Analysis of a Belgian Chocolate Stabilization Problem

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

Abstract

We give a detailed numerical and theoretical analysis of a stabilization problem posed by V. Blondel in 1994. Our approach illustrates the effectiveness of a new "gradient sampling" algorithm for finding local optimizers of nonsmooth, nonconvex optimization problems arising in control, as well as the power of nonsmooth analysis for understanding variational problems involving polynomial roots and eigenvalues.

I. Introduction

More than a decade ago, Blondel [Blo94, p.150] offered a prize of a kilogram of Belgian chocolate for the solution of the following stabilization problem. By a stable polynomial we mean a real polynomial with all its roots in the open left half-plane.

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 $ax + by$

J.V. Burke, Department of Mathematics, University of Washington, Seattle, WA 98195, USA. Email: burke@math.washington.edu. Research supported in part by National Science Foundation Grant DMS-0203175.

D. Henrion, LAAS-CNRS, 7 Avenue du Colonel Roche, 31077 Toulouse, France and Department of Control Engineering, Czech Technical University in Prague, Technická 2, 16627 Prague, Czech Republic. Email: henrion@laas.fr. Research supported in part by Project ME 698/2003 of the Ministry of Education, Czech Republic.

A.S. Lewis, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, USA. Email: aslewis@orie.cornell.edu. Research supported in part by National Science and Engineering Research Council of Canada at Simon Fraser University, Burnaby, B.C., Canada.

M.L. Overton, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA. Email: overton@cs.nyu.edu. Research supported in part by National Science Foundation Grant DMS-0412049 and in part by Université Paul Sabatier, Toulouse, France.

is stable. (In the language of control, find the range of real values for δ for which there exists a stable, minimum phase rational controller y/x that stabilizes the plant b/a.)

He also offered a kilo of Belgian chocolate for a solution of a special case:[1]

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 [PDV02] found a solution with $\deg(x) = \deg(y) = 11$, using a randomized search method.

Problem 1.1 remains unsolved. According to [PDV02], it follows from results in [BG93], [BGMR94] that there exists a number δ^* such that stabilization is possible for all $\delta < \delta^*$ and is not possible for $\delta > \delta^*$, and furthermore, that $\delta^* < 0.9999800002$. Patel et al. [PDV02] also demonstrated 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

[1]There is a well known error in the original statement of Problem 1.2 that is evident from the implication that it is a special case of Problem 1.1.

which any small perturbation to x or y moves at least one root 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 model 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. 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 in state space.

It is broadly agreed that stabilization by low-order controllers is both difficult and of great practical importance. A principal difficulty is that the cones of stable polynomials and stable matrices are not convex. As a result, 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, as we explain in Section II. Then in Section III we present our local optimality analysis, using key theoretical properties of the cone of stable polynomials.

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,

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

(interpreted as $-\infty$ if p is a nonzero constant). Problem 1.1 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. We thus consider the problem: for fixed real δ and integers n and m with $m \leq n+1$, minimize $\alpha(xy(ax+by))$ over $x \in \mathbf{MP}_R^{n+1}$ and $y \in \mathbf{P}_R^m$. This is a nonconvex optimization problem in $n+m+2$ real variables.

For a given x and y, we say that a root (zero) of x, y 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. Consequently, standard local optimization methods such as steepest descent are inapplicable (when tried, they typically "jam" at a non-optimal point where the objective function is not differentiable). 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. Three of the authors have developed a method based on *gradient sampling* that is very effective in practice and for which a local convergence theory has been established [BLO05]. 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 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 $\langle A, B\rangle = \mathrm{Re}\,\mathrm{tr} A^*B$) 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 eigenvalue whose real part equals the spectral abscissa, and it 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 small values of n, 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

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

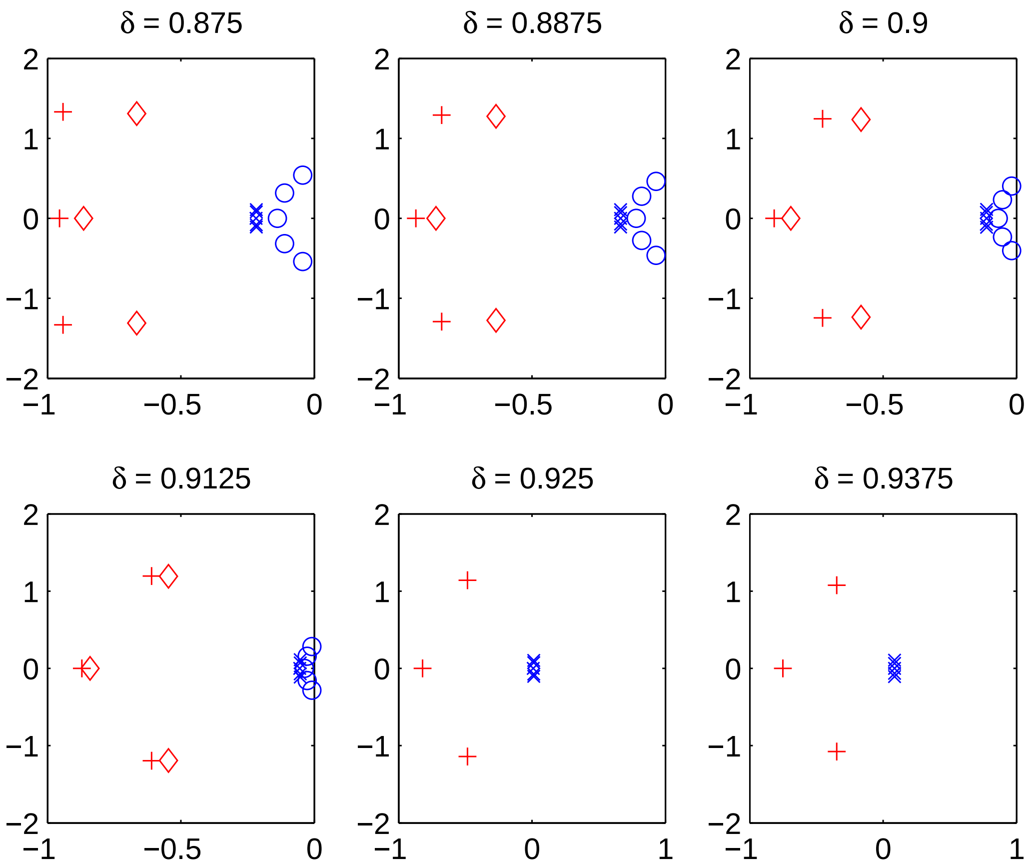

Fig. 1. Optimal roots for $\deg(x) = 3$, plotted in the complex plane for various values of δ. Crosses (blue) and plus signs (red) are respectively roots of $ax + by$ and x when the abscissa function $\alpha(x(ax + by))$ is minimized over monic cubic x and scalar y. Circles (blue) and diamonds (red)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

reduced to a negative value when $n \geq 2$ ($\deg(x) \geq 3$), 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 Figure 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 other, indicating the likelihood of coalescence to a single root for the exact local optimizer, while the three roots of x are well

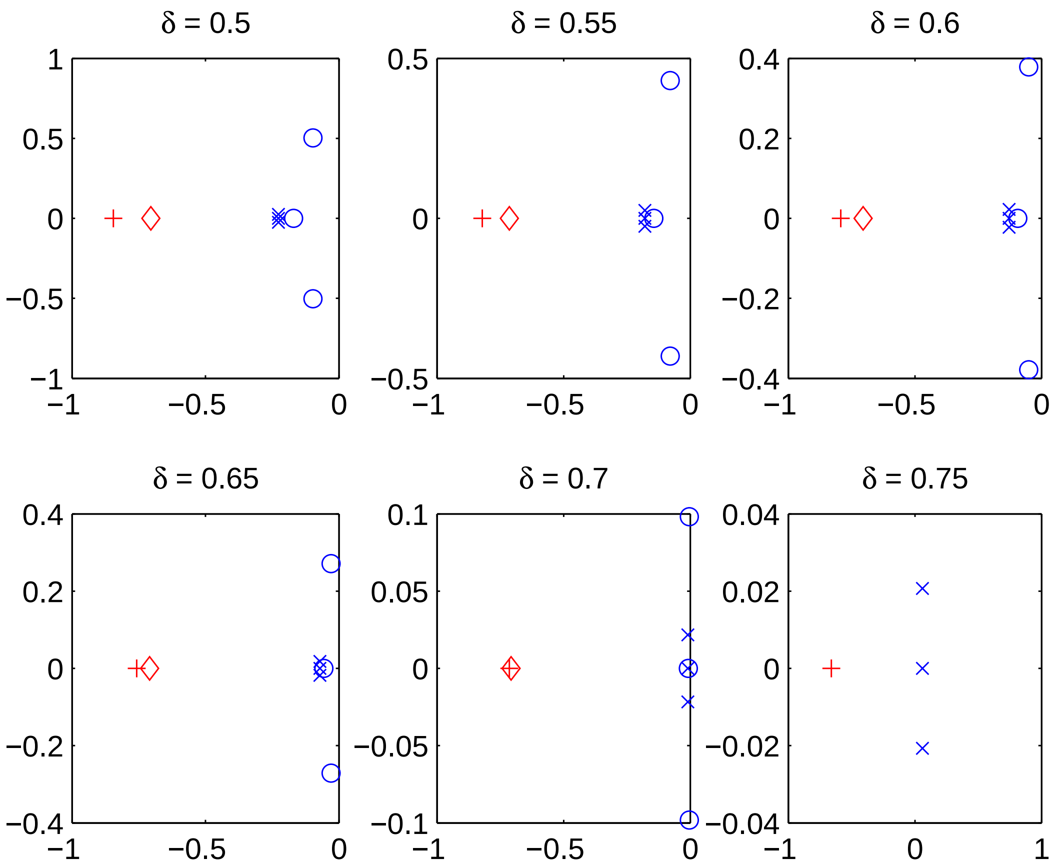

Fig. 2. Optimal roots for $\deg(x) = 1$ (see Figure 1 for details)

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 larger values of δ, this root is to the right of the imaginary axis, so the minimized value of $\alpha(x(ax + by))$ is positive, indicating that stabilization is *not* achieved in these cases.

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

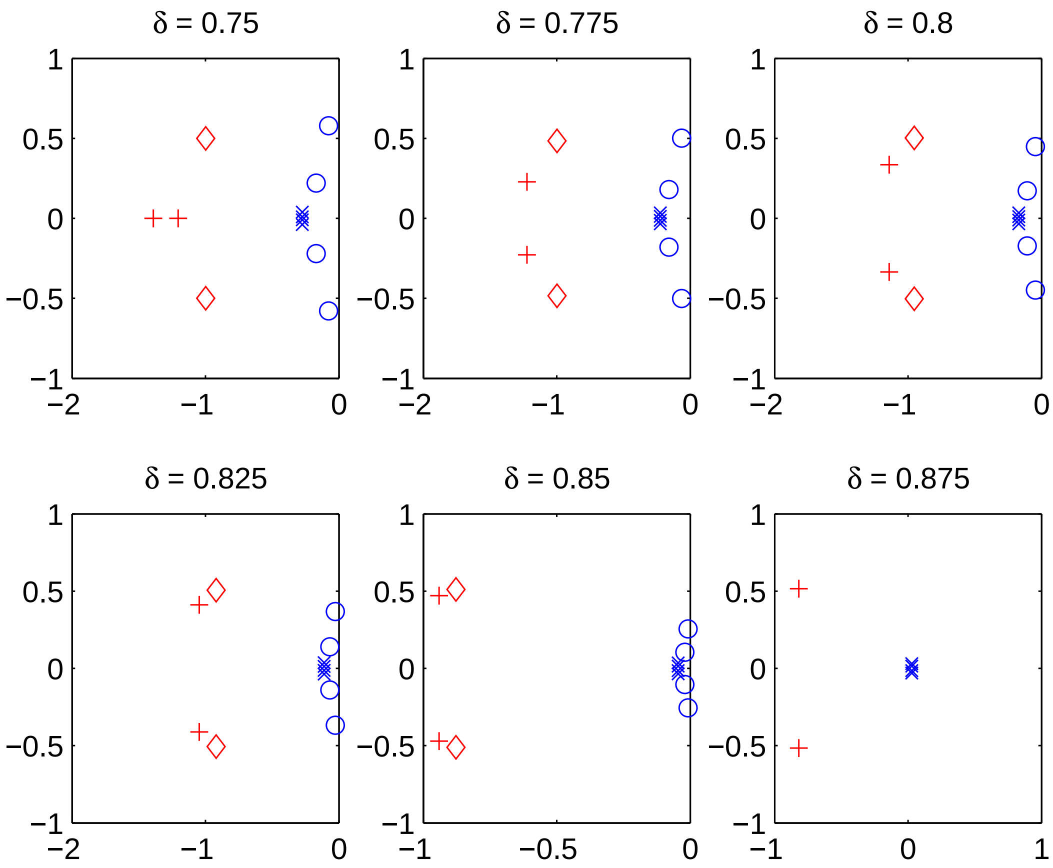

Fig. 3. Optimal roots for $\deg(x) = 2$ (see Figure 1 for details)

in $\mathbf{MP}^k$,

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

Here the norm is just the 2-norm of the coefficient vector. The quantity $\beta(p)$ can be computed by standard software. 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.

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

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

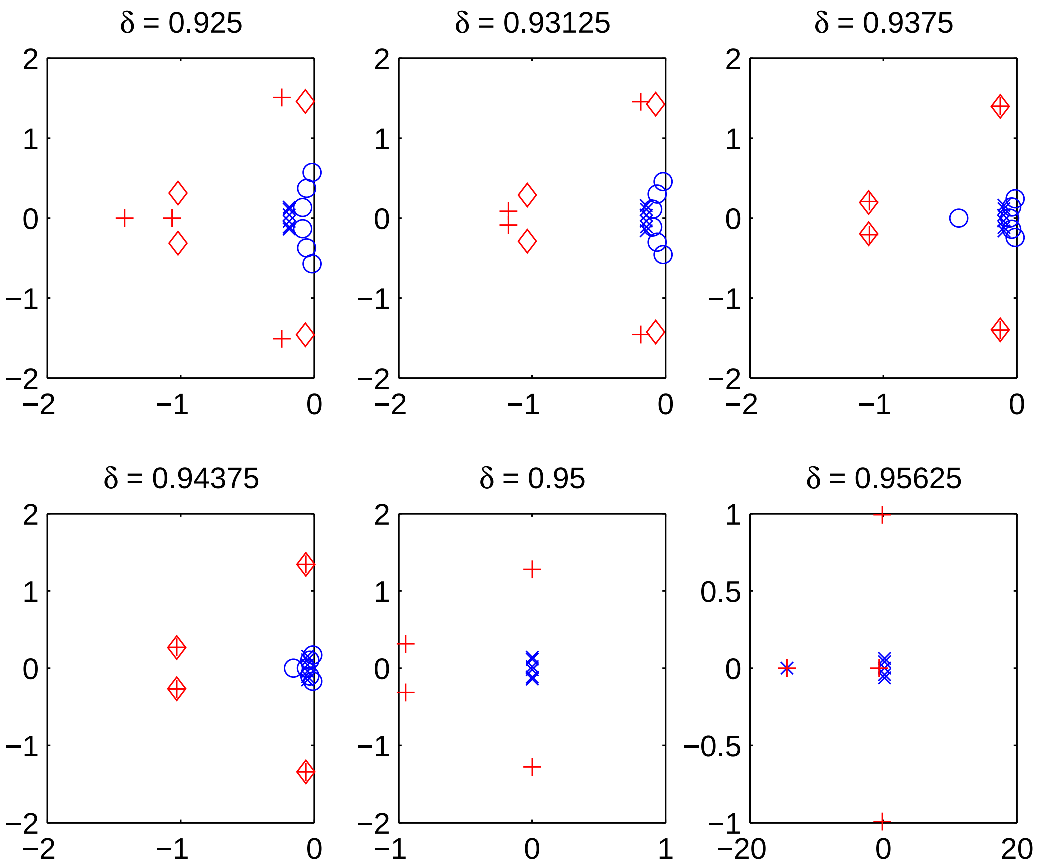

Fig. 4. Optimal roots for $\deg(x) = 4$ (see Figure 1 for details)

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 $ax + by$ does *not* have multiple roots, as expected; 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 Figure 1. For the two largest values of δ, stabilization was not achieved, so optimization of the stability radius could not be initialized.

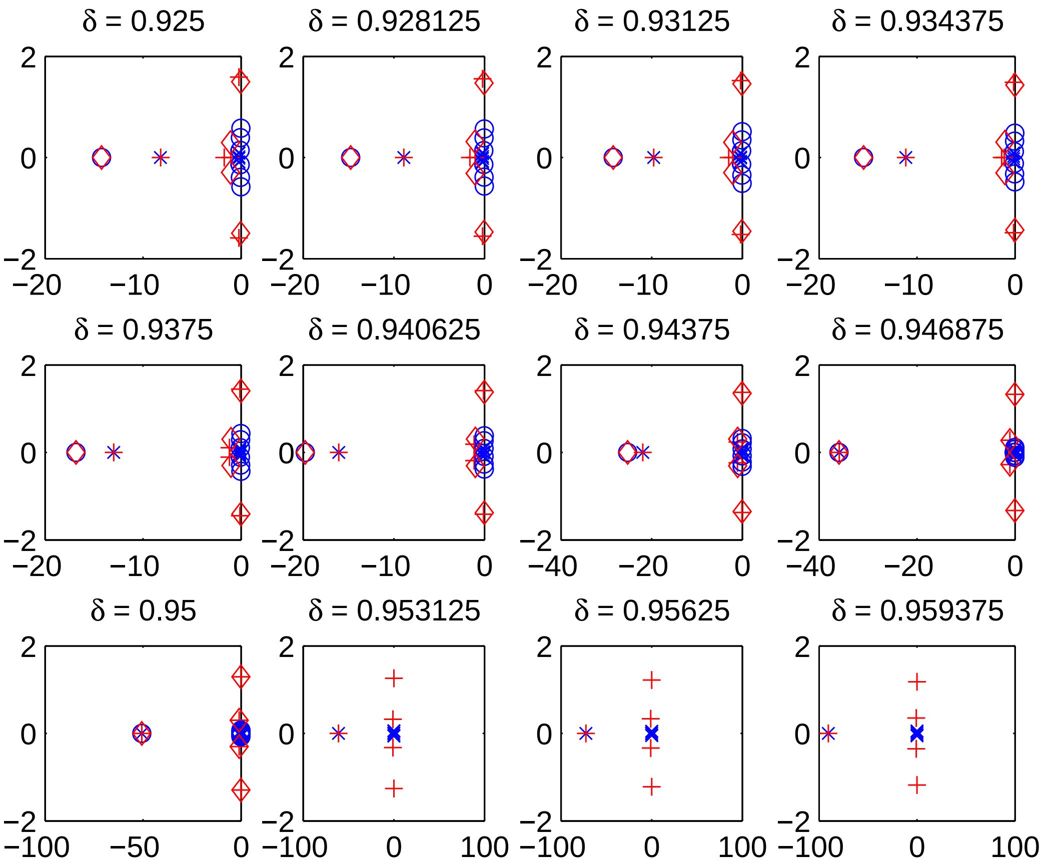

Fig. 5. Optimal roots for $\deg(x) = 5$ (see Figure 1 for details)

For $\delta = 0.9$, the approximately optimal abscissa was attained by $y = 1.8867980$ and

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

and the approximately optimal stability radius was attained by $y = 1.5774578$ and

$$x(s) = s^3 + 2.0105665\, s^2 + 2.8509924\, s + 1.5780819.$$

We next investigated for what values of δ stabilization is possible for $n = 0$ ($\deg(x) = 1$) and $n = 1$ ($\deg(x) = 2$). As earlier, we found that the optimal y is a scalar. Figures 2 and 3 show the roots of $ax + by$ (crosses) and x (plus signs) obtained by minimizing the abscissa function $\alpha(x(ax + by))$ for $n = 0$ and $n = 1$ respectively, for various values of δ. As previously, the optimal $ax + by$ evidently has only one root (triple when $n = 0$ and quadruple when $n = 1$), and

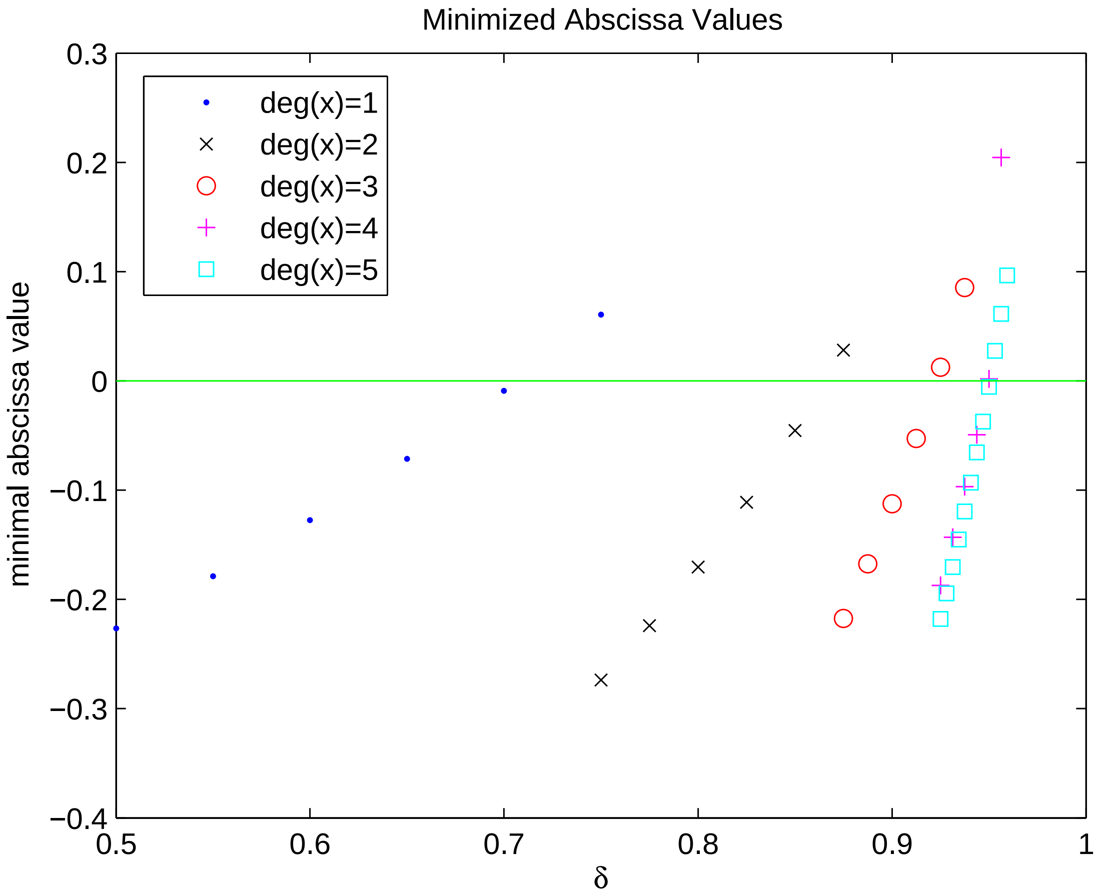

Fig. 6. Summary of Minimized Abscissa Values

this is the only active root, as the roots of x are well to its left. In the cases where stabilization was achieved, the roots of $ax + by$ and x obtained by maximizing the stability radius function $\tilde{\beta}(x, y)$ are shown as circles and diamonds respectively.

Having already covered the case $n = 2$, we now turn to $n = 3$. We found the same pattern, shown in Figure 4: the optimal y is a scalar, and for all δ for which stabilization is achieved, the optimal $ax + by$ 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

$$\begin{aligned} x(s) &= s^4 + 2.1853347\, s^3 + 3.1991472\, s^2 \\ &\quad + 3.8629224\, s + 2.0465529 \end{aligned}$$

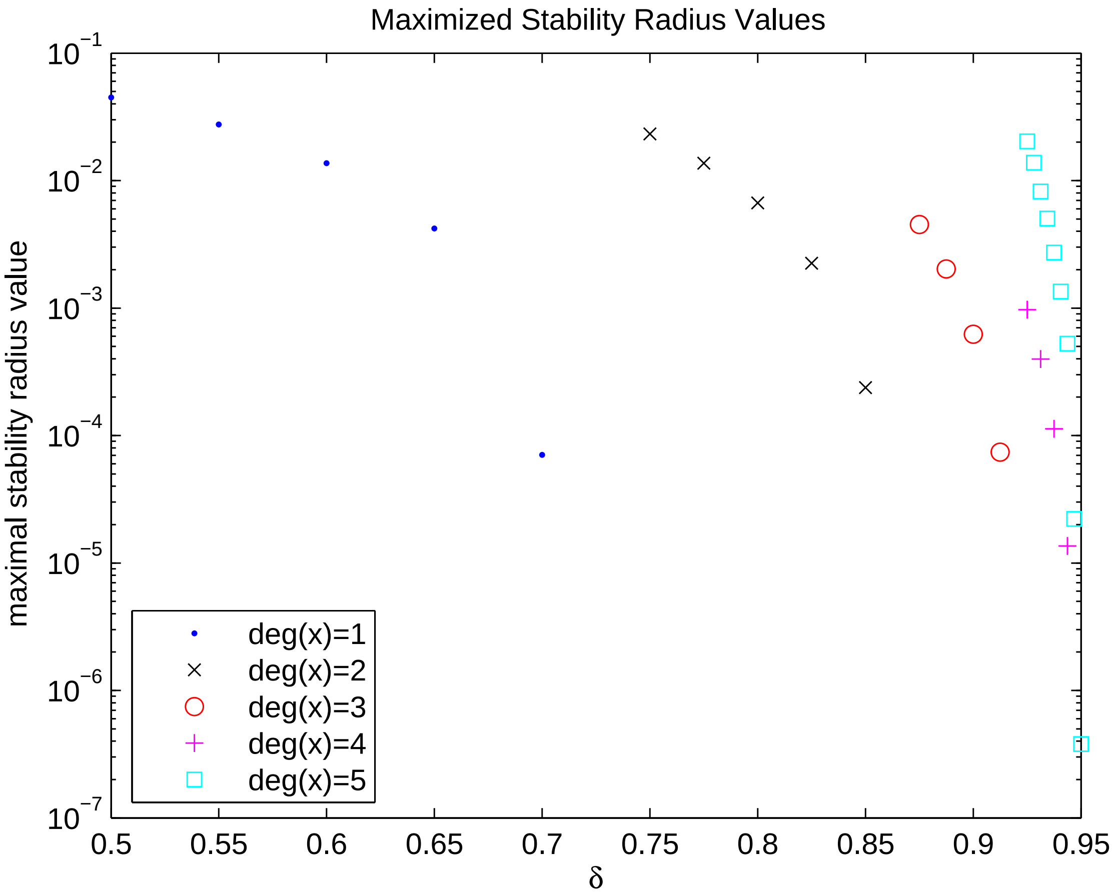

Fig. 7. Summary of Maximized Stability Radius Values

and the approximately optimal stability radius was attained by $y = 2.0461507$ and

$$x(s) = s^4 + 2.1875331\, s^3 + 3.1987968\, s^2 + 3.8622071\, s + 2.0461537.$$

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. In fact, as discussed at the end of Section III-B, these events occur at exactly the same critical value $\tilde{\delta}$. Beyond this value the 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.

In order to achieve stabilization beyond $\delta = \tilde{\delta}$, we must increase n to 4 ($\deg(x) = 5$). Now the structure of the local optimizer of $\alpha(x(ax + by))$, with $\deg(y)$ still set to 0, has changed, as shown in Figure 5. 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 up to $\delta = 0.95$. 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.

Figures 6 and 7 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 $\deg(x) \leq 4$, and how the curve for $\deg(x) = 5$ is very close to the one for $\deg(x) = 4$. This is a consequence of the change in structure of the optimal solution when the degree of x is raised from 4 to 5.

III. Theoretical Analysis

We now present a theoretical analysis inspired by the experimental results reported in the previous section. We observed that, for $0 \leq n \leq 3$ and for certain ranges of δ that depend on n, local minimizers (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

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

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 below; see [RW98, Chap. 8] and [BLO01].

Essential to our local optimality analysis is the following result of Burke and Overton [BO01]. 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

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

We can identify $\mathbf{P}^n$ with the Euclidean space $\mathbf{C}^{n+1}$, with the inner product $\langle u, v\rangle = \text{Re}\ \sum_{j=0}^{n} u_j^* v_j$. 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

$$\partial\gamma(0) = \Big\{ \sum_j c_j e_j : c_n = -\frac{1}{n+1},\ \text{Re}\, c_{n-1} \le 0 \Big\}$$

$$\partial^\infty\gamma(0) = \Big\{ \sum_j c_j e_j : c_n = 0,\ \text{Re}\, c_{n-1} \le 0 \Big\}.$$

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 equation (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

$$\begin{aligned} A_\delta(q,r)(t) &= q(t^2 + 2(z_\delta - \delta)t + (z_\delta^2 - 2\delta z_\delta + 1)) \\ &\quad + r(t^2 + 2z_\delta t + (z_\delta^2 - 1)). \end{aligned}$$

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 equation (3.3).

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

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

and in particular,

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

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

$$A_\delta^*\Big(\sum_j c_j e_j\Big) = 0 \ \text{ and } \ c_2 = 0 \quad \Rightarrow \quad c = 0.$$

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

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

Consequently we can apply the nonsmooth chain rule [BLO01, Lemma 4.4] to deduce that the composite function $\gamma \circ A_\delta$ is subdifferentially regular at zero, with subdifferential

$$\begin{aligned} &\partial(\gamma \circ A_\delta)(0) = A_\delta^* \partial\gamma(0) = \\ &\quad \Big\{ \begin{bmatrix} -\frac{1}{3} + 2(z_\delta - \delta)c_1 + (z_\delta^2 - 2\delta z_\delta + 1)c_0 \\ -\frac{1}{3} + 2z_\delta c_1 + (z_\delta^2 - 1)c_0 \end{bmatrix} : \\ &\qquad\qquad \mathrm{Re}\, c_1 \le 0,\ c_0 \in \mathbf{C} \Big\}. \end{aligned}$$

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 [BLO01, 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

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

The identity (3.2) becomes

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

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 $\bar{\delta} = \frac{1}{2}\sqrt{2 + \sqrt{2}}$ the equation

$$\begin{aligned} &\delta z^5 - 5z^4 + 10\delta z^3 + 10(1 - 2\delta^2)z^2 \\ &+ 5\delta(4\delta^2 - 3)z + (-1 + 8\delta^2 - 8\delta^4) \;=\; 0 \end{aligned}$$

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

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

We now present the main result of the 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 $\bar\delta = \frac{1}{2}\sqrt{2+\sqrt{2}}$ 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 $\delta < \hat\delta$.

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).

We therefore 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

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

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

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

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

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

We have

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

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

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

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

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

is invertible, and the inverse image of zero has $\operatorname{Re} c_3 < 0$. Consequently, by equation (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. □

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

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

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. The first is that 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. The second is that stability objectives and constraints can be analyzed theoretically using recent results on nonsmooth analysis of the cone of stable polynomials. These approaches can be extended to encompass other key quantities of great practical interest, such as optimization of $\mathbf{H}_\infty$ performance. We intend to address these issues in future work. Our gradient sampling code is freely available now, and we hope in the future to provide an interface for it that is more convenient for control engineers.

References

[BG93] V. Blondel and M. Gevers. Simultaneous stabilization of three linear systems is rationally undecidable. *Mathematics of Control, Signals Systems*, 6:135–145, 1993.

[BGMR94] 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 Journal on Control and Optimization*, 32:572–590, 1994.

[Blo94] V. Blondel. *Simultaneous stabilization of linear systems*. Springer, Berlin, 1994. Lecture Notes in Control and Information Sciences 191.

[BLO01] J.V. Burke, A.S. Lewis, and M.L. Overton. Optimizing matrix stability. *Proceedings of the American Mathematical Society*, 129:1635–1642, 2001.

[BLO05] J.V. Burke, A.S. Lewis, and M.L. Overton. A robust gradient sampling algorithm for nonsmooth, nonconvex optimization. *SIAM Journal on Optimization*, 2005. To appear.

[BO01] J.V. Burke and M.L. Overton. Variational analysis of the abscissa mapping for polynomials. *SIAM Journal on Control and Optimization*, 39:1651–1676, 2001.

[BRS95] V. Blondel, R. Rupp, and H. Shapiro. On zero and one points of analytic functions. *Complex Variables: Theory and Applications*, 28:182–192, 1995.

[PDV02] V.V. Patel, G. Deodhare, and T. Viswanath. Some applications of randomized algorithms for control system design. *Automatica*, 38:2085–2092, 2002.

[RW98] R.T. Rockafellar and R.J.B. Wets. *Variational Analysis*. Springer, New York, 1998.