

Nonlinear Physical Systems

Spectral Analysis, Stability and Bifurcations

Edited by
Oleg N. Kirillov
Dmitry E. Pelinovsky

Series Editor
Noël Challamel

ISTE

WILEY

First published 2014 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

© ISTE Ltd 2014

The rights of Oleg N. Kirillov and Dimtry E. Pelinovsky to be identified as the author of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Control Number: 2013950133

British Library Cataloguing-in-Publication Data

A CIP record for this book is available from the British Library

ISBN: 978-1-84821-420-0



Printed and bound in Great Britain by CPI Group (UK) Ltd., Croydon, Surrey CR0 4YY

Chapter 16

Stability Optimization for Polynomials and Matrices

Suppose that the coefficients of a monic polynomial or the entries of a square matrix depend affinely on parameters, and consider the problem of minimizing the root radius (maximum of the moduli of the roots) or root abscissa (maximum of their real parts) in the polynomial case, and the spectral radius or spectral abscissa in the matrix case. These functions are not convex and they are typically not locally Lipschitz near minimizers. We first address polynomials, for which some remarkable analytical results are available in one special case, and then consider the more general case of matrices, focusing on the static output feedback (SOF) problem arising in control of linear dynamical systems. We also briefly discuss some spectral radius optimization problems arising in the analysis of the transient behavior of a Markov chain and the design of smooth surfaces using subdivision algorithms.

16.1. Optimization of roots of polynomials

Optimization of roots of polynomials can arise in many contexts, but perhaps the most important application area is feedback control in the frequency domain [DOR 99]. Consider the problem

$$\min_{p \in P} \max_{\lambda \in C} \{\operatorname{Re} \lambda \mid p(\lambda) = 0\}$$

where

$$P = \{(\lambda^4 + 2\lambda^2)(x_0 + x_1\lambda + \lambda^2) + y_0 + y_1\lambda + y_2\lambda^2 \mid x_0, x_1, y_0, y_1, y_2 \in \mathbf{R}\}.$$

This arises in maximizing, over all linear feedback controllers of order two, the asymptotic decay rate for a two-mass-spring dynamical system with one input (an actuator positioning the first mass) and one output (the measured position of the second mass) [HEN 06]. The polynomials in P are those that are admissible as the denominators of the relevant rational closed-loop system transfer function, and their roots must be in the left half of the complex plane for the system to be stable. Note that P is a set of monic polynomials with degree 6 whose coefficients depend affinely on five parameters. In [HEN 06], a formulation was given of a polynomial in P of the form $(\lambda - \lambda_0)^6$, with just one distinct negative real root λ_0 of multiplicity 6, and its local optimality (the property that no *nearby* polynomial in P has all its roots to the left of λ_0) was established. It was subsequently discovered that the solution constructed in [HEN 06] is globally optimal (*no* polynomial in P has all its roots to the left of λ_0). This fact is a special case of a remarkable general theory that we summarize below, extracting the main results from a recent paper by the author and others [BLO 12]. As noted there, it turns out that part of this general theory was first given in a PhD thesis by Raymond Chen [CHE 79b].

16.1.1. Root optimization over a polynomial family with a single affine constraint

Every monic polynomial of degree n may be represented by a point in \mathbf{C}^n , representing the coefficients of the monomials $\lambda^{n-1}, \dots, \lambda, 1$. Let \mathbf{r} denote the *root radius* of such a polynomial:

$$\mathbf{r}(p) = \max \{|\lambda| : p(\lambda) = 0, \lambda \in \mathbf{C}\}.$$

The polynomial p is said to be *Schur stable* if $\mathbf{r}(p) < 1$. Let \mathbf{a} denote the *root abscissa*:

$$\mathbf{a}(p) = \max \{\operatorname{Re}(\lambda) : p(\lambda) = 0, \lambda \in \mathbf{C}\}.$$

The polynomial p is said to be *Hurwitz stable* if $\mathbf{a}(p) < 0$.

As functions of the polynomial coefficients, the radius \mathbf{r} and abscissa \mathbf{a} are not convex. They are continuous, but not Lipschitz continuous near a polynomial p with a multiple root whose modulus or real part, respectively, equals $\mathbf{r}(p)$ or $\mathbf{a}(p)$. So, in general, global minimization of the radius or abscissa over an affine family of monic

polynomials, pushing the roots as far as possible toward the origin or left in the complex plane, seems hard. Indeed, variations on the question of whether a given polynomial family contains one that is stable (has roots inside the unit circle or in the left-half plane) have been studied for decades [BLO 95]. But if an affine family of monic polynomials of degree n has $n - 1$ free parameters, this question can be answered efficiently. Equivalently, there is a *single affine constraint* on the coefficients.

16.1.2. The root radius

We begin by optimizing the root radius over real coefficients. Informally, we want to push all the roots as close to zero as possible. By compactness, an optimal polynomial must exist; the following result states an explicit form for the solution.

THEOREM 16.1.– [BLO 12, theorem 1]– Let b_0, b_1, \dots, b_n be real scalars (with b_1, \dots, b_n not all zero) and consider the affine family

$$P = \{ \lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n : b_0 + \sum_{j=1}^n b_j a_j = 0, a_i \in \mathbf{R} \}.$$

The optimization problem

$$\mathbf{r}^* := \inf_{p \in P} \mathbf{r}(p)$$

has a globally optimal solution of the form

$$p^*(\lambda) = (\lambda - \gamma)^{n-k} (\lambda + \gamma)^k \in P$$

for some integer k with $0 \leq k \leq n$, where $\gamma = \mathbf{r}^*$.

This leads immediately to an algorithm for computing k and γ :

COROLLARY 16.1.– [BLO 12, Corollary 2]– Let γ be the globally optimal value whose existence is asserted in theorem 16.1 and consider the set

$$\Xi = \{ r \in \mathbf{R} : g_k(r) = 0 \text{ for some } k \in \{0, 1, \dots, n\} \},$$

where

$$g_k(\lambda) = b_0 v_0 + b_1 v_1 \lambda + \dots + b_{n-1} v_{n-1} \lambda^{n-1} + b_n v_n \lambda^n$$

and (v_0, \dots, v_n) is the convolution of the vectors

$$\left(\binom{n-k}{0}, \binom{n-k}{1}, \dots, \binom{n-k}{n-k} \right) \text{ and } \left(\binom{k}{0}, -\binom{k}{1}, \dots, (-1)^k \binom{k}{k} \right)$$

for $k = 0, \dots, n$. Then, $-\gamma$ and γ are elements of Ξ with smallest magnitude.

Now, we consider the same problem with complex coefficients.

THEOREM 16.2.— [BLO 12, theorem 6]— Let b_0, b_1, \dots, b_n be complex scalars (with b_1, \dots, b_n not all zero) and consider the affine family

$$P = \left\{ \lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n : b_0 + \sum_{j=1}^n b_j a_j = 0, a_i \in \mathbf{C} \right\}.$$

The optimization problem

$$\mathbf{r}^* := \inf_{p \in P} \mathbf{r}(p)$$

has an optimal solution of the form

$$p^*(\lambda) = (\lambda - \gamma)^n \in P$$

with $-\gamma$ given by a root of smallest magnitude of the polynomial

$$h(\lambda) = b_n \lambda^n + b_{n-1} \binom{n}{n-1} \lambda^{n-1} + \dots + b_1 \binom{n}{1} \lambda + b_0.$$

Compared to the case of real coefficients, this result is somewhat simpler to state: there is an optimal polynomial all of whose roots coincide at one point γ , compared to two real roots $\pm\gamma$ in the real case. However, somewhat surprisingly, the proof in the complex case is much more complicated than in the real case, requiring an elaborate induction on n (see [BLO 12, Appendix A] for details).

16.1.3. The root abscissa

Now, we consider root abscissa optimization in the real case: instead of trying to push all the roots as close to zero as possible, we wish to push them as far as possible into the left half-plane. Consequently, we no longer have a compact domain for the coefficients, and hence there is no guarantee that an optimal polynomial will exist. The next result states exactly when it exists and what form it has.

THEOREM 16.3.— [BLO 12, theorem 7]— Let b_0, b_1, \dots, b_n be real scalars (with b_1, \dots, b_n not all zero) and consider the affine family

$$P = \{\lambda^n + a_1\lambda^{n-1} + \dots + a_{n-1}\lambda + a_n : b_0 + \sum_{j=1}^n b_j a_j = 0, a_i \in \mathbf{R}\}.$$

Let $k = \max\{j : b_j \neq 0\}$ and define the polynomial of degree k

$$h(\lambda) = b_n\lambda^n + b_{n-1}\binom{n}{n-1}\lambda^{n-1} + \dots + b_1\binom{n}{1}\lambda + b_0.$$

The optimization problem

$$\mathbf{a}^* := \inf_{p \in P} \mathbf{a}(p).$$

has the infimal value

$$\mathbf{a}^* = \min \left\{ \beta \in \mathbf{R} : h^{(i)}(-\beta) = 0 \text{ for some } i \in \{0, \dots, k-1\} \right\},$$

where $h^{(i)}$ is the i th derivative of h . Furthermore, the optimal value \mathbf{a}^* is attained by a minimizing polynomial p^* if and only if $-\mathbf{a}^*$ is a root of h (as opposed to one of its derivatives), and in this case, we can take

$$p^*(\lambda) = (\lambda - \gamma)^n \in P$$

with $\gamma = \mathbf{a}^*$.

The formula for the optimal value was given in [CHE 79b], as was the characterization of the optimal polynomial when $-\mathbf{a}^*$ is a root of h . The fact that there is no optimal polynomial when $-\mathbf{a}^*$ is a root of one of the higher derivatives of h was established in [BLO 12]. In this case, we may consider instead the problem of finding a sequence of polynomials whose abscissa converges to the optimal value. The following result displays a specific structure for such a sequence:

THEOREM 16.4.– [BLO 12, theorem 13]– Assume that $-\mathbf{a}^*$ is not a root of h . Let ℓ be the smallest integer $i \in \{1, \dots, k-1\}$ for which $-\mathbf{a}^*$ is a root of $h^{(i)}$. Then, for all sufficiently small $\epsilon > 0$, there exists a real scalar M_ϵ for which

$$p_\epsilon(\lambda) := (\lambda - M_\epsilon)^m (\lambda - (\mathbf{a}^* + \epsilon))^{n-m} \in P,$$

where $m = \ell$ or $\ell + 1$, and $M_\epsilon \rightarrow -\infty$ as $\epsilon \rightarrow 0$.

Thus, as in the real radius case, two roots play a role, but only one is finite. There is a formula for M_ϵ , but it is too complicated to include here and depends on whether $m = \ell$ or $m = \ell + 1$. In practice, it is a bad idea to make ϵ too small: then $|M_\epsilon|$ becomes large.

We observed that, in the real case, the optimal value is not attained when one of the *derivatives of h* has a real root to the right of the *rightmost real root* of h . However, it is not possible that a derivative of h has a complex root to the right of the *rightmost complex root* of h . This follows immediately from the Gauss–Lucas theorem [BUR 05b, MAR 66]. Consequently, we might guess that the optimal abscissa value is always attained in the complex case. Indeed, this is the case, as we now state.

THEOREM 16.5.– [BLO 12, theorem 14]– Let b_0, b_1, \dots, b_n be complex scalars (with b_1, \dots, b_n not all zero) and consider the affine family

$$P = \left\{ \lambda^n + a_1 \lambda^{n-1} + \dots + a_{n-1} \lambda + a_n : b_0 + \sum_{j=1}^n b_j a_j = 0, a_i \in \mathbf{C} \right\}.$$

The optimization problem

$$\mathbf{a}^* := \inf_{p \in P} \mathbf{a}(p)$$

has an optimal solution of the form

$$p^*(\lambda) = (\lambda - \gamma)^n \in P$$

with $-\gamma$ given by a root with largest real part of the polynomial

$$h(\lambda) = b_n \lambda^n + b_{n-1} \binom{n}{n-1} \lambda^{n-1} + \dots + b_1 \binom{n}{1} \lambda + b_0.$$

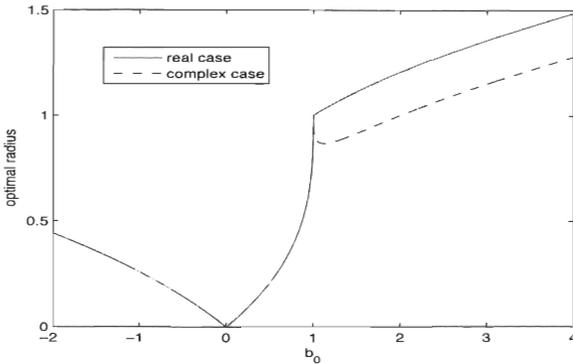


Figure 16.1. Optimal radius value for $\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3$ subject to the constraint $b_0 + a_1 + a_2 + a_3 = 0$, for $b_0 \in [-2, 4]$. The solid curve shows the optimal value when the coefficients a_i are required to be real while the dashed curve shows the optimal value when the a_i are allowed to be complex

16.1.4. Examples

EXAMPLE 16.1.– [BLO 12, BUR 01c]– Consider the problem of minimizing the root abscissa \mathbf{a} over the monic polynomials of degree n with the constraint that the coefficients of λ^{n-1} and λ^{n-2} must sum to zero. The polynomial h defined in theorems 16.3 and 16.5 is given by

$$h(\lambda) = \binom{n}{2}\lambda^2 + n\lambda = \binom{n}{2}\lambda \left(\lambda + \frac{2}{n-1} \right).$$

The roots of h are 0 and $\frac{-2}{n-1}$ and the only root of the derivative $h^{(1)}$ is $\frac{-1}{n-1}$, so $\mathbf{a}^* = 0$ and λ^n is a global optimizer. Theorem 16.5 proves global optimality over $a_i \in \mathbf{R}$ and theorem 16.5 proves global optimality over $a_i \in \mathbf{C}$. Thus, no polynomial with the constraint $a_1 + a_2 = 0$ can be Hurwitz stable.

EXAMPLE 16.2.– Consider the problem of minimizing the root radius \mathbf{r} over the monic polynomials of degree n with the constraint that $b_0 + a_1 + \dots + a_n = 0$, where $b_0 \in \mathbf{R}$. When $b_0 = 0$, all the coefficients can be taken to be zero, so the optimal polynomial is $p^* = \lambda^n$ with $\mathbf{r}^* = 0$. When $b_0 = 1$ and the coefficients are restricted to be real, we find that the set Ξ defined in corollary 16.1 is just $\{-1, 1\}$, so $\mathbf{r}^* = 1$. When $b_0 = 1$ and the coefficients a_i are allowed to be complex, the polynomial h defined in theorem 16.2 is $h(\lambda) = (\lambda + 1)^n$, so its only root is -1 and again the optimal value is $\mathbf{r}^* = 1$, with the optimal polynomial $p^* = (\lambda - 1)^n$. For $b_0 < 1$, the optimal value drops rapidly toward zero in both the real and complex cases, but for

$b_0 > 1$, the behavior is quite different: in the complex case, \mathbf{r}^* drops rapidly before turning around and increasing again, while in the real case, \mathbf{r}^* increases steadily as b_0 increases. Figure 16.1 shows the optimal value for $n = 3$ and $b_0 \in [-2, 4]$. The solid curve shows the optimal value in the real case and the dashed curve shows the optimal value in the complex case. Thus, for all b_0 near to but not equal to one, there is a Schur stable polynomial whose coefficients sum to one if they are allowed to be complex, but if they are required to be real, no Schur stable polynomial exists when $b_0 \geq 1$.

There is a significant class of frequency domain stabilization problems that can be solved using theorem 16.3, as explained in [RAN 89]; see also [BLO 12, section I]. These include the two-mass-spring stabilization example mentioned in section 16.1. Some other examples may be found in [BLO 12, section IV]. More may readily be explored using a publicly available MATLAB code implementing the constructive algorithms implicit in the theorems above.¹

We should emphasize that multiple roots are very sensitive to perturbation. More specifically, a random perturbation of size ϵ to the coefficients of a polynomial with a non-zero root with multiplicity k moves the corresponding roots by $O(\epsilon^{1/k})$. In practice, we might want to locally optimize a more robust measure of stability, as mentioned in section 16.3. Also, the monomial basis is a poor choice numerically unless the polynomial has very small degree. Nonetheless, the optimal *value* can be computed accurately even if n is quite large.

16.1.5. Polynomial root optimization with several affine constraints

A natural question is whether the results given above extend to cases where there are several affine constraints on the coefficients. Figure 16.2 shows contour plots of \mathbf{r} for two randomly generated monic polynomial families with coefficients depending on two real parameters. In Figure 16.2(a) [BLO 10], the polynomials have degree $n = 3$, so affine dependence on two parameters is equivalent to imposing a single affine constraint on the coefficients, and hence the global minimizer can be computed explicitly by theorem 16.1. The global minimizer $(\lambda - \gamma_1)^3$, where $\gamma_1 \approx 0.541$, occurs on the right side of the plot. There is also a local minimizer, $(\lambda - \gamma_2)^3$, with $\gamma_2 \approx -0.567$, on the left side. Both $-\gamma_1$ and $-\gamma_2$ are roots of the polynomial g_0 defined in corollary 16.1 and so both are elements of Ξ , and indeed, it is generally true that all local minimizers must have roots $\pm\gamma \in \Xi$, and so elements of Ξ often (though not always) correspond to local minimizers. In contrast, in Figure 16.2(b), the polynomials have degree $n = 4$, so affine dependence on two parameters is equivalent to two affine constraints, and the theorems of the previous section are not applicable. Again global and local minimizers appear on the right and left, respectively, but these were approximated by numerical optimization; more on

¹ www.cs.nyu.edu/overton/software/affpoly/.

this below. Both panels clearly illustrate both the non-convexity and non-Lipschitz behavior of r , particularly around the global and local minimizers.

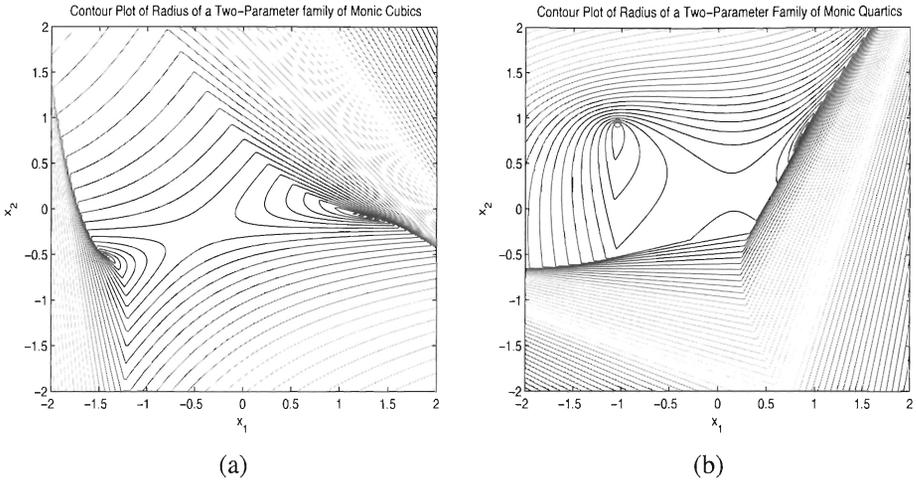


Figure 16.2. Contour plot of the root radius r for two randomly generated monic polynomial families with coefficients depending affinely on two real parameters x_1, x_2 . For each, the global minimizer occurs on the right side of the plot and a local minimizer occurs on the left. a) A cubic family for which the minimizers were obtained using theorem 16.1. b) A quartic family for which minimizers were approximated numerically

Let us define the *active roots* of a polynomial p as those whose modulus equals $r(p)$ or real part equals $a(p)$, depending on the context; the others are said to be *inactive*. We know from the results of section 16.1.2 that when there is one affine constraint, the root radius always has a global minimizer with no inactive roots and in fact with at most two distinct active roots (one in the complex case). In the case of the quartic family plotted in the right panel, for which there are implicitly two affine constraints on the coefficients, the global minimizer that was found numerically is $(\lambda - \zeta_1)^3(\lambda - \zeta_2)$ where $\zeta_1 \approx -0.983$ and $\zeta_2 \approx 0.905$, with one triple active root and one inactive root. The local minimizer that was found apparently has the form $(\lambda - \zeta_3)^2(\lambda - \bar{\zeta}_3)^2$, with $\zeta_3 \approx -0.053 + 1.07i$, so it has a double conjugate pair of active roots with no inactive roots. The global minimizer is Schur stable but the local minimizer is not.

Now consider the general problem of minimizing r subject to κ affine constraints on the coefficients. Based on extensive numerical experiments, we conjecture that there always exists a global minimizer with *at most* $\kappa - 1$ inactive roots. However, there does not seem to be a useful bound on the number of possible *distinct* active roots. Thus, computing global optimizers appears to be difficult.

16.1.6. Variational analysis of the root radius and abscissa

Given a polynomial p that is a candidate for a local minimizer of r or a , discovered through numerical optimization or in some other way, how may we verify that it is locally optimal, or at least a stationary point in a sense that is suitable for non-smooth functions? When the coefficients are parametrized by two real variables, we can be reasonably confident simply by looking at a contour plot, but this is not possible for many variables. Standard optimality conditions such as those used for constrained optimization, based on Lagrange multipliers [NOC 06], or for convex optimization, based on subgradients and other concepts from convex analysis [ROC 70], are not applicable to polynomial root optimization problems. However, there is a whole body of work known as non-smooth analysis or variational analysis [CLA 83, ROC 98, BOR 05] that *is* applicable. Following the development in [ROC 98], formulas for subgradients of the root abscissa a , valid even at polynomials with multiple active roots where a is not Lipschitz, were presented in [BUR 01b], where a key result was proved: a is globally *subdifferentially regular*. Subsequently, this work was refined in [BUR 05b] and extended to more general max functions of the roots, including the root radius r , in [BUR 12]. These results can, in principle, be used to establish necessary or sufficient conditions for a candidate polynomial to be a local minimizer of the radius r or abscissa a even in the presence of multiple active roots, exploiting a chain rule from non-smooth analysis to take account of the affine parametrization; an abscissa example is worked out in detail in [BUR 06b]. We write “in principle” above, because if the optimization problem or the candidate minimizer is not sufficiently simple, or the candidate minimizer is not known exactly, establishing these conditions is complicated or impossible. We will mention a far less rigorous but more practical alternative approach to assessing approximate local optimality in section 16.2.3.

16.1.7. Computing the root radius and abscissa

In MATLAB, roots of a given polynomial $p(\lambda) = \lambda^n + a_1\lambda^{n-1} + \dots + a_n$ are computed numerically by determining the eigenvalues of the corresponding companion matrix

$$M(p) = \begin{bmatrix} -a_1 & 1 & & & \\ -a_2 & & 1 & & \\ \cdot & & & \cdot & \\ \cdot & & & & \cdot & \\ \cdot & & & & & 1 \\ -a_n & 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix},$$

the rationale being that the method used to solve the eigenvalue problem is highly reliable and efficient. So, in our root optimization experiments, we computed the root

radius $r(p)$ or root abscissa $a(p)$ by computing the spectral radius or abscissa of $M(p)$. The gradients needed for numerical optimization were then obtained by differentiating the spectral radius or abscissa (more on this below), applying the ordinary chain taking account of the companion matrix structure to yield gradients in the polynomial coefficient space, and applying the chain rule again to recover gradients in the parameter space. Although the gradients are valid only at polynomials with just one active root (or conjugate pair of active roots in the real case) with multiplicity one, appropriate numerical methods using these gradients approximate local minimizers of r and a with multiple active roots surprisingly effectively, as will be discussed further in the next section.

16.2. Optimization of eigenvalues of matrices

Now let us turn our attention to eigenvalues of $n \times n$ matrices. Let $\rho : \mathbf{C}^{n \times n} \rightarrow \mathbf{R}$ define the *spectral radius*:

$$\rho(X) = \max \{ |\lambda| : \det(\lambda I - X) = 0, \lambda \in \mathbf{C} \}$$

and $\alpha : \mathbf{C}^{n \times n} \rightarrow \mathbf{R}$ the *spectral abscissa*:

$$\alpha(X) = \max \{ \operatorname{Re}(\lambda) : \det(\lambda I - X) = 0, \lambda \in \mathbf{C} \}.$$

An eigenvalue of X is said to be *active* if its modulus (real part) equals the spectral radius (abscissa). The spectral functions ρ and α are not convex and are not Lipschitz near a matrix with an active multiple eigenvalue. We say that an eigenvalue is *non-derogatory* if it is associated with only one right eigenvector. Equivalently, it is associated with a single block in the Jordan form, whose dimension is its *multiplicity*. Of those matrices with an eigenvalue with given multiplicity, those for which the eigenvalue is non-derogatory are most generic [ARN 71].

The spectral radius and abscissa of a matrix X are the root radius and abscissa of the characteristic polynomial $\det(\lambda I - X)$, but the results of section 1 do not extend to the more general case of an affine family of $n \times n$ matrices depending on $n - 1$ parameters. For example, consider the matrix family

$$A(\xi) = \begin{bmatrix} \xi & 1 \\ -1 & \xi \end{bmatrix}.$$

This matrix depends affinely on a single parameter $\xi \in \mathbf{R}$, but the coefficients of its characteristic polynomial $\det(\lambda I - A(\xi))$ do not, so the results of the previous section do not apply. The minimal spectral radius of $A(\xi)$ is attained by $\xi = 0$, for which the eigenvalues are $\pm i$.

16.2.1. Static output feedback

As control feedback problems in the frequency domain are a source of applications for polynomial root optimization problems, so control feedback problems in state space are a source of applications for eigenvalue optimization problems. To be specific, let us consider the SOF stabilization problem. Given the discrete-time dynamical system with control input and measured output

$$x_{k+1} = Fx_k + Gu_k, \quad y_k = Hx_k$$

where $F \in \mathbf{R}^{n \times n}$, $G \in \mathbf{R}^{n \times p}$, $H \in \mathbf{R}^{m \times n}$, the SOF problem is to find a controller $K \in \mathbf{R}^{p \times m}$ so that, setting $u = Ky$, all solutions of

$$x_{k+1} = (F + GKH)x_k$$

converge to zero, that is all eigenvalues of $F + GKH$ are inside the unit disk, or prove that this is not possible. The analogous continuous-time system is

$$\dot{x} = Fx + Gu, \quad y = Hx$$

and then the SOF problem is to find K so that the eigenvalues of $F + GKH$ are in the left half-plane. In general, SOF is a major open problem in control [BLO 95, KIM 94]. However, if $m = n$ and H is non-singular (without loss of generality, we can take $H = I$, the identity matrix, so that $y = x$), the SOF greatly simplifies and becomes the *state feedback* stabilization problem. Then, provided the pair (F, G) is controllable, a property that holds generically, a standard “pole placement” procedure [WON 85] defines K so that the eigenvalues of $F + GK$ have *any* desired values, so stabilization is trivial. In fact, stabilization via pole placement extends generically to the case $n < mp$ [WAN 96, WIL 97] (and to $n = mp$ in the complex case [HER 77]).

When $n > mp$, the SOF problem seems hard. A natural approach is to formulate the stabilization problem as an eigenvalue optimization problem: minimize either the spectral radius ρ or spectral abscissa α over the affine matrix parametrization $F + GKH$ for $K \in \mathbf{R}^{pm}$. But, as pointed out by Chen [CHE 79b, p. 4], [CHE 79a, p. 412], there is an interesting simplification in the case $p = 1$, i.e. when the dynamical system has a single control input, which depends on the following lemma due to Cauchy [HOR 13, equation [0.8.5.11]].

LEMMA 16.1.– Let $A \in \mathbf{C}^{n \times n}$, $b \in \mathbf{C}^{n \times 1}$ and $c \in \mathbf{C}^{1 \times n}$. Then,

$$\det(A + bc) = \det(A) + c \operatorname{adj}(A) b$$

where adj is variously known as the classical adjoint, the adjugate or the transposed cofactor matrix.

It follows that when $p = 1$, substituting $\lambda I - F$ for A , G for b and $-KH$ for c , one has²

$$\det(\lambda I - (F + GKH)) = \det(\lambda I - F) - KH \text{adj}(\lambda I - F)G. \quad [16.1]$$

The entries in $\text{adj}(\lambda I - F)$ are cofactors of $\lambda I - F$, which are polynomials in λ with degree $n - 1$. So, when $p = 1$, the coefficients of the characteristic polynomial of $F + GKH$ depend affinely on the m entries of the row vector $K \in \mathbf{R}^{1 \times m}$, and hence optimizing the spectral radius or abscissa reduces to an affinely parametrized polynomial root radius or root abscissa optimization problem. Then, if we further assume $m = n - 1$ (so that the dynamical system has $n - 1$ outputs), it follows that the SOF problem can be solved explicitly using the results given in sections 16.1.2 and 16.1.3. Indeed, this was a key motivation for Chen's PhD thesis [CHE 79b]. However, in general, optimizing the spectral abscissa or radius of $F + GKH$ over $K \in \mathbf{R}^{pm}$ does *not* reduce to polynomial root optimization, so we tackle the eigenvalue optimization problem directly.

16.2.2. Numerical methods for non-smooth optimization

In developing numerical methods suitable for searching for local minima of non-smooth functions, we take the following viewpoint: although the optimization objective function may not be differentiable or even Lipschitz at minimizers, we evaluate it at only a finite number of points and we can reasonably expect it to be differentiable at all of them. Indeed, any locally Lipschitz function is differentiable almost everywhere on its domain, and this is also true of the non-locally-Lipschitz but semi-algebraic functions ρ , α , \mathbf{r} and \mathbf{a} . However, the standard method of steepest descent (choosing each new iterate by moving along the direction of the negative gradient from the current iterate) fails badly on non-smooth functions, typically generating iterates that converge to a point where the function is not differentiable but that is nowhere near a local minimizer.

To be able to approximate local minimizers of non-smooth functions, a method must be able to exploit the gradient information obtained at several points, not just at one point. One such method, the gradient sampling algorithm [BUR 05a], has a very satisfactory convergence theory for locally Lipschitz functions, although this has not been extended to non-locally-Lipschitz functions. This was the method that we used in our first numerical experiments with eigenvalue optimization [BUR 02], but the

² There is a sign error in Chen's version of this equation.

computational cost is substantial. Another method, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton algorithm with an Armijo–Wolfe line search, is a standard workhorse for minimizing smooth functions even in the absence of convexity, but it has not generally been considered suitable for functions that are non-differentiable, let alone non-Lipschitz, at their local minimizers. However, although the theoretical properties of the BFGS method in this context are not well understood, in practice we have found that it is an efficient and reliable method for identifying local minimizers of non-smooth functions, particularly in the locally Lipschitz case [LEW 13]. Our MATLAB code HANSO combines the gradient sampling and BFGS methods together in one package.³

Both methods require repeated evaluation of the objective function and its gradient at a sequence of points. As already noted, we take the view that with high probability, the function will be differentiable at all these points, and hence the gradient will be well defined, but in any case, using floating-point computer arithmetic it makes little sense to try to determine whether or not a function is differentiable at a given point. In the case of the spectral radius ρ , the matrix at which it is being evaluated will normally have only one active eigenvalue (or one complex conjugate pair of active eigenvalues in the real case), with multiplicity one, implying that ρ is differentiable. If there happens to be a tie for the maximum value, we take the view that it is broken arbitrarily; when ρ is evaluated again at other nearby points, eventually the tie will be broken differently, providing more gradient information.

Thus, in minimizing the spectral radius ρ or the spectral abscissa α of $X(K) = F + GKH$ over the parameter matrix $K \in \mathbf{R}^{p \times m}$, all we need to do is to provide a routine to compute $\rho(\tilde{X})$ or $\alpha(\tilde{X})$ together with its gradient with respect to $\tilde{X} \in \mathbf{R}^{n \times n}$ at a given matrix $\tilde{X} = X(\tilde{K}) = F + G\tilde{K}H$, and then use the ordinary chain rule to obtain the gradient of $\rho(X(K))$ or $\alpha(X(K))$ with respect to the parameter matrix K at a given \tilde{K} . Assume that only one real eigenvalue or conjugate pair of eigenvalues of \tilde{X} is active, with multiplicity one. Since $X(K)$ is real for all real K , we can confine our attention to eigenvalues with non-negative imaginary part. Let the active eigenvalue and its right and left eigenvectors be denoted by μ , v and u respectively, that is $\tilde{X}v = \mu v$ and $u^* \tilde{X} = \mu u^*$, normalized so that $u^*v = 1$. Then, exploiting well-known formulas [HOR 85, theorem 6.3.12] we have

$$\nabla \alpha(\tilde{X}) = \operatorname{Re} uv^* \quad \text{and} \quad \nabla \rho(\tilde{X}) = \operatorname{Re} \frac{\mu}{|\mu|} uv^*.$$

Under the assumption that μ is simple, the normalization $u^*v = 1$ is always possible, but if μ is close to a multiple eigenvalue, the resulting gradient norm $\|u\| \|v\|$ may be arbitrarily large. This is to be expected, because α and ρ are not Lipschitz near a matrix with an active multiple eigenvalue.

³ www.cs.nyu.edu/overton/software/hanso.

16.2.3. Numerical results for some SOF problems

We define a class of SOF problems as follows. Let F be a Grcar matrix [WRI 02] scaled by $1/2$, defined as a Toeplitz matrix whose non-zeros consist of the number 0.5 on the main diagonal and first three superdiagonals, and the number -0.5 on the first subdiagonal. We start by setting its order $n = 8$ and $p = 1$, with $G = [1, \dots, 1]^T$, and consider m ranging from 0 to 8 , setting H to the matrix whose rows are the first m rows of the identity matrix. When $m = 0$, the matrices H and K are empty, so we interpret $X = F + GKH$ as $X = F$. The top half of Figure 16.3 shows eigenvalues in the complex plane as small circles. The first panel, for $m = 0$, shows the eigenvalues of F . The other panels show, for each value of m between 1 and 8 , the eigenvalues of a matrix \tilde{X} obtained from numerical minimization of the spectral radius of $X = F + GKH$ over $K \in \mathbf{R}^{1 \times m}$. The optimization was conducted by initializing the BFGS method at 100 different starting points whose entries were randomly generated from the standard normal distribution, taking up to 1,000 BFGS iterations for each run, with early termination when the line search failed to obtain a reduction in ρ . Then, \tilde{X} was defined as the final matrix with the smallest spectral radius over all 100 runs. The large circle on each panel has its center at the origin and radius $\rho(\tilde{X})$, so the small circles superimposed on the large circle show the active eigenvalues of \tilde{X} while those inside it correspond to inactive eigenvalues.

In the first panel, we see that F has just one conjugate pair of active eigenvalues, and in the second panel, that after optimizing ρ over $K \in \mathbf{R}$, there are two conjugate pairs of active eigenvalues whose modulus is slightly reduced below $\rho(F)$. In the next panel, we see that by optimizing over $K \in \mathbf{R}^{1 \times 2}$, we are able to obtain a Schur stable matrix, with a real active eigenvalue as well as two active conjugate pairs. For $m = 3$, we see active multiple eigenvalues for the first time. Of course, these are not *exactly* multiple eigenvalues, but we observe one active conjugate pair of double eigenvalues coinciding up to plotting accuracy as well as another active conjugate pair of simple eigenvalues. Note that since non-derogatory eigenvalues are the most generic, the multiple eigenvalues obtained from optimizing spectral functions are expected to be non-derogatory, unless some special structure, such as matrix symmetry, is present. That is not the case here.

For $m = 4$, we observe an active conjugate pair of triple eigenvalues. For $m = 5$, we see a double active conjugate pair and a triple active real eigenvalue, leaving just one inactive eigenvalue. For $m = 6$, we have one quadruple active conjugate pair, with no inactive eigenvalues. As explained in section 16.2.1, since $p = 1$, these SOF spectral radius optimization problems are actually equivalent to root radius optimization problems with $n - m$ affine constraints on the polynomial coefficients. Therefore, according to the conjecture in section 16.1.5, for each m , there should exist an optimal solution with most $n - m - 1$ inactive roots. As $n = 8$, Figure 16.3 supports this conjecture, displaying 4, 3, 2, 2, 1 and 0 inactive roots for $m = 1, 2, 3, 4, 5$ and 6 respectively.

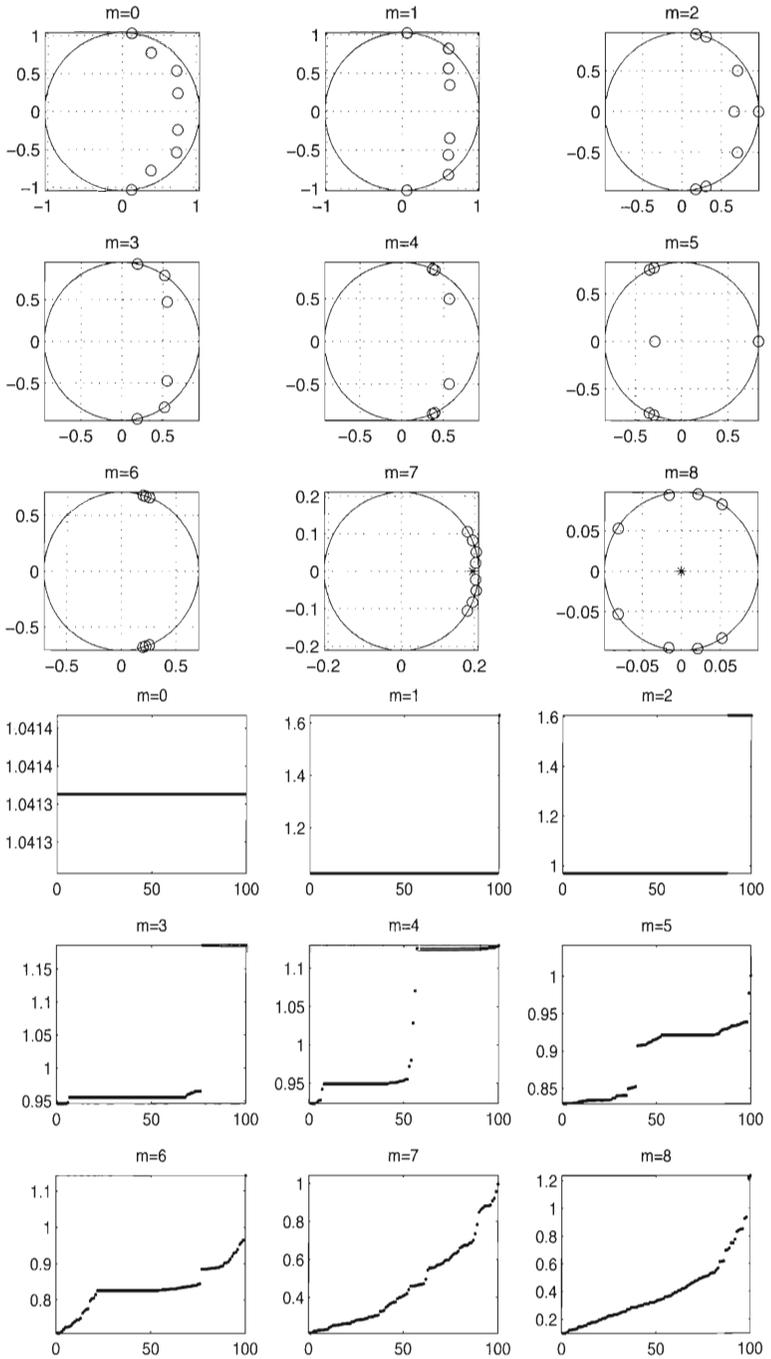


Figure 16.3. Static output feedback via spectral radius optimization with $n = 8$, $p = 1$. Top: optimized eigenvalues. Bottom: sorted final spectral radius values for 100 runs of BFGS

The scaling of the axes changes for each m , but for $m = 7$, we have a dramatic change in scaling, with all 8 eigenvalues nearly coalescing to one real eigenvalue of multiplicity 8. In this case, the eigenvalue optimization problem is equivalent to a root radius optimization problem with just one affine constraint, whose solution can be computed explicitly by the results in section 16.1.2. Applying corollary 16.1, we find that the optimal polynomial is indeed of the form $(\lambda - \gamma)^8$, and the asterisk marked on the panel for $m = 7$ marks the location of the multiple root γ . Finally, for $m = 8$, SOF reduces to the state feedback problem, and since the pair (F, G) is controllable, this can be solved by placing all the eigenvalues at the origin, so the optimal spectral radius is zero, as marked by the asterisk.

For both $m = 7$ and $m = 8$, we observe that BFGS is unable to reduce the spectral radius to close to its known optimal value. Instead, for $m = 7$, we find that the final eigenvalues are gathered on an arc of a circle with a radius somewhat greater than the optimal value of the spectral radius, and for $m = 8$, although the spectral radius is reduced by another 50% compared to $m = 7$, it is still not very small. These results are not surprising given the sensitivity of non-derogatory eigenvalues with such high multiplicity. For example, suppose \tilde{X} were an upper Jordan block of order 8 with a zero eigenvalue perturbed by placing the entry 10^{-8} in the lower left corner; then, the eigenvalues of \tilde{X} would lie equally spaced around the circle with radius 0.1. From this point of view, the results achieved by BFGS are remarkably good.

This naturally raises the question: how close to optimal are the results for $m = 1$ through $m = 6$? Although we do not know the optimal values in these cases, we can get a good idea by looking at the lower half of Figure 16.3. The first panel ($m = 0$) simply displays the constant value $\rho(F)$. Each subsequent panel ($m = 1$ through $m = 8$) displays the *sorted* final spectral radius values found by the 100 randomly initiated runs of the BFGS algorithm applied to minimize $\rho(F + GKH)$ over $K \in \mathbf{R}^{1 \times m}$. For $m = 1$, we find that all 100 final values are the same at least to plotting accuracy. For $m = 2$, two locally optimal values are found: the lower one is less than one indicating Schur stability, but the other is substantially greater than $\rho(F)$. For $m = 3$, we see that several different locally optimal values were found by BFGS, each one corresponding to a “plateau” in the plotted values. The one found most frequently is the second smallest. It seems a reasonable guess that the smallest is globally optimal in this case. We see plateaus for $m = 4, 5$ and 6 too, indicating the existence of several locally minimal values, but, especially for $m = 6$, there is no clear plateau at the left end of the plot, indicating that the globally minimal value is not well approximated. Again, this is not surprising, given the sensitivity of the multiple eigenvalues that were found in these cases. For $m = 7$, we see no convincing plateaus, but nonetheless, the irregularity of the optimal value curve at the top right may indicate approximate convergence to local minimizers for some runs. Finally, for $m = 8$ we see no plateaus at all, which is to be expected since the pole placement theorem implies that minimizers of ρ must have all eigenvalues equal to zero.

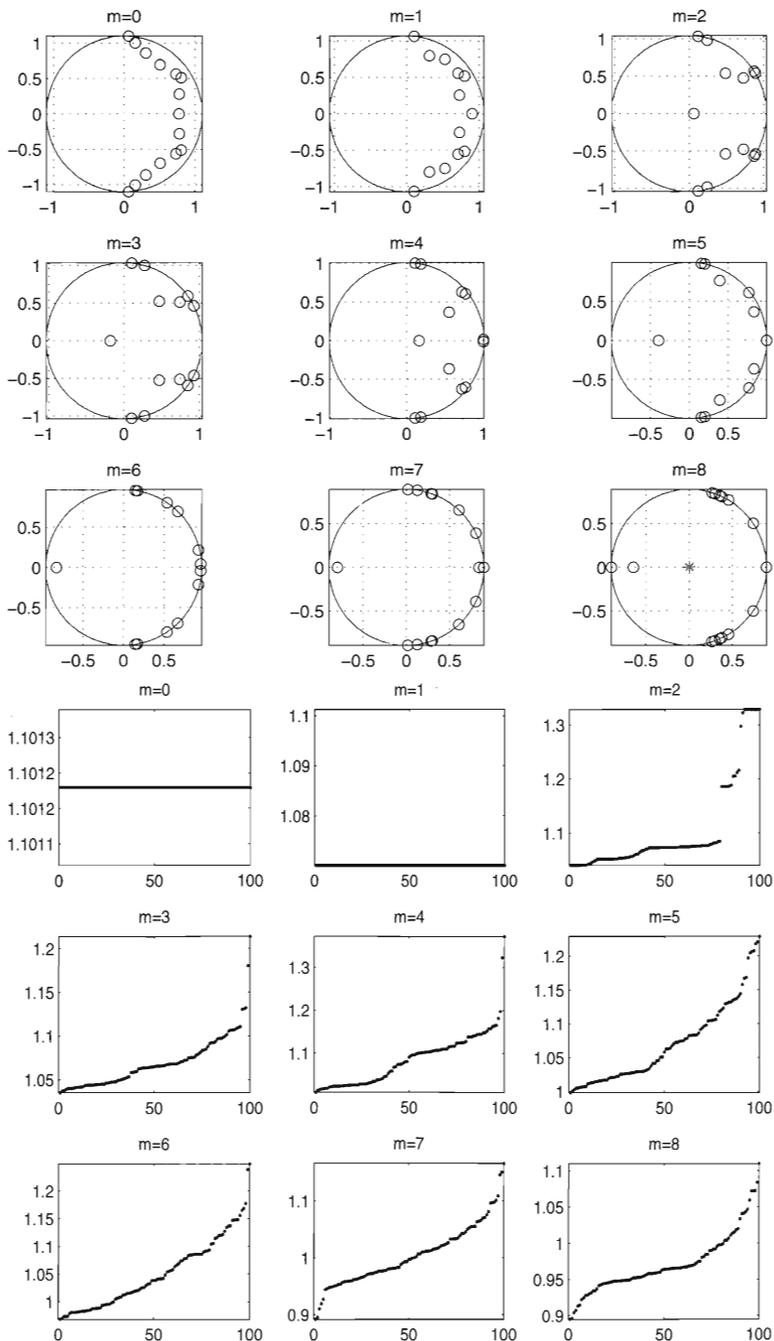


Figure 16.4. Static output feedback via spectral radius optimization with $n = 15$, $p = 2$. Top: optimized eigenvalues. Bottom: sorted final spectral radius values for 100 runs of BFGS

Now let us consider a larger problem, setting F to the Grcar matrix of order $n = 15$ scaled by 0.5 and setting G to the 15×2 matrix with the first column $[1, 1, \dots, 1]^T$ and the second column $[1, -1, 1, \dots, -1, 1]^T$. Equation [16.1] no longer applies, so the eigenvalue optimization problem no longer reduces to a polynomial root optimization problem with affine constraints on the coefficients. The top part of Figure 16.4 shows the eigenvalues obtained by using BFGS to minimize the spectral radius, for m ranging from 1 to 8. Every panel indicates the presence of active multiple eigenvalues, even for $m = 1$, for which there is an active double conjugate pair. For $m = 8$, we know that the optimal spectral radius is zero for generic (F, G, H) as $mp > n$. Turning to the sorted final function values in the lower half of Figure 16.4, there is just one locally minimal value for $m = 1$, but there are several clear plateaus of locally optimal values for $m = 2$. However, BFGS is unable to approximate the optimal values accurately for $m \geq 3$, no doubt because of the high eigenvalue multiplicities, so it is hard to judge how many local minimizers were approximated.

16.2.4. The Diaconis–Holmes–Neal Markov chain

An eigenvalue optimization problem with a very different character is provided by the analysis of a non-reversible Markov chain for Monte Carlo simulation [DIA 00]. For $\xi \in [0, 1]$, the associated doubly stochastic transition matrix $A(\xi) \in \mathbf{R}^{2n \times 2n}$ is

$$\begin{bmatrix} 0 & 1 - \xi & & & & & \xi & & 0 \\ & & 1 - \xi & & & & & \xi & \\ & & & \ddots & & & & & \ddots \\ & & & & 1 - \xi & \xi & & & \\ & & & & \xi & 1 - \xi & & & \\ & & & & & & \ddots & & \\ & & & & & & & 1 - \xi & \\ \xi & \xi & & & & & & 1 - \xi & \\ 1 - \xi & & & & & & & & \xi \end{bmatrix}.$$

Diaconis *et al.* [DIA 00] showed that for $\xi = 1/n$, the non-reversible Markov chain defined by this transition matrix reaches a stationary state in $O(n)$ steps, compared to $O(n^2)$ steps for a similar reversible chain. (Non-reversible chains have non-symmetric transition matrices, while reversible chains have symmetric transition matrices.)

The rate of convergence of the chain is determined by

$$\tilde{\rho}(A(\xi)) = \max \{ |\lambda| : \det(\lambda I - A(\xi)) = 0, \lambda \in \mathbf{C}, \lambda \neq 1 \}.$$

We call this quantity the *reduced spectral radius*, as it is the spectral radius after the eigenvalue one is removed. We call an eigenvalue active if its modulus equals the reduced spectral radius and inactive if it is smaller (the eigenvalue one is neither active nor inactive). Using formulas for the eigenvalues of $A(\xi)$ given in [DIA 00], it is easy to prove that $\tilde{\rho}(A(\xi))$ is minimized over $\xi \in [0, 1]$ by

$$\tilde{\xi} = \frac{\sin(\pi/n)}{1 + \sin(\pi/n)} > \frac{1}{n}.$$

For $\xi < \tilde{\xi}$, there are $n - 1$ active conjugate pairs and one inactive eigenvalue. For $\xi = \tilde{\xi}$, two conjugate pairs coalesce to two double real non-derogatory eigenvalues, resulting in two active double real eigenvalues and $n - 3$ active simple conjugate pairs. For $\xi > \tilde{\xi}$, the double eigenvalues each split into two real eigenvalues, increasing $\tilde{\rho}$ by a term proportional to $|\xi - \tilde{\xi}|^{1/2}$, with two active simple eigenvalues. The eigenvalues of $A(\xi)$ are shown in Figure 16.5(a), along with a circle centered at the origin with radius $\tilde{\rho}(A(\xi))$ for $n = 10$ and for $\xi = 1/n < \tilde{\xi}$, $\xi = \tilde{\xi}$ and $\xi > \tilde{\xi}$. Plots of $\tilde{\rho}(A(\xi))$ are shown in Figure 16.5(b) for $\xi \in [0, 1]$ for both $n = 10$ and $n = 100$.

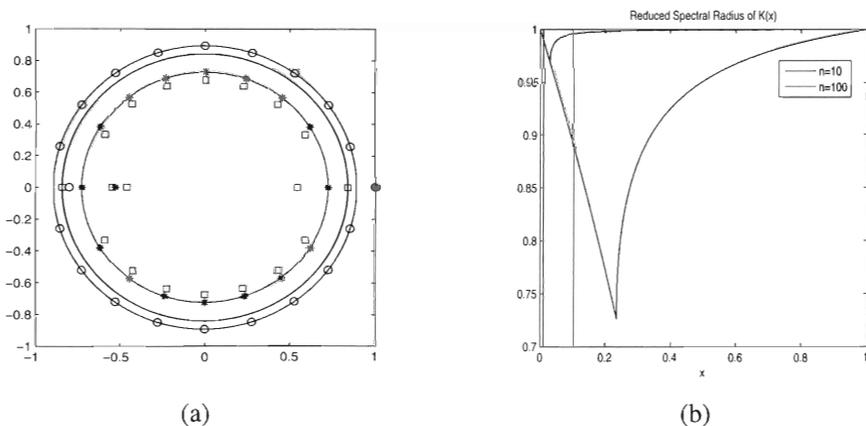


Figure 16.5. a) Eigenvalues of the $2n \times 2n$ transition matrix $A(\xi)$ for $n = 10$, along with circles of radius $\tilde{\rho}(A(\xi))$ centered at the origin. Small circles: eigenvalues when $\xi = 1/n$ (with $n - 1$ active complex conjugate pairs). Asterisks: eigenvalues when $\xi = \tilde{\xi}$ (two conjugate pairs have coalesced to double real eigenvalues). Small squares: eigenvalues when $\xi > \tilde{\xi}$ (the double eigenvalues have each split into a real pair, with $\tilde{\rho}$ increasing rapidly). b) Plots $\tilde{\rho}(A(\xi))$ for $\xi \in [0, 1]$, for $n = 10$ and $n = 100$. Note that $\tilde{\rho}(A(\tilde{\xi})) < \tilde{\rho}(A(1/n))$ and that $\tilde{\rho}(A(\tilde{\xi}))$ approaches one as n increases

It is not surprising that with one free parameter, we can only make one pair of eigenvalues coalesce. So, let us change the problem to have multiple parameters: for $x = [x_1, \dots, x_n]^T$, $x_j \in [0, 1]$, $j = 1, \dots, n$, we define the transition matrix $A(x) \in \mathbf{R}^{2n \times 2n}$ as

having order 2, 1 and 1. In the second example, $A(\tilde{x})$ again apparently has one active zero eigenvalue but with four Jordan blocks, respectively, having order 5, 3, 2 and 2. In derogatory cases like this, variational analysis of the spectral abscissa or spectral radius becomes very difficult even for simple examples. The special case with two Jordan blocks with order 2 and 1, respectively, is analyzed in detail in [GRU 13].

16.3. Concluding remarks

We have seen that the optimization problems discussed in this chapter typically lead to polynomials with multiple roots or matrices with non-derogatory multiple eigenvalues, and we have observed that the higher their multiplicity, the more these multiple roots or eigenvalues are sensitive to small perturbations; furthermore, computing these minimizers numerically is difficult. Instead of optimizing eigenvalues, we could consider optimizing *pseudospectra*. The ϵ -pseudospectrum of a matrix A , denoted $\sigma_\epsilon(A)$, is the set of points in the complex plane that are eigenvalues of matrices within ϵ in norm of a given matrix [TRE 05]. Fortunately, for the 2-norm, there is a more convenient equivalent definition using the singular value decomposition: $\sigma_\epsilon(A)$ is the set of points $z \in \mathbf{C}$ for which the smallest singular value of $A - zI$ is no greater than ϵ . Then, it is natural to define the pseudospectral radius $\rho_\epsilon(A)$ and pseudospectral abscissa $\alpha_\epsilon(A)$, respectively, as the largest of the moduli and largest of the real parts of the points in $\sigma_\epsilon(A)$. Algorithms to compute these functions are given in [MEN 05, BUR 03b]. It turns out that these functions are locally Lipschitz with respect to both A and ϵ [LEW 08, GÜR 12], although the pseudospectrum itself is not. For this reason, neglecting the cost of computing the pseudospectral radius and abscissa, local optimization of these functions is less difficult than for the spectral radius and abscissa: gradient sampling is known to converge to non-smooth stationary points in the locally Lipschitz case, and while no such result is known for BFGS, the boundedness of the gradients in the Lipschitz case allows for the use of a practical termination condition described in [LEW 13, section 6.3], instead of running the BFGS iteration until the line search fails or a maximum iteration count is exceeded, as described in section 16.2.3.

See [BUR 03a] for some examples of optimizing the pseudospectral abscissa for some SOF problems arising in applications. In [BUR 02], we used a different approach for replacing the spectral abscissa by a robust alternative exploiting the well-known Lyapunov characterization of stability, but this required solving nonconvex optimization problems in a much larger variable space. Finally, we mention that using many of the ideas discussed in this chapter, we have built a software package for solving practical problems that arise in controller design [BUR 06a].⁴

⁴ www.cs.nyu.edu/overton/software/hifoo.

16.4. Acknowledgments

The author thanks O. Kirillov for the kind invitation to the Banff workshop where this chapter was presented. He also thanks J. Willems for providing references to papers on generic solutions to the SOF problem when $mp < n$. Financial support from the U.S. National Science Foundation under Grant DMS-1317205 is also gratefully acknowledged.

16.5. Bibliography

- [ARN 71] ARNOLD V.I., “On matrices depending on parameters”, *Russian Mathematical Surveys*, vol. 26, pp. 29–43, 1971.
- [BLO 95] BLONDEL V., GEVERS M., LINDQUIST A., “Survey on the state of systems and control”, *European Journal of Control*, vol. 1, pp. 5–23, 1995.
- [BLO 10] BLONDEL V.D., GÜRBÜZBALABAN M., MEGRETSKI A., OVERTON M.L., “Explicit solutions for root optimization of a polynomial family”, *49th IEEE Conference on Decision and Control (CDC), 2010*, IEEE, pp. 485–488, 2010.
- [BLO 12] BLONDEL V.D., GÜRBÜZBALABAN M., MEGRETSKI A., OVERTON M.L., “Explicit solutions for root optimization of a polynomial family with one affine constraint,” *IEEE Transactions on Automatic Control*, vol. 57, pp. 3078–3089, 2012.
- [BOR 05] BORWEIN J.M., LEWIS A.S., “*Convex Analysis and Nonlinear Optimization: Theory and Examples, 2nd ed.*”, Springer, New York, 2005.
- [BUR 01a] BURKE J.V., OVERTON M.L., “Variational analysis of non-Lipschitz spectral functions”, *Mathematical Programming*, vol. 90, pp. 317–352, 2001.
- [BUR 01b] BURKE J.V., OVERTON M.L., “Variational analysis of the abscissa map for polynomials”, *SIAM Journal on Control and Optimization*, vol. 39, pp. 1651–1676, 2001.
- [BUR 01c] BURKE J.V., LEWIS A.S., OVERTON M.L., “Optimizing matrix stability”, *Proceedings of the American Mathematical Society*, vol. 129, pp. 1635–1642, 2001.
- [BUR 02] BURKE J.V., LEWIS A.S., OVERTON M.L., “Two numerical methods for optimizing matrix stability”, *Linear Algebra and its Applications*, vol. 351–352, pp. 117–145, 2002.
- [BUR 03a] BURKE J.V., LEWIS A.S., OVERTON M.L., “A nonsmooth, nonconvex optimization approach to robust stabilization by static output feedback and low-order controllers”, *Fourth IFAC Symposium on Robust Control Design*, Milan, 2003.
- [BUR 03b] BURKE J.V., LEWIS A.S., OVERTON M.L., “Robust stability and a criss-cross algorithm for pseudospectra”, *IMA Journal of Numerical Analysis*, vol. 23, pp. 359–375, 2003.
- [BUR 05a] BURKE J.V., LEWIS A.S., OVERTON M.L., “A robust gradient sampling algorithm for nonsmooth, nonconvex optimization”, *SIAM Journal on Optimization*, vol. 15, pp. 751–779, 2005.

- [BUR 05b] BURKE J.V., LEWIS A.S., OVERTON M.L., “Variational analysis of functions of the roots of polynomials”, *Mathematical Programming*, vol. 104, pp. 263–292, 2005.
- [BUR 06a] BURKE J.V., HENRION D., LEWIS A.S., OVERTON M.L., “HIFOO - a MATLAB package for fixed-order controller design and H_∞ optimization”, *Fifth IFAC Symposium on Robust Control Design*, Toulouse, 2006.
- [BUR 06b] BURKE J.V., HENRION D., LEWIS A.S., OVERTON M.L., “Stabilization via nonsmooth, nonconvex optimization”, *IEEE Transactions on Automatics Control*, vol. 51, pp. 1760–1769, 2006.
- [BUR 12] BURKE J.V., EATON J., “On the subdifferential regularity of max root functions for polynomials”, *Nonlinear Analysis: Theory, Methods and Applications*, vol. 75, p. 1168–1187, 2012.
- [CHE 79a] CHEN R., “On the problem of direct output feedback stabilization”, *Proceedings of MTNS*, pp. 412–414, 1979.
- [CHE 79b] CHEN R., Output feedback stabilization of linear systems, PhD Thesis, University of Florida, 1979.
- [CLA 83] CLARKE F.H., *Optimization and Nonsmooth Analysis*, John Wiley, New York, 1983. Reprinted by SIAM, Philadelphia, PA, 1990.
- [DIA 00] DIACONIS P., HOLMES S., NEAL R.M., “Analysis of a nonreversible Markov chain sampler”, *Annals of Applied Probability*, vol. 10, pp. 726–752, 2000.
- [DOR 99] DORATO P., *Analytic Feedback System Design: An Interpolation Approach*, Thomson Publishing Services, UK, 1999.
- [GAD 07] GADE K.K., OVERTON M.L., “Optimizing the asymptotic convergence rate of the Diaconis-Holmes-Neal sampler”, *Advances in Applied Mathematics*, vol. 38, no. 3, pp. 382–403, 2007.
- [GÜR 12] GÜRBÜZBALABAN M., OVERTON M.L., “On Nesterov’s nonsmooth Chebyshev-Rosenbrock functions”, *Nonlinear Analysis, Theory, Methods and Applications*, vol. 75, pp. 1282–1289, 2012.
- [GRU 11] GRUNDEL S., Eigenvalue optimization in C^2 subdivision and boundary subdivision, PhD Thesis, New York University, 2011. Available at <http://cs.nyu.edu/overton/phdtheses/sara.pdf>.
- [GRU 13] GRUNDEL S., OVERTON, M.L., “Variational analysis of the spectral abscissa at a matrix with a nongeneric multiple eigenvalue”, *Set-Valued and Variational Analysis*, 2013.
- [HER 77] HERMANN R., MARTIN C.F., “Applications of algebraic geometry to system theory, Part I”, *IEEE Transactions on Automatic Control*, vol. 22, pp. 19–25, 1977.
- [HEN 06] HENRION D., OVERTON, M.L., Maximizing the closed loop asymptotic decay rate for the two-mass-spring control problem, Technical Report 06342, LAAS-CNRS, March 2006. Available at <http://homepages.laas.fr/henrion/Papers/massspring.pdf>.
- [HOR 85] HORN R.A., JOHNSON C.R., *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge, UK, 1985.

- [HOR 13] HORN, R.A., JOHNSON C.R., *Matrix Analysis*, 2nd ed., Cambridge University Press, Cambridge, UK, 2013.
- [KIM 94] KIMURA H., “Pole assignment by output feedback: a longstanding open problem”, *33rd IEEE Conference on Decision and Control (CDC)*, 1994, IEEE, pp. 2101–2105, 1994.
- [LEW 03] LEWIS A.S., “The mathematics of eigenvalue optimization”, *Mathematical Programming*, vol. 97, nos. 1–2, Ser. B, pp. 155–176, ISMP, Copenhagen, 2003.
- [LEW 08] LEWIS A.S., PANG C.H.J. “Variational analysis of pseudospectra”, *SIAM Journal on Optimization*, 19:1048–1072, 2008.
- [LEW 13] LEWIS A.S., OVERTON M.L., “Nonsmooth optimization via quasi-Newton methods”, *Mathematical Programming*, vol. 141, pp. 135–163, 2013.
- [MAR 66] MARDEN M., *Geometry of Polynomials*, American Mathematical Society, 1966.
- [MEN 05] MENGI E., OVERTON M.L., “Algorithms for the computation of the pseudospectral radius and the numerical radius of a matrix”, *IMA Journal on Numerical Analysis*, 25:648–669, 2005.
- [NOC 06] NOCEDAL J., WRIGHT S.J., *Nonlinear Optimization*, 2nd ed., Springer, New York, 2006.
- [RAN 89] RANTZER A., “Equivalence between stability of partial realizations and feedback stabilization-applications to reduced order stabilization”, *Linear Algebra and its Applications*, vol. 122/123/124, pp. 641–653, 1989.
- [ROC 70] ROCKAFELLAR R.T., *Convex Analysis*, Princeton University Press, Princeton, 1970.
- [ROC 98] ROCKAFELLAR R.T., WETS R.J.B., *Variational Analysis*, Springer, New York, 1998.
- [TRE 05] TREFETHEN L.N., EMBREE M., *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, 2005.
- [WAN 96] WANG X.A., “Grassmannian, central projection, and output feedback pole assignment of linear systems”, *IEEE Transactions on Automatic Control*, vol. 41, no. 6, pp. 786–794, 1996.
- [WIL 97] WILLEMS. J.C., “Generic eigenvalue assignability by real memoryless output feedback made simple”, in PAULRAJ A., ROYCHOWDHURY V., SCHAPER C.D. (eds), *Communications, Computation, Control and Signal Processing*, Kluwer, pp. 343–354, 1997.
- [WON 85] WONHAM W.M., *Linear Multivariable Control: A Geometric Approach*, 3rd ed., Springer-Verlag, 1985.
- [WRI 02] WRIGHT T.G., “EigTool: a graphical tool for nonsymmetric eigenproblems”, 2002. Available at <http://web.comlab.ox.ac.uk/pseudospectra/eigtool/>.