We believe that our techniques are quite broadly applicable. But we also believe that it is very useful to focus on a specific, challenging example, to demonstrate the potential of the approach. More than a decade ago, Blondel [2, p. 150] [3] offered a prize of a kilogram of Belgian chocolate for the solution of the following stabilization problem.

Problem 1.1: Let $a(s) = s^2 - 2\delta s + 1$ and $b(s) = s^2 - 1$. Find the range of real values for δ for which there exist stable polynomials $x(s)$ and $y(s)$ with $\deg(x) \geq \deg(y)$ such that $ax + by$ is stable. Equivalently, in the language of control, find the range of real values for δ for which there exists a stable, minimum phase, proper rational controller y/x (one whose poles and zeros are in the open left-half plane) that stabilizes the plant b/a (for which the closed-loop transfer function $by/(ax + by)$ has all its poles in the open left-half plane).

Blondel also offered a kilo of Belgian chocolate for a solution of a special case.

[1] We speak of roots, not zeros, of polynomials, to avoid confusion with zeros of a transfer function.

Manuscript received February 28, 2005; revised November 22, 2005 and March 25, 2006. Recommended by Associate Editor A. Astolfi. The work of J. V. Burke was supported in part by National Science Foundation under Grant DMS-0505712. The work of D. Henrion was supported in part by Project 102/05/0011 of the Grant Agency of the Czech Republic and by Project ME 698/2003 of the Ministry of Education of the Czech Republic. The work of A. S. Lewis was supported in part by National Science Foundation under Grant DMS-0504032. The work of M. L. Overton was supported in part by National Science Foundation under Grant DMS-0412049, and in part by Université Paul Sabatier, Toulouse, France.

J. V. Burke is with the Department of Mathematics, University of Washington, Seattle, WA 98195 USA (e-mail: burke@math.washington.edu).

D. Henrion is with the LAAS-CNRS, 31077 Toulouse, France, and also with the Department of Control Engineering, Czech Technical University, 16627 Prague, Czech Republic (e-mail: henrion@laas.fr).

A. S. Lewis is with the School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: aslewis@orie.cornell.edu).

M. L. Overton is with the Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 USA (e-mail: overton@cs.nyu.edu).

Color versions of Figs. 1–5 are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2006.884944

Problem 1.2: Show whether or not 0.9 is in the range of values for δ for which stabilization is possible.

For $\delta = 1, ax + by$ is not stable for any x and y because $s-1$ is a common factor of a and b. Conversely, stabilization is easy for $\delta \leq 0.5$, say. But for $\delta = 0.9$ stabilization is surprisingly difficult. Problem 1.2 went unsolved for eight years until Patel *et al.* [13] found a solution with $\deg(x) = \deg(y) = 11$, using a randomized search method.

Problem 1.1 remains unsolved. However, it follows from results in [5] that there exists a number δ^* such that stabilization is possible for all $\delta < \delta^*$ and is not possible for $\delta > \delta^*$. Thus, Problem 1.1 reduces to determining δ^*. Results on the range of analytic functions proved in [6] imply that $\delta^* < 0.99998$ [3], [4], and it is known from the experiments reported in [13] that $\delta^* > 0.937$.

In this paper, we give a solution to Problem 1.2 with far lower degree than had previously been thought possible. Specifically, we show stabilization is possible with $\deg(x) = 3$ and $\deg(y) = 0$ for

$$\delta < \bar{\delta} = \frac{\sqrt{2+\sqrt{2}}}{2} = 0.924\ldots$$

and with $\deg(x) = 4$ and $\deg(y) = 0$ for

$$\delta < \tilde{\delta} = \frac{\sqrt{10+2\sqrt{5}}}{4} = 0.951\ldots$$

Furthermore, the controllers have a systematic structure that we describe in detail. Stabilization is still possible for some $\delta > \tilde{\delta}$, but becomes much more difficult as the structure of the controllers changes. We still do not know the answer to Problem 1.1, but we know that $\delta^* > 0.96$.

These stabilizing controllers were obtained by application of a new numerical method for nonsmooth, nonconvex optimization called *gradient sampling*. The controllers that we found are locally optimal in a specific sense. Although they were found experimentally, we *prove* their local optimality for the case $\deg(x) = 3$. In particular, for the boundary case $\delta = \bar{\delta}$, we exhibit a stable cubic polynomial x and scalar y for which $ax + by$ is exactly the monomial s^5, and for which any small perturbation to x or y moves at least one root of $ax + by$ into the open right-half plane. Our theoretical analysis builds on recent work on nonsmooth analysis of the cone of stable polynomials.

We believe that our work is significant not because the Belgian chocolate problem is important by itself, but because the solution of a challenging problem by new techniques suggests that the same ideas should be useful in a far broader context and presents an illustrative and intuitive example that can be easily understood. Indeed, the reason that the chocolate problem is so difficult is that for δ near 1, a and b have unstable roots that are nearly equal to each other, or in the language of control, that an unstable pole and unstable zero of b/a (the transfer function of the open loop system) nearly cancel each other. Exactly this phenomenon arises in physically relevant engineering problems, such as the X-29 prototype aircraft design problem or Klein's bicycle design problem mentioned in [1]. It is our hope, as was Blondel's, that a detailed analysis of the chocolate problem will provide insight that is useful in many other contexts. We note also that our techniques are not limited to polynomials, and that indeed much of our work is oriented towards stabilization of matrices. Furthermore, our interests are not limited to stabilization. We are currently developing a MATLAB toolbox for fixed-order controller design which allows specification of various optimization objectives, including $\mathbf{H}_\infty$ performance.

The remainder of this paper is organized as follows. In Section II, we discuss our computational approach to the chocolate problem and present numerical results. In Section III, we present our local optimality analysis, using key theoretical properties of the cone of stable polynomials. In Section IV, we make some concluding remarks.

II. Experimental Analysis

Let $\mathbf{P}^n$ (respectively, $\mathbf{P}^n_R$) denote the space of polynomials with complex (respectively real) coefficients and with degree less than or equal to n, and let $\mathbf{MP}^n$ and $\mathbf{MP}^n_R$ denote the corresponding subsets of monic polynomials with degree n. For $p \in \mathbf{P}^n$, let $\alpha(p)$ denote the *abscissa* of p

$$\alpha(p) = \max\{\mathrm{Re}\, s : p(s) = 0\}$$

(interpreted as $-\infty$ if p is a nonzero constant). Problem 1.2 asks for what range of δ do there exist polynomials x and y with $\deg(x) \geq \deg(y)$ such that $\alpha(xy(ax+by)) < 0$, and Problem 1.2 addresses the case $\delta = 0.9$.

Now, consider the problem of choosing polynomials x and y to minimize $\alpha(xy(ax+by))$. For convenience we restrict x, but not y, to be monic. Thus, we consider the problem: For fixed real δ and integers n and m with $m \leq n+1$, minimize $\alpha(xy(ax+by))$ over $x \in \mathbf{MP}^{n+1}_R$ and $y \in \mathbf{P}^m_R$. This is a nonconvex optimization problem in $n+m+2$ real variables.

For a given x and y, we say that a root of, or $ax + by$ is an *active* root if its real part equals $\alpha(xy(ax+by))$. The objective function α is, as we shall see, typically not differentiable at local minimizers, either because there are two or more active roots, or because there is a multiple active root, or both. Furthermore, it is not the case that α is an ordinary "max function," that is the pointwise maximum of a finite number of smooth functions. On the contrary, α is not even Lipschitz because of the possibility of multiple roots.

Reliable software for both smooth, nonconvex optimization and for nonsmooth, convex optimization is widely available, but there are not many options for tackling nonsmooth, nonconvex optimization problems. We have developed a method based on *gradient sampling* that is very effective in practice and for which a local convergence theory has been established [10], [12]. This method is intended for finding local minimizers of functions f that are continuous and for which the gradient exists and is readily computable *almost everywhere* on the design parameter space, even though the gradient may (and often does) fail to exist at a local optimizer. Briefly, the method generates a sequence of points in the parameter space, say $\mathbf{R}^N$, as follows. Given ξ^ν, the gradient ∇f is computed at ξ^ν and at randomly generated points near ξ^ν within a sampling diameter η, and the convex combination of these gradients with smallest 2-norm, say d, is computed

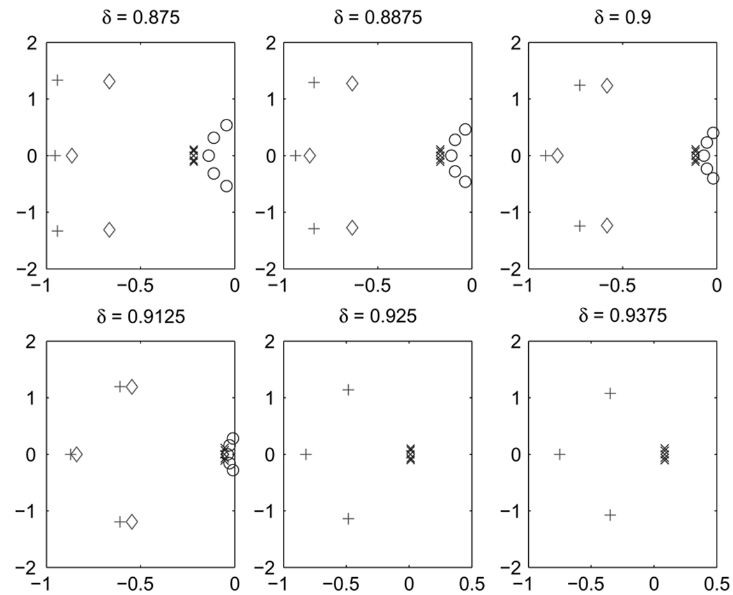


Fig. 1. Optimal roots for $\deg(x) = 3$, plotted in the complex plane for various values of δ. Crosses and plus signs are, respectively, roots of $ax + by$ and x when the abscissa $\alpha(x(ax + by))$ is minimized over monic cubic x and scalar y. Circles and diamonds are, respectively, roots of $ax + by$ and x when stabilization is achieved and the stability radius $\min(\beta(x), \beta(ax + by))$ is maximized. The third panel shows that Problem 1.2 is solved by an order 3 controller.

by solving a quadratic program. One should view $-d$ as a kind of stabilized steepest descent direction. A line search is then used to obtain $\xi^{\nu+1} = \xi^{\nu} - td/||d||$, with $f(x^{\nu+1}) < f(x^{\nu})$, for some $t \leq 1$. If $||d||$ is below a prescribed tolerance, or a prescribed iteration limit is exceeded, the sampling diameter η is reduced by a prescribed factor, and the process is repeated. For the numerical examples to be discussed, we used sampling diameters $10^{-j}, j = 1, \ldots, 6$, with a maximum of 100 iterates per sampling diameter and a tolerance 10^{-6} for $||d||$, and we set the number of randomly generated sample points to $2N$ (twice the number of design variables) per iterate. Besides its simplicity and wide applicability, a particularly appealing feature of the gradient sampling algorithm is that it provides approximate "optimality certificates": $||d||$ being small for a small sampling diameter η suggests that a local minimizer has been approximated. A MATLAB implementation of the gradient sampling algorithm is freely available.[2]

The abscissa of a polynomial is the spectral abscissa (largest of the real parts of the eigenvalues) of its companion matrix, and so $\alpha(xy(ax + by))$ is the spectral abscissa of a block diagonal matrix, with blocks that are companion matrices for x, y and $ax+by$ respectively. Computing the gradient of the spectral abscissa in matrix space is convenient, because the gradient of a simple eigenvalue λ (with respect to the real trace inner product $\langle AB \rangle = \mathrm{Retr}(A^*B)$ is the rank-one matrix uv^*, where u and v are respectively the left and right eigenvectors corresponding to λ, normalized so that $u^*v = 1$. The ordinary chain rule then easily yields the gradient of the spectral abscissa with respect to the relevant coefficients of x and y when it exists, which is exactly when there is only one real eigenvalue or one conjugate pair whose real part equals the spectral abscissa, and that eigenvalue or conjugate pair is simple. The gradient of α on polynomial space depends on the inner product we choose: nothing *a priori* in the problem defines our choice. For our numerical experiments, simply for computational convenience, we define the inner product to coincide, for monic polynomials, with the inner product of the corresponding companion matrices.

We now summarize the numerical results that we obtained when we applied the gradient sampling algorithm to minimize $\alpha(xy(ax + by))$ for various values of δ, n and m. We began with $\delta = 0.9$ and $m = n + 1$. We soon found negative optimal values for $\alpha(xy(ax + by))$ for small values of n, thus solving Problem 1.2. Furthermore, we observed that the leading coefficient of the non-monic polynomial y converged to zero as the apparent local optimizer was approached. This led to numerical difficulties since constructing a companion matrix requires normalizing the polynomial to be monic; hence, the norm of the companion matrix blows up as the leading coefficient of the polynomial goes to zero. These difficulties were avoided by explicitly reducing m, the degree of y and the size of its corresponding companion matrix block, when it was realized that the leading coefficient was converging to zero, restarting the optimization in a smaller parameter space. This phenomenon was observed again for smaller values of m, and we soon became quite confident that, for $\delta = 0.9$ and $n \leq 3$, the function $\alpha(xy(ax + by))$ is minimized when $m = 0$, that is, the polynomial y is a scalar, so $\alpha(xy(ax + by)) = \alpha(x(ax + by))$. Further experimentation showed that $\alpha(x(ax + by))$ could be reduced to a negative value when $n = 2(\deg(x) = 3)$, but not when $n < 2$. Furthermore, the structure of the minimizer is striking: the polynomial $ax + by$ evidently has only one (distinct) root, which, for $n = 2$, is a quintuple root (multiplicity 5), since a is quadratic. This is the only active root; the roots of x have smaller real part. The structure is clearly seen in the third panel of Fig. 1, where the roots of the polynomials $ax + by$ and x obtained by minimizing $\alpha(x(ax + by))$ are shown as crosses and plus signs respectively; disregard the circles and diamonds in the plot for the moment. The five roots of $ax + by$ are very close to each

[2]See http://www.cs.nyu.edu/overton/papers/gradsamp/alg/.

other, indicating the likelihood of coalescence to a single root for the exact local optimizer, while the three roots of x are well to their left. The other panels in the same figure show similar results for δ ranging from 0.875 to 0.9375 (still with $n = 2$ and $m = 0$). In all six cases, the approximately optimal $ax + by$ has a nearly multiple (quintuple) root, but for the two largest values of δ, this root is to the right of the imaginary axis, causing a change in the automatic scaling of the horizontal axes in the last two panels, and indicating that stabilization was not achieved. For $\delta = 0.9$, the approximately optimal abscissa was attained by $y = 1.8867980$ and

$$x(s) = s^3 + 2.3628818\, s^2 + 3.3978859\, s + 1.8868496.$$

This was the best result found in 100 runs of the gradient sampling algorithm starting from randomly generated starting points.

The appearance of the multiple root at a local optimizer is a very interesting phenomenon that we discuss further in the next section. However, it is well known that the roots of a polynomial with a nominally stable multiple root are highly sensitive to perturbation and, therefore, such a polynomial has poor stability properties in a practical setting. For this reason, we also consider a more robust measure than the abscissa, namely the *complex stability radius* of a monic polynomial in $\mathbf{MP}^k$

$$\beta(p) = \sup\{\epsilon : \alpha(q) < 0, \text{ for all} q \in \mathbf{MP}^k \text{ with } \|p - q\| \le \epsilon\}.$$

Here, the norm is just the 2-norm of the coefficient vector. The quantity $\beta(p)$ can be computed by standard software.[3] It is the reciprocal of the $\mathbf{H}_\infty$ norm for the state–space realization (A, B, C, D), where A, B, C and D are respectively the companion matrix for p (with its negated coefficients in the first column), the identity matrix, the first row of the identity matrix, and a zero row, since then $C(sI - A)^{-1}B + D = [s^{k-1} \ldots s^2 s 1]/p(s)$. Like the abscissa, the complex stability radius is differentiable almost everywhere and its gradient is easily computed [9].

When $\deg(y) = 0$, a natural maximization objective is

$$\tilde{\beta}(x, y) = \min(\beta(x), \beta(ax + by)).$$

We applied the gradient sampling algorithm to minimize $-\tilde{\beta}(x, y)$ over $x \in \mathbf{MP}_R^{n+1}$ and $y \in \mathbf{P}_R^0$, using the same values for δ and n as earlier. A key point is that the complex stability radius is identically zero in a small neighborhood of any polynomial with a root in the open right half-plane. We therefore used the locally optimal x and y found by minimizing α (as already described above) to initialize minimization of $-\tilde{\beta}$ over the same parameter space. This optimization produced locally optimal x and y for which the roots of $ax + by$ are well separated; for $n = 2$ and $m = 0$, the roots of the optimal $ax + by$ and x are shown as circles and diamonds respectively in the first four panels of Fig. 1. For the two largest values of δ, stabilization was not achieved, so optimization of the stability radius could not be initialized.

[3]We used the norm function in the MATLAB Control System Toolbox.

In order to achieve stabilization for larger values of δ, we increased n to $3(\deg(x) = 4)$. Results are shown in Fig. 2: The optimal y is again a scalar, and for all δ for which stabilization is achieved, the optimal $ax + by$ apparently has only one root, which is hextuple (multiplicity 6), with the roots of x inactive. We were able to achieve stabilization up to $\delta = 0.94375$, for which the approximately optimal abscissa was attained by $y = 2.0465513$ and

$$x(s) = s^4 + 2.1853347\, s^3 + 3.1991472\, s^2 + 3.8629224\, s + 2.0465529.$$

As in Fig. 1, the circles and diamonds in the first four panels of Fig. 2 show the roots of the optimal $ax+by$ and x when the complex stability radius is optimized instead of the abscissa; these roots are well separated. As δ increases, a complex conjugate pair of roots of x moves to the right, and we see that this pair becomes active, having the same real part as the hextuple root of $ax + by$, at approximately the same critical value of δ beyond which stabilization is not possible. In other words, the trajectories of the rightmost conjugate pair of roots of the optimizing x and hextuple root of the corresponding $ax + by$ as a function of δ reach the imaginary axis at approximately the same value of δ (approximately 0.95). In fact, as discussed at the end of Section III-B, these events occur at exactly the same critical value $\tilde{\delta} = 0.951\ldots$. Beyond this value the structure of the local optimizer changes; a conjugate pair of roots of x is active and the hextuple root of $ax + by$ splits into a quintuple root and a simple root. This simple root of $ax + by$ and a corresponding root of x shoot off together into the left half-plane (see the final panel of Fig. 2, for which the automatic scaling of the horizontal axis changes abruptly in order to fit all the roots in the panel).

In order to achieve stabilization for larger δ, we increased n to 4 $(\deg(x) = 5)$, with $\deg(y)$ still set to 0. The results are shown in Fig. 3. The structure of the optimizers remains consistent with the final panel of Fig. 2: A conjugate pair of roots of x is active, and $ax + by$, which has degree 7, has one active hextuple root. Both $ax + by$ and x have a simple root well into the left half-plane, causing a change in the automatic scaling of the real axis, and the other roots all appear to be very close to the imaginary axis as a result. Stabilization was achieved for $\delta = 0.95$, but not for larger δ. Stabilization was possible for somewhat larger values by further raising the degrees of both x and y, for example, with $\delta = 0.96, n = 7(\deg(x) = 8)$ and $m = \deg(y) = 2$, but the numerical optimization problem is much more difficult than it is for smaller values of δ.

Figs. 4 and 5 summarize the numerical experiments, respectively showing the optimal values of the abscissa $\alpha(x(ax + by))$ and stability radius $\tilde{\beta}(x, y)$ (the latter on a log scale) as a function of δ, for $\deg(x)$ ranging from 1 to 5 (n from 0 to 4), with $m = \deg(y) = 0$. Note the way the underlying curves are regularly spaced for $\delta \le 0.95$, with each increment in $\deg(x)$ providing a substantial increase in the range of δ for which stabilization is possible, while for $\delta > 0.95$, little is gained by increasing $\deg(x)$. This is a consequence of the change in structure of the optimal solution when δ increases beyond $\tilde{\delta} = 0.951\ldots$

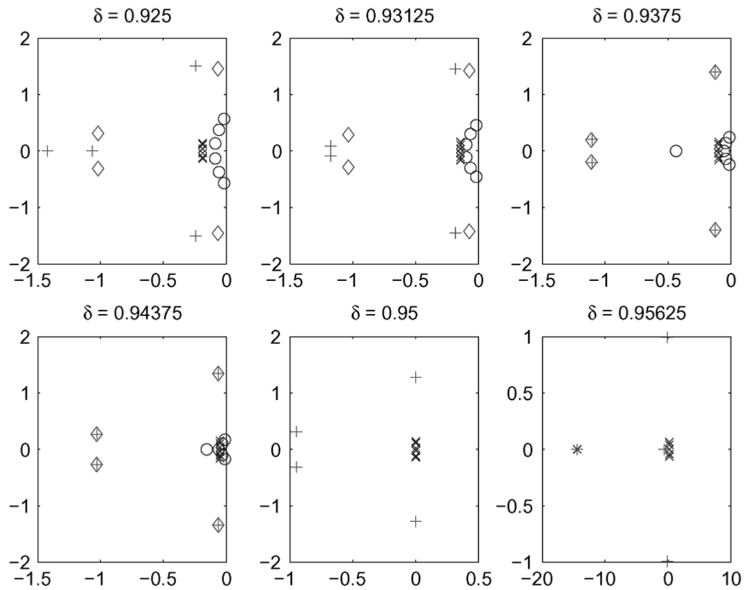


Fig. 2. Optimal roots for $\deg(x) = 4$, plotted in the complex plane for various values of δ. Crosses and plus signs are, respectively, roots of $ax + by$ and x when the abscissa $\alpha(x(ax + by))$ is minimized over monic quartic x and scalar y. Circles and diamonds are, respectively, roots of $ax + by$ and x when stabilization is achieved and the stability radius $\min(\beta(x), \beta(ax + by))$ is maximized. The structure of the abscissa optimizer changes in the last panel, causing a root of $ax + by$ and of x to shoot off into the left half-plane.

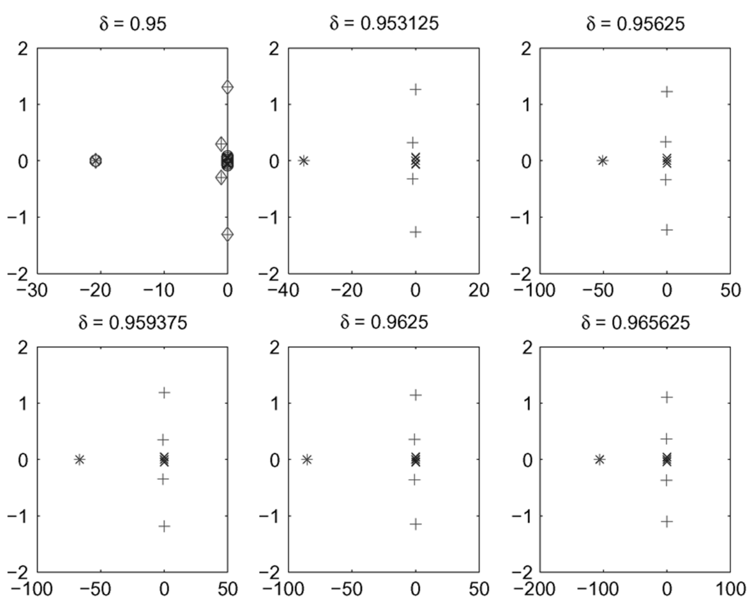


Fig. 3. Optimal roots for $\deg(x) = 5$, plotted in the complex plane for various values of δ. Crosses and plus signs are, respectively, roots of $ax + by$ and x when the abscissa $\alpha(x(ax + by))$ is minimized over monic quintic x and scalar y. Stabilization is achieved only in the first panel. The structure of the optimizer is consistent with the final panel of Fig. 3, with a root of $ax + by$ and of x moving further into the left half-plane as δ is increased.

III. Theoretical Analysis

We now present a theoretical analysis inspired by the experimental results reported in the previous section. We observed that, for $0 \le n \le 3$ and for certain ranges of δ that depend on n, local minimizers[4] (x^δ, y^δ) of $\alpha(x(ax + by))$ on $\mathbf{MP}_R^{n+1} \times \mathbf{P}_R^0$ apparently have a very special property, namely, that $ax^\delta + by^\delta$ has only one (distinct) root. Since the polynomial $ax + by$ is

$$s \mapsto (s^2 - 2\delta s + 1)x(s) + (s^2 - 1)y(s) \tag{3.1}$$

we can write this observed optimality property explicitly as the polynomial identity

$$(s^2 - 2\delta s + 1)\left(s^{n+1} + \sum_{k=0}^{n} w_k^\delta s^k\right) + (s^2 - 1)v_\delta \equiv (s - z_\delta)^{n+3} \tag{3.2}$$

[4]The use of a superscript δ indicating the dependence of the minimizer on δ should not be confused with use of superscripts to mean exponentiation elsewhere.

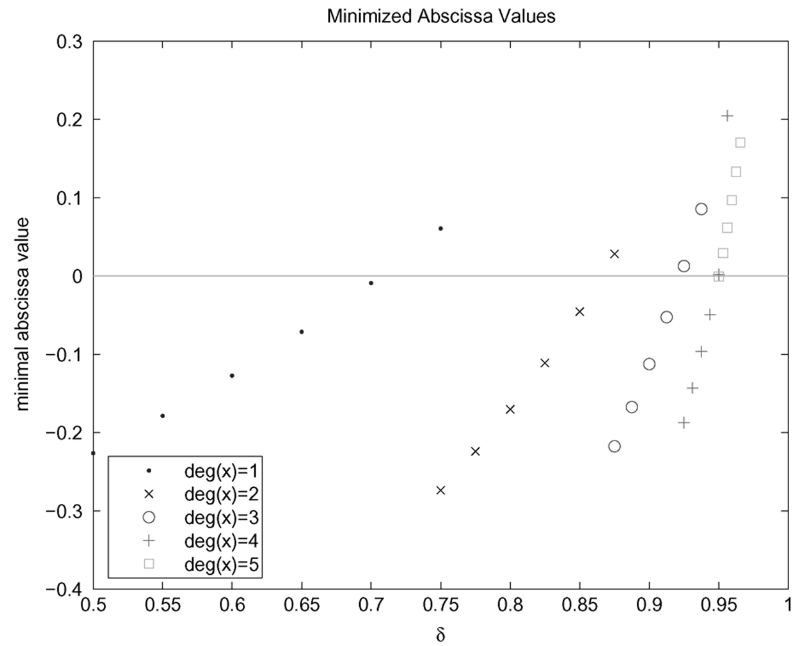


Fig. 4. Summary of minimized abscissa values. The pattern changes abruptly as δ is increased beyond 0.95, reflecting a change in the structure of the optimizing solution.

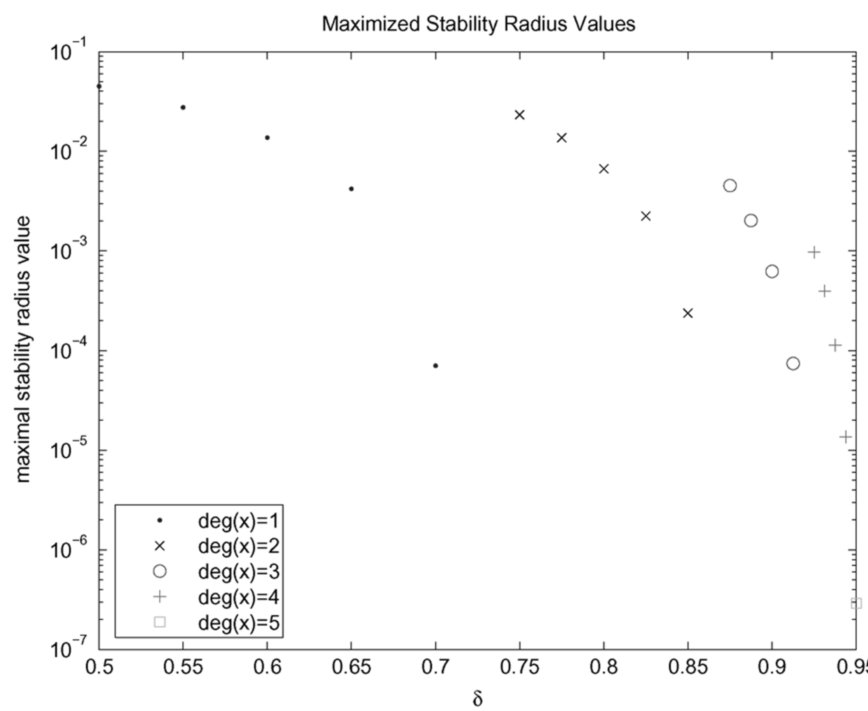


Fig. 5. Summary of maximized stability radius values for the cases that stabilization was achieved.

where z_δ is the root, w_k^δ are the coefficients of x^δ and v_δ is the constant (and only) coefficient of y^δ. The dependence of the coefficients and the root on δ is expressed explicitly, but the dependence on n is suppressed. Using the identity (3.2), it is not difficult to derive, for $n = 0, \ldots, 3$, a formula for the critical value δ for which $z_\delta = 0$ and to observe that for smaller values of δ, we have z_δ less than 0 and greater than the real part of any root of x^δ. The real contribution of our analysis is a proof that, for δ sufficiently near its critical value, (x^δ, y^δ) is indeed strictly locally optimal, which we present for $n = 0$ (the simplest case) and $n = 2$ (covering the simplest solution to Problem 1.2).

In what follows we make use of the terminology *subdifferential* (set of subgradients), *horizon subdifferential* (set of horizon subgradients) and *subdifferentially regular*, all standard notions of nonsmooth analysis, as is the nonsmooth chain rule we use later; see [14, Ch. 8] and [8].

Essential to our local optimality analysis is the following result of Burke and Overton [11, pp. 1668–1673]. The result was originally stated for the abscissa map α on the affine space $\mathbf{MP}^{n+1}$. However, it is more convenient to work with a related map on the linear space $\mathbf{P}^n$, namely

$$\gamma(p) = \max\{\mathrm{Re}s : s^{n+1} + p(s) = 0\}. \tag{3.3}$$

We can identify $\mathbf{P}^n$ with the Euclidean space $\mathbf{C}^{n+1}$, with the inner product $\langle uv \rangle = \mathrm{Re}\sum_{j=0}^{n} u_j^* v_j$. For $j = 0, 1, 2, \ldots$, we define the polynomial e_j by

$$e_j(s) = s^j.$$

Theorem 3.4 (Abscissa Subdifferential): The map γ defined in (3.3) is everywhere subdifferentially regular. The subdifferential and horizon subdifferential at 0 are, respectively, given by

$$\partial\gamma(0) = \left\{ \sum_j c_j e_j : c_n = -\frac{1}{n+1}, \mathrm{Re}c_{n-1} \leq 0 \right\}$$
$$\partial^\infty\gamma(0) = \left\{ \sum_j c_j e_j : c_n = 0, \mathrm{Re}c_{n-1} \leq 0 \right\}.$$

A. The Simplest Case

In the case $n = 0$, the polynomial (3.1) reduces to

$$s \mapsto (s^2 - 2\delta s + 1)(s + w) + (s^2 - 1)v \tag{3.5}$$

writing $x(s) = s + w$ and $y(s) = v$. Identity (3.2) reduces to

$$(s^2 - 2\delta s + 1)(s + w_\delta) + (s^2 - 1)v_\delta \equiv (s - z_\delta)^3 \tag{3.6}$$

where we have abbreviated w_0^δ to w_δ. Multiplying out factors and equating terms leads to the following result.

Lemma 3.7 (Condition for Triple Root): Identity (3.6) holds if and only if

$$w_\delta = \delta - \frac{3}{2}z_\delta - \frac{1}{2}z_\delta^3$$
$$v_\delta = \delta - \frac{3}{2}z_\delta + \frac{1}{2}z_\delta^3$$

and z_δ solves the equation

$$\delta z^3 - 3z^2 + 3\delta z + 1 - 2\delta^2 = 0. \tag{3.8}$$

The next lemma follows from the implicit function theorem. For technical reasons associated with the nonsmooth chain rule we use, we will in fact allow w and v to be *complex* variables. Consequently, we may as well also allow the parameter δ to be complex.

Lemma 3.9 (Definition of z_δ, Linear Case): For complex δ near $\hat{\delta} = 1/\sqrt{2}$, the (3.8) has a unique solution z_δ near 0, depending analytically on δ. For real δ near $\hat{\delta}$, the solution z_δ is real, and increases strictly with δ, with $z_{\hat{\delta}} = 0$.

Equipped with these lemmas, we can proceed to our main result for the case $n = 0$.

Theorem 3.10 (Minimizing the Abscissa, Linear Case): Consider the problem of choosing a monic linear polynomial x and scalar y to minimize the maximum of the real parts of the roots of the polynomial $x(ax + by)$, where $a(s) = s^2 - 2\delta s + 1$ and $b(s) = s^2 - 1$. For all complex δ near $\hat{\delta} = 1/\sqrt{2}$ this problem has a strict local minimizer at the unique pair (x, y) for which $ax + by$ has a triple root near 0. Furthermore, x is stable, and for δ real, $ax + by$ is stable if and only if $\delta < \hat{\delta}$.

Proof: Define z_δ as in Lemma 3.9. The unique pair (x, y) in the theorem statement is therefore given by $x(s) = x^\delta(s) = s + w_\delta$ and $y = v_\delta$, where w_δ and v_δ are given by Lemma 3.7. Notice $w_{\hat{\delta}} = \hat{\delta} > 0$, so x^δ is stable and $\alpha(x(ax + by)) = \alpha(ax + by)$ for all (x, y) near (x^δ, v_δ).

Consider the polynomial (3.5), and make the following changes of variables:

$$t = s - z_\delta \quad q = w - w_\delta \quad r = v - v_\delta.$$

With this notation, a calculation shows that minimizing $\alpha(ax + by)$ is equivalent to minimizing the abscissa of the polynomial

$$t \mapsto t^3 + A_\delta(q, r)(t)$$

where the linear map $A_\delta : \mathbf{C}^2 \to \mathbf{P}^2$ is given by

$$A_\delta(q, r)(t) = q\left(t^2 + 2(z_\delta - \delta)t + \left(z_\delta^2 - 2\delta z_\delta + 1\right)\right) + r\left(t^2 + 2z_\delta t + \left(z_\delta^2 - 1\right)\right).$$

We, therefore, need to prove that the point $(0, 0)$ is a strict local minimizer of the composite function $\gamma \circ A_\delta$, where the function γ is defined by (3.3).

The adjoint map $A_\delta^* : \mathbf{P}^2 \to \mathbf{C}^2$ is given by

$$A_\delta^*\left(\sum_j c_j e_j\right) = \begin{bmatrix} c_2 + 2(z_\delta - \delta)c_1 + (z_\delta^2 - 2\delta z_\delta + 1)c_0 \\ c_2 + 2z_\delta c_1 + \left(z_\delta^2 - 1\right)c_0 \end{bmatrix}$$

and, in particular

$$A_{\hat{\delta}}^*\left(\sum_j c_j e_j\right) = \begin{bmatrix} c_2 - 2\hat{\delta}c_1 + c_0 \\ c_2 - c_0 \end{bmatrix}.$$

Notice that when $\delta = \hat{\delta}$ we have the implication

$$A_{\hat{\delta}}^*\left(\sum_j c_j e_j\right) = 0 \quad \text{and} \quad c_2 = 0$$
$$\Rightarrow c = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = 0.$$

Since the map A_δ depends continuously on δ, the same implication holds for all δ near $\hat{\delta}$. We are now in a position to apply Theorem 3.4 (abscissa subdifferential). First, we observe the constraint qualification

$$N(A_\delta^*) \cap \partial^\infty\gamma(0) = \{0\}$$

where N denotes null space. Consequently we can apply the nonsmooth chain rule [8, Lemma 4.4] to deduce that the composite function $\gamma \circ A_\delta$ is subdifferentially regular at zero, with subdifferential

$$\partial(\gamma \circ A_\delta)(0) = A_\delta^* \partial\gamma(0) = \left\{ \begin{bmatrix} -\frac{1}{3} + 2(z_\delta - \delta)c_1 + \left(z_\delta^2 - 2\delta z_\delta + 1\right)c_0 \\ -\frac{1}{3} + 2z_\delta c_1 + \left(z_\delta^2 - 1\right)c_0 \end{bmatrix} : \times \mathrm{Re} c_1 \leq 0, c_0 \in \mathbf{C} \right\}.$$

The matrix

$$B_\delta = \begin{bmatrix} 2z_\delta - 2\delta & z_\delta^2 - 2\delta z_\delta + 1 \\ 2z_\delta & z_\delta^2 - 1 \end{bmatrix}$$

depends continuously on δ, and, since $z_{\hat{\delta}} = 0$, the vector

$$\begin{bmatrix} c_1 \\ c_0 \end{bmatrix} = B_{\hat{\delta}}^{-1} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} = -\begin{bmatrix} \frac{1}{3\hat{\delta}} \\ \frac{1}{3} \end{bmatrix}$$

satisfies $\mathrm{Re} c_1 < 0$. Hence, by continuity, the subdifferential $\partial(\gamma \circ A_\delta)(0)$ contains zero in its interior. Together with subdifferential regularity, this implies [8, Prop. 4.3] that the function $\gamma \circ A_\delta$ has a "sharp" local minimizer at zero: it grows at least linearly at this point. □

This result proves that any small perturbation to the locally optimal polynomials x or y splits the triple root of $ax + by$ and moves at least one root strictly to the right, and into the open right half-plane when $\delta = \hat{\delta}$. A simple argument based on the Routh–Hurwitz conditions shows that stabilization by a first-order controller is not possible when $\delta > \hat{\delta}$, thus providing a global optimality certificate for our local optimizer when $\delta = \hat{\delta}$.

B. The Chocolate Problem

We now turn to the case $n = 2$, providing what is almost certainly the simplest possible solution to Problem 1.2. The polynomial $ax + by$ is now

$$s \mapsto (s^2 - 2\delta s + 1)\left(s^3 + \sum_{k=0}^{2} w_k s^k\right) + (s^2 - 1)v. \quad (3.11)$$

The identity (3.2) becomes

$$(s^2 - 2\delta s + 1)\left(s^3 + \sum_{k=0}^{2} w_k^\delta s^k\right) + (s^2 - 1)v_\delta \equiv (s - z_\delta)^5. \quad (3.12)$$

Multiplying out factors and equating terms leads to an analogue of Lemma 3.7 with explicit formulas for w_k^δ and v_δ; for brevity, we omit the details and proceed to the following result. The proof uses the implicit function theorem.

Lemma 3.13 (Definition of z_δ, Cubic Case): For complex δ near $\bar{\delta} = (1/2)\sqrt{2 + \sqrt{2}}$ the equation

$$\delta z^5 - 5z^4 + 10\delta z^3 + 10(1 - 2\delta^2)z^2 + 5\delta(4\delta^2 - 3)z + (-1 + 8\delta^2 - 8\delta^4) = 0$$

has a unique solution z_δ near 0, depending analytically on δ. For real δ near $\bar{\delta}$, the solution z_δ is real, and increases strictly with δ, with $z_{\bar{\delta}} = 0$. Furthermore, there exist analytic functions of δ, namely $w^\delta \in \mathbf{C}^3$ and $v_\delta \in \mathbf{C}$, for which the identity (3.12) holds. Finally, the polynomial

$$s^3 + \sum_{k=0}^{2} w_k^\delta s^k$$

is stable, with

$$w^{\bar{\delta}} = \left[2\bar{\delta}, 4\bar{\delta}^2 - 1, 2\bar{\delta} - \frac{1}{2\bar{\delta}}\right]^T$$

and

$$v_{\bar{\delta}} = 2\bar{\delta} - \frac{1}{2\bar{\delta}}.$$

We now present the main result of this paper.

Theorem 3.14 (Minimizing the Abscissa, Cubic Case): Consider the problem of choosing a monic cubic polynomial x and a scalar y to minimize the maximum of the real parts of the roots of the polynomial $x(ax + by)$, where $a(s) = s^2 - 2\delta s + 1$ and $b(s) = s^2 - 1$. For all complex δ near the value $\bar{\delta} = (1/2)\sqrt{2 + \sqrt{2}}$ this problem has a strict local minimizer at the unique pair (x, y) for which $ax + by$ has a quintuple root near 0. Furthermore, x is stable, and for δ real, $ax + by$ is stable if and only if $\delta < \bar{\delta}$.

Proof: Define z_δ as in Lemma 3.13. The unique pair (x, y) in the theorem statement is given by $x(s) = x^\delta(s) = s^3 + \sum_{k=0}^{2} w_k^\delta s^k$ and $y = v_\delta$. By the lemma, the polynomial $x^\delta(s)$ is stable and $\alpha(x(ax + by)) = \alpha(ax + by)$ for all (x, y) near (x^δ, v_δ).

Therefore, we wish to check that, for all complex δ close to $\bar{\delta}$, choosing $x = x^\delta$ and $v = v_\delta$ gives a strict local minimum for $\alpha(ax + by)$. To verify this, we first check the case $\delta = \bar{\delta}$, and then, as in the previous section, but with fewer details, appeal to a continuity argument.

We make the change of variables

$$q = w - w^{\bar{\delta}} \in \mathbf{C}^3, r = v - v_{\bar{\delta}} \in \mathbf{C}.$$

With this notation, minimizing $\alpha(ax + by)$ is equivalent to minimizing the abscissa of the polynomial

$$s \mapsto s^5 + A_{\bar{\delta}}(q, r)(s)$$

where

$$A_{\bar{\delta}}(q, r)(s) = (s^2 - 2\bar{\delta}s + 1) \times (q_2 s^2 + q_1 s + q_0) + (s^2 - 1)r.$$

So, we wish to show that $(q, r) = (0, 0)$ is a strict local minimizer of the function $\gamma \circ A_{\bar{\delta}}$. A calculation shows that the adjoint map $A_{\bar{\delta}}^* : \mathbf{P}^4 \to \mathbf{C}^4$ is given by

$$A_{\bar{\delta}}^*\left(\sum_{j=0}^{4} c_j e_j\right) = \begin{bmatrix} c_2 - 2\bar{\delta}c_1 + c_0 \\ c_3 - 2\bar{\delta}c_2 + c_1 \\ c_4 - 2\bar{\delta}c_3 + c_2 \\ c_2 - c_0 \end{bmatrix}.$$

We have

$$A_{\bar\delta}^{*}\left(\sum_{j=0}^{4} c_j e_j\right) = 0 \quad \text{and} \quad c_4 = 0 \Rightarrow c = 0.$$

We now use Theorem 3.4 (abscissa subdifferential). First, we observe the constraint qualification

$$N(A_{\bar\delta}^{*}) \cap \partial^{\infty}\gamma(0) = \{0\}.$$

Hence, the nonsmooth chain rule holds

$$\partial(\gamma \circ A_{\bar\delta})(0) = A_{\bar\delta}^{*}\partial\gamma(0)$$

yielding

$$\partial(\gamma \circ A_{\bar\delta})(0) = \left\{ A_{\bar\delta}^{*}\left(-\frac{1}{5}e_4 + \sum_{j=0}^{3} c_j e_j\right) : \operatorname{Re} c_3 \le 0 \right\}. \tag{3.15}$$

A straightforward check shows the affine map from $\mathbf{C}^4$ to $\mathbf{C}^4$ defined by

$$(c_3, c_2, c_1, c_0) \mapsto A_{\bar\delta}^{*}\left(-\frac{1}{5}e_4 + \sum_{j=0}^{3} c_j e_j\right)$$

is invertible, and the inverse image of zero has $\operatorname{Re} c_3 < 0$. Consequently, by (3.15), we have the condition for a sharp minimizer

$$0 \in \operatorname{int}\partial(\gamma \circ A_{\bar\delta})(0).$$

A continuity argument now completes the proof. $\square$

It follows from this theorem that any small perturbation to the locally optimal polynomials x or y splits the quintuple root of $ax+by$ and moves at least one root strictly to the right, and into the open right half-plane when $\delta = \bar\delta$.

We now turn briefly to the case $n = 3$. The identity (3.2) reduces to

$$(s^2 - 2\delta s + 1)\left(s^4 + \sum_{k=0}^{3} w_k^{\delta} s^k\right) + (s^2 - 1)v_\delta \equiv (s - z_\delta)^6. \tag{3.16}$$

Multiplying out factors and equating terms leads to the formula

$$\tilde\delta = \frac{\sqrt{10 + 2\sqrt{5}}}{4} \approx 0.951\ldots$$

for which $z_{\tilde\delta} = 0$. This value is slightly larger than we observed numerically; given the sensitivity of the roots, it is not surprising that the optimization method was unable to find a stabilizing solution for $\delta = 0.95$. We verified that, as observed in our numerical experiments, a remarkable coincidence occurs: the real part of the rightmost conjugate pair of roots of $x^{\delta}(s) = s^4 + \sum_{k=0}^{3} w_k^{\delta} s^k$ is less than zero for $\delta < \tilde\delta$ and equal to zero for $\delta = \tilde\delta$. Consequently, for $\delta > \tilde\delta$, the structure of the optimal solution changes. The minimizer of the abscissa of $x(ax+by)$ is no longer a minimizer of the abscissa of $ax+by$, as a conjugate pair of roots of x is active. In principle, one could apply a parallel analysis to the new optimal structure for $n = 4$, but this has diminishing returns, especially as it seems likely that the optimal structure would change further as δ and $\deg(x)$ (and perhaps also $\deg(y)$) are increased further.

IV. Concluding Remarks

This paper has two messages. First, the gradient sampling method provides a very effective way to find local minimizers of challenging nonsmooth, nonconvex optimization problems of the kind that frequently arise in control. Second, stability objectives and constraints can be analyzed theoretically using recent results on nonsmooth analysis of the cone of stable polynomials. These approaches extend to encompass other key quantities of great practical interest, such as optimization of $\mathbf{H}_\infty$ performance. We are addressing these issues in ongoing work. In particular, we are developing a MATLAB toolbox called HIFOO ($\mathbf{H}_\infty$ Fixed-Order Optimization) that will allow engineers convenient free access to our techniques via a friendly user interface [7].

References

[1] K. J. Åström, "Limitations on control system performance," *Eur. J. Control*, vol. 6, no. 1, pp. 2–20, 2000.
[2] V. Blondel, *Simultaneous Stabilization of Linear Systems*, ser. Lecture Notes in Control and Information Sciences 191. Berlin, Germany: Springer-Verlag, 1994.
[3] V. Blondel, "Open problems in mathematical systems and control theory," in *Simultaneous Stabilization of Linear Systems and Interpolation with Rational Functions*, V. D. Blondel, E. D. Sontag, M. Vidyasagar, and J. C. Willems, Eds.. London, U.K.: Springer-Verlag, 1999, pp. 53–56.
[4] V. Blondel, 2005, Private communication.
[5] V. Blondel, M. Gevers, R. Mortini, and R. Rupp, "Simultaneous stabilization of three or more systems: Conditions on the real axis do not suffice," *SIAM J. Control Optim.*, vol. 32, pp. 572–590, 1994.
[6] V. Blondel, R. Rupp, and H. Shapiro, "On zero and one points of analytic functions," *Complex Variables: Theory Appl.*, vol. 28, pp. 189–192, 1995.
[7] J. V. Burke, D. Henrion, A. S. Lewis, and M. L. Overton, "HIFOO—A MATLAB package for fixed-order controller design and H_∞ optimization," in *Proc. 5th IFAC Symp. Robust Control Design*, Toulouse, France, 2006.
[8] J. V. Burke, A. S. Lewis, and M. L. Overton, "Optimizing matrix stability," in *Proc. Amer. Math. Soc.*, 2001, vol. 129, pp. 1635–1642.
[9] 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," in *Proc. 4th IFAC Symp. Robust Control Design*, Milan, Italy, 2003, pp. 175–181.
[10] J. V. Burke, A. S. Lewis, and M. L. Overton, "A robust gradient sampling algorithm for nonsmooth, nonconvex optimization," *SIAM J. Optim.*, vol. 15, pp. 751–779, 2005.
[11] J. V. Burke and M. L. Overton, "Variational analysis of the abscissa map for polynomials," *SIAM J. Control Optim.*, vol. 39, pp. 1651–1676, 2001.
[12] K. C. Kiwiel, "Convergence of the gradient sampling algorithm for nonsmooth nonconvex optimization," *SIAM J. Optim.*, 2006, to be published.

[13] V. V. Patel, G. Deodhare, and T. Viswanath, "Some applications of randomized algorithms for control systems design," *Automatica*, vol. 38, pp. 2085–2092, 2002.
[14] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*. New York: Springer-Verlag, 1998.

James V. Burke was born in New Jersey in 1955. He received the B.S. degree from Knox College, Galesburg, IL, and the Ph.D. degree in mathematics from the University of Illinois at Urbana-Champaign in 1983.

He was an Assistant Professor of Mathematics at the University of Kentucky, Lexington, from 1983 to 1985. He then moved to the University of Washington, Seattle, in 1985, where he is now a Professor of Mathematics holding adjunct positions in applied mathematics, statistics, and bioengineering. He is past Director of the Applied and Computational Mathematical Sciences Program at the University of Washington, and is currently the team leader for mathematical and statistical research at the Resource Facility for Population Kinetics in the Department of Bioengineering. His research interests include variational analysis and theoretical, numerical, and applied optimization, with a current focus on eigenvalue optimization and numerical methods for large-scale optimization.

Didier Henrion was born in Creutzwald, France, in 1971. He received the M.S. and Eng. degrees from the National Institute for Applied Sciences (INSA), Toulouse, France, in 1994, the Ph.D. degree from the Academy of Sciences of the Czech Republic, Prague, in 1998, and the Ph.D. degree from INSA in 1999, all in the area of control.

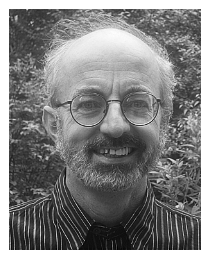
Since 2000, he has been a Researcher of CNRS (French National Center for Scientific Research) at LAAS (Laboratory of Analysis and Architecture of Systems), Toulouse, France. He is also holding a secondary appointment as a Researcher at the Faculty of Electrical Engineering of the Czech Technical University, Prague. His research interests are in numerical algorithms for polynomial matrices, convex optimization over linear matrix inequalities, robust multivariable control and computer-aided control system design.

Dr. Henrion was awarded the Bronze Medal from CNRS in 2004. He is currently an Associate Editor of IFAC *Automatica*, *European Journal of Control*, and the IEEE TRANSACTIONS ON AUTOMATIC CONTROL.

Adrian S. Lewis was born in England in 1962. He received the B.A., M.A., and Ph.D. degrees from Cambridge University, Cambridge, U.K., in 1983 and 1987, respectively.

He is a Professor in the School of Operations Research and Industrial Engineering at Cornell University, Ithaca, NY. Following his Research Fellowships at Queens' College, Cambridge, and Dalhousie University, Canada, he was with the University of Waterloo (1989–2001) and Simon Fraser University (2001–2004). He coauthored *Convex Analysis and Nonlinear Optimization* (New York: Springer-Verlag, 2006) with J. M. Borwein. His research concerns variational analysis and nonsmooth optimization, with a particular interest in optimization problems involving eigenvalues.

Dr. Lewis is an Associate Editor of the *SIAM Journal on Optimization*, *Mathematics of Operations Research*, and the *SIAM/MPS Book Series on Optimization*, as well as a Co-Editor of *Mathematical Programming*. He received the 1995 Aisenstadt Prize from the Canadian Centre de Recherches Mathematiques, the 2003 Lagrange Prize for Continuous Optimization from SIAM and the Mathematical Programming Society, and an Outstanding Paper Award from SIAM in 2005.



Michael L. Overton was born in England in 1953. He received his undergraduate education at the University of British Columbia, Canada, and received the Ph.D. degree in computer science from Stanford University, Stanford, CA, in 1979.

He then joined the Faculty of the Courant Institute of Mathematical Sciences, New York University, New York, where he now holds the position of Professor of Computer Science and Mathematics. He currently serves as Editor-in-Chief of the book series published jointly by the Mathematical Programming Society and SIAM. He is the author of *Numerical Computing with IEEE Floating Point Arithmetic* (Philadelphia, PA: SIAM, 2001). He has a special interest in optimization problems involving eigenvalues and related functions of matrices.

Dr. Overton served on the Council and Board of the Society for Industrial and Applied Mathematics (SIAM) for 15 years, and was Chair of the Board during 2004–2005. He was Editor-in-Chief of the *SIAM Journal on Optimization* during 1995–1999, and is currently on the editorial boards of the *SIAM Journal on Optimization*, the *SIAM Journal on Matrix Analysis and Applications*, *SIAM Review*, the *IMA Journal of Numerical Analysis*, the *ESAIM Journal on Control, Optimization and Calculus of Variations*, and *Numerische Mathematik*.