On Properties of Univariate Max Functions at Local Maximizers

Tim Mitchell* Michael L. Overton†

August 17, 2021

Abstract

More than three decades ago, Boyd and Balakrishnan established a regularity result for the two-norm of a transfer function at maximizers. Their result extends easily to the statement that the maximum eigenvalue of a univariate real analytic Hermitian matrix family is twice continuously differentiable, with Lipschitz second derivative, at all local maximizers, a property that is useful in several applications that we describe. We also investigate whether this smoothness property extends to max functions more generally. We show that the pointwise maximum of a finite set of q-times continuously differentiable univariate functions must have zero derivative at a maximizer for $q = 1$, but arbitrarily close to the maximizer, the derivative may not be defined, even when $q = 3$ and the maximizer is isolated.

Keywords: univariate max functions · eigenvalues of Hermitian matrix families · H-infinity norm · numerical radius · optimization of passive systems

MSC (2020): 49J52 · 65F99

1 Introduction

Let $\mathbb{H}^n$ denote the space of $n \times n$ complex Hermitian matrices, let $\mathcal{D} \subseteq \mathbb{R}$ be open, and let $H : \mathcal{D} \to \mathbb{H}^n$ denote an analytic Hermitian matrix family in one real variable, i.e., for all $x \in \mathcal{D}$ and all $i, j \in \{1, \ldots, n\}$, there exist coefficients $a_0, a_1, a_2, \ldots$ such that the power series $\sum_{k=0}^{\infty} a_k(t-x)^k$ converges to $H_{ij}(t) = \overline{H_{ji}(t)}$ for all t in a neighborhood of x. For a generic family H, the eigenvalues of $H(t)$ are simple for all $t \in \mathcal{D}$; often known as the von Neumann-Wigner crossing-avoidance rule [vNW29], this phenomenon is emphasized in [Lax07, section 9.5], where it is also illustrated on the front cover. The reason is simple: the real codimension of the subspace of Hermitian matrices with an eigenvalue of multiplicity m is $m^2 - 1$, so to obtain a double eigenvalue one would need three parameters generically; when the matrix family is real symmetric, the analogous codimension is $\frac{m(m+1)}{2} - 1$, so one would need two parameters generically. When there are no multiple eigenvalues, the ordered eigenvalues of $H(t)$, say, $\mu_j(t)$ for $j = 1, \ldots, n$, are all real analytic functions.

Let $\lambda_{\max} : \mathbb{H}^n \to \mathbb{R}$ and $\lambda_{\min} : \mathbb{H}^n \to \mathbb{R}$ denote algebraically largest and smallest eigenvalue, respectively. In the absence of multiple eigenvalues, $\lambda_{\max} \circ H$ and $\lambda_{\min} \circ H$ are both smooth functions of t. However, for the nongeneric family $H(t) = \operatorname{diag}(t, -t)$, a double eigenvalue occurs at $t = 0$. By a theorem of Rellich, given in Section 4, the eigenvalues can be written as two real analytic functions, $\mu_1(t) = t$ and $\mu_2(t) = -t$, but we must give up the property that these functions are ordered near zero. Consequently, the function $\lambda_{\max} \circ H$ is not differentiable at its minimizer $t = 0$.

In contrast, the function $\lambda_{\max} \circ H$ is *unconditionally* $\mathcal{C}^2$, i.e., twice continuously differentiable, with Lipschitz second derivative, near all its local *maximizers*, regardless of eigenvalue multiplicity at these maximizers. As we explain below, this observation is a straightforward extension of a well-known result of Boyd and Balakrishnan [BB90] established more than three decades ago. One purpose of this paper is to bring attention to the more general result, as it is useful in a number of applications. We also investigate whether this smoothness property extends to max functions more generally. We show that

*Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, 39106 Germany, mitchell@mpi-magdeburg.mpg.de, ORCID: 0000-0002-8426-0242.

†Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA. mo1@nyu.edu, ORCID: 0000-0002-6563-6371. Supported in part by U.S. National Science Foundation Grant DMS-2012250.

the pointwise maximum of a finite set of continuously differentiable univariate functions must have zero derivative at a maximizer. However, arbitrarily close to the maximizer, the derivative may not be defined, even if the functions are three times continuously differentiable and the maximizer is isolated.

2 Properties of max functions at local maximizers

Let $\mathcal{D} \subset \mathbb{R}$ be open, $\mathcal{I} = \{1, \ldots, n\}$, and $f_j : \mathcal{D} \to \mathbb{R}$ be continuous for all $j \in \mathcal{I}$, and define

$$f_{\max}(t) \coloneqq \max_{j \in \mathcal{I}} f_j(t). \tag{2.1}$$

Lemma 2.1. *Let $x \in \mathcal{D}$ be any local maximizer of $f_{\max}$ with $f_{\max}(x) = \gamma$ and let $\mathcal{I}_\gamma = \{j \in \mathcal{I} : f_j(x) = \gamma\}$. Then*

(i) *for all $j \in \mathcal{I}_\gamma$, x is a local maximizer of f_j and*

(ii) *for all $j \in \mathcal{I} \setminus \mathcal{I}_\gamma$, $f_j(x) < \gamma$.*

We omit the proof as it is elementary.

We now consider adding additional assumptions on the smoothness of the f_j, writing $f_j \in \mathcal{C}^q$ to mean f_j is q-times continuously differentiable. Clearly, assuming that the f_j are $\mathcal{C}^0$ is not sufficient to obtain differentiability at maximizers (e.g., $f_{\max}(t) = f_1(t) = -|t|$), but $\mathcal{C}^1$ is sufficient.

Theorem 2.2. *Let $x \in \mathcal{D}$ be any local maximizer of $f_{\max}$ with $f_{\max}(x) = \gamma$. Suppose that for all $j \in \mathcal{I}$, f_j is $\mathcal{C}^1$ at x. Then $f_{\max}$ is differentiable at x with $f'_{\max}(x) = 0$.*

Proof. Since the functions f_j are continuous, clearly $f_{\max}$ is also continuous, and without loss of generality, we can assume that $\gamma = 0$. Suppose that $f'_{\max}$ does not exist at x or does not equal zero, i.e., there exists some sequence $\{\varepsilon_k\}$ with $\varepsilon_k \to 0$ such that $\lim_{k\to\infty} \frac{f_{\max}(x+\varepsilon_k)}{\varepsilon_k}$ does not exist or is not zero. Since $\mathcal{I}$ is finite, there exist a $j \in \mathcal{I}$ and a subsequence $\{\varepsilon_{k_\ell}\}$ such that $f_{\max}(x + \varepsilon_{k_\ell}) = f_j(x + \varepsilon_{k_\ell})$ for all k_ℓ, which implies that f'_j either does not exist or is not zero at x. However, as f_j is $\mathcal{C}^1$ and with local maximizer x by Lemma 2.1, it must be that $f'_j(x) = 0$; hence, we have a contradiction. □

Assuming that the f_j are $\mathcal{C}^1$ at (or near) a maximizer is not sufficient to obtain that $f_{\max}$ is twice differentiable at this point. For example, if

$$f_1(t) = \begin{cases} -t^2 & \text{if } t \le 0 \\ -3t^2 & \text{if } t > 0 \end{cases} \qquad \text{and} \qquad f_2(t) = -2t^2,$$

then the second derivative of $f_{\max} = \max(f_1, f_2)$ does not exist at the maximizer $t = 0$, as $f'_{\max}(t) = -2t$ on the left and $-4t$ on the right, so $\lim_{t\to 0} \frac{f'_{\max}(t)}{t}$ does not exist at $t = 0$. In this example, $f_{\max}$ is continuously differentiable at $t = 0$, but this does not hold in general, even when assuming that the f_j are $\mathcal{C}^3$ near a maximizer; see Remark 2.4 below. However, we do have the following result.

Theorem 2.3. *Let $x \in \mathcal{D}$ be any local maximizer of $f_{\max}$ with $f_{\max}(x) = \gamma$. Suppose that for all $j \in \mathcal{I}$, f_j is $\mathcal{C}^3$ near x. Then for all sufficiently small $|\varepsilon|$,*

$$f_{\max}(x + \varepsilon) = \gamma + M\varepsilon^2 + O(|\varepsilon|^3), \tag{2.2}$$

where $M = \frac{1}{2}\left(\max_{j\in\mathcal{I}_\gamma} f''_j(x)\right) \le 0$. If the $\mathcal{C}^3$ assumption is reduced to $\mathcal{C}^2$, then $f_{\max}(x+\varepsilon) = \gamma + \mathcal{O}(\varepsilon^2)$.

Proof. Let $\gamma = f_{\max}(x)$ and let $\mathcal{I}_\gamma = \{j \in \mathcal{I} : f_j(x) = \gamma\}$. By Lemma 2.1, we have that x is also a local maximizer of f_j for all $j \in \mathcal{I}_\gamma$ and $f_j(x) < \gamma$ for all $j \in \mathcal{I} \setminus \mathcal{I}_\gamma$. Since the f_j are Lipschitz near x,

$$f_{\max}(x + \varepsilon) = \max_{j\in\mathcal{I}_\gamma} f_j(x + \varepsilon)$$

holds for all sufficiently small $|\varepsilon|$. For each $j \in \mathcal{I}_\gamma$, by Taylor's Theorem we have that

$$\begin{aligned} f_j(x + \varepsilon) &= f_j(x) + f'_j(x)\varepsilon + \tfrac{1}{2}f''_j(x)\varepsilon^2 + \tfrac{1}{6}f'''_j(\tau_j)\varepsilon^3 \\ &= \gamma + \tfrac{1}{2}f''_j(x)\varepsilon^2 + O(|\varepsilon|^3) \end{aligned}$$

for τ_j between x and $x + \varepsilon$. Taking the maximum of the equation above over all $j \in \mathcal{I}_\gamma$ yields (2.2). The proof for the $\mathcal{C}^2$ case follows analogously. □

Remark 2.4. *Even with the $\mathcal{C}^3$ assumption, $f_{\max}$ is not necessarily continuously differentiable at maximizers, let alone twice differentiable. A simple counterexample is given by $f_1(t) = t^8(\sin(\frac{1}{t}) - 1)$ and $f_2(t) = t^8(\sin(\frac{1}{2t}) - 1)$, with $f_1(0) = f_2(0) = 0$, where f_1 and f_2 are $\mathcal{C}^3$ but not $\mathcal{C}^4$ at $t = 0$. However, in this case the maximizer $t = 0$ of $f_{\max}$ is not an isolated maximizer. In contrast, in Section 3, we construct a counterexample where the f_j are $\mathcal{C}^3$ functions, and for which $f_{\max}$ has an isolated maximizer, yet the derivative of $f_{\max}$ does not exist at points arbitrarily close to this maximizer. It seems that this counterexample can be extended to apply to $\mathcal{C}^q$ functions for any $q \geq 1$. The key point of both of these counterexamples is not that the f_j are insufficiently smooth* per se*, but that the f_j cross each other infinitely many times near maximizers.*

In light of Remark 2.4, we now make a much stronger assumption.

Theorem 2.5. *Given a maximizer x of $f_{\max}$, suppose there exist $j_1, j_2 \in \mathcal{I}$, possibly equal, such that, for all sufficiently small $\varepsilon > 0$, $f_{\max}(x - \varepsilon) = f_{j_1}(x - \varepsilon)$ and $f_{\max}(x + \varepsilon) = f_{j_2}(x + \varepsilon)$, with f_{j_1} and f_{j_2} both $\mathcal{C}^3$ near x. Then $f_{\max}$ is twice continuously differentiable, with Lipschitz second derivative, near x.*

Proof. It is clear that $f_{j_1}(x) = f_{j_2}(x) = \gamma$ and $f'_{j_1}(x) = f'_{j_2}(x) = 0$. By Theorem 2.3, both $f''_{j_1}(x)$ and $f''_{j_2}(x)$ are equal to M, so $f_{\max}$ is locally described by two $\mathcal{C}^3$ pieces whose function values and first and second derivatives agree at x. Hence, $f_{\max}$ is $\mathcal{C}^2$ with Lipschitz second derivative near x. □

The assumptions of Theorem 2.5 hold if the f_j are real analytic [KP02, Corollary 1.2.7]. In particular, this holds if the f_j are eigenvalues of a univariate real analytic Hermitian matrix function, as we discuss in Section 4. First, we present the $\mathcal{C}^3$ counterexample mentioned above.

3 An example with $\mathcal{C}^3$ functions f_j and an isolated maximizer for which $f_{\max}$ is not continuously differentiable at x

Let $l_k = \frac{1}{2^k}$, and $f_1 : [-1, 1] \to \mathbb{R}$ be defined by

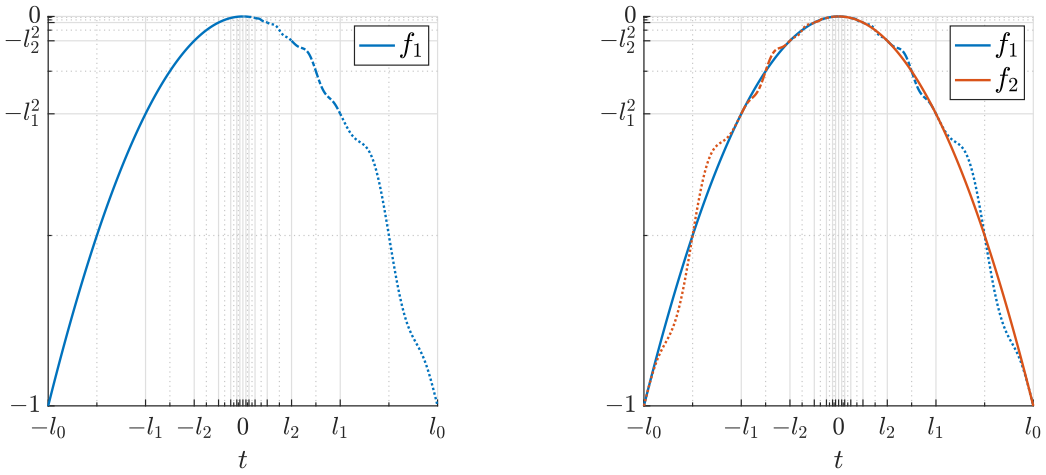
$$f_1(t) = \begin{cases} p_k(t) & \text{if } t \in [l_{k+1}, l_k] \text{ for } k = 0, 1, 2, \ldots \\ -t^2 & \text{if } t \in [-1, 0] \end{cases} \tag{3.1}$$

where p_k is a (piece of a) degree-nine polynomial chosen such that at

1. l_{k+1} (the left endpoint), p_k and $-t^2$ agree up to and including their respective third derivatives,
2. l_k (the right endpoint), p_k and $-t^2$ agree up to and including their respective third derivatives,
3. $t_k = \frac{1}{2}(l_{k+1} + l_k)$ (the midpoint), p_k and $-t^2$ agree, but the first derivative of p_k is $s_k \neq 1$ times the value of the first derivative of $-t^2$.

For any k, the degree-nine polynomial p_k is uniquely determined by the ten algebraic constraints given above. If we choose $s_k = 1$, then p_k is simply $-t^2$. However, by choosing $s_k > 1$ but sufficiently close to 1, then p_k must be strictly decreasing between its endpoints l_{k+1} and l_k and cross $-t^2$ at t_k. If this is done for all k, it follows that $t = 0$ must be an isolated maximizer of f_1. See Figure 1a for a plot of f_1 with $s_k = 2$ for all k; the choice $s_k = 2$ is not close to 1 but was chosen to make the features of f_1 easily seen.

Now define $f_2(t) = f_1(-t)$, i.e., the graph of f_2 is a reflection of the graph of f_1 across the vertical line $t = 0$. Figure 1b shows f_1 and f_2 plotted together, again with $s_k = 2$, showing how they cross at every t_k. Recall that by our construction, their respective first three derivatives match at each l_k, but their first derivatives do not match at any t_k. Figure 2 shows plots of the first three derivatives of f_1 for two different sequences $\{s_k\}$ respectively defined by $s_k = 1 + 2^{-k}$ and $s_k = 1 + 2^{-2k}$. The rightmost plots in Figure 2 indicate that the first choice for sequence $\{s_k\}$ does not converge to 1 fast enough for f'''_1 to exist and be continuous at $t = 0$, but that the second sequence does. In fact, for this latter choice of sequence, we have the following pair of theorems respectively proving that f_1 is indeed $\mathcal{C}^3$ with $t = 0$ being an isolated maximizer. We defer the proofs to Appendix A as they are a bit technical, and in Appendix B, we discuss why $s_k = 1 + 2^{-k}$ does not converge to 1 sufficiently fast for $f'''_1(0)$ to exist.



(a) (b)

Figure 1: Plots of f_1 and f_2 with $s_k = 2$; their $-t^2$ parts are shown in solid, while their p_k parts are shown in dotted for k even and dash-dot for k odd.

Theorem 3.1. *For f_1 defined in (3.1), if $s_k = 1 + 2^{-2k}$, then f_1 is $\mathcal{C}^3$ on its domain $[-1, 1]$.*

Theorem 3.2. *For f_1 defined in (3.1), if $s_k = 1 + 2^{-2k}$, then $t = 0$ is an isolated maximizer of f_1, as well as an isolated maximizer of $f_{\max} = \max(f_1, f_2)$.*

Theorem 2.2 shows that $f_{\max} = \max(f_1, f_2)$ is differentiable at $t = 0$ with $f'_{\max}(0) = 0$. However, even though f_1 and f_2 are $\mathcal{C}^3$ and $t = 0$ is an isolated maximizer of $f_{\max}$ with the choice of $s_k = 1+2^{-2k}$, by construction, we have that (i) $t_k \to 0$ as $k \to \infty$, and (ii) $f_{\max}$ is nondifferentiable at every t_k. Hence, although $f_{\max}$ is differentiable at $t = 0$, it is not $\mathcal{C}^1$ at this point, let alone twice differentiable. Plots of $f_{\max}$ and its first and second derivatives are shown in Figure 3, where we see the discontinuities in $f'_{\max}$ for all t_k and $-t_k$.

Remark 3.3. *For any $q \geq 1$, it seems that the same argument extends to show that $f_{\max}$ is not necessarily $\mathcal{C}^1$ at $t = 0$ when defined by functions f_j that are $\mathcal{C}^q$, using polynomials p_k of degree $2q+3$. From computational investigations for $q \in \{1, 2, 3, 4, 5\}$, we conjecture that $s_k = 1 + 2^{-(k+1)}$ for $q = 1$ and $s_k = 1 + 2^{-(q-1)k}$ for $q \geq 2$ are suitable choices in general to obtain that f_1 is $\mathcal{C}^q$ with $t = 0$ being an isolated maximizer. It is not clear how to extend such an argument to the $\mathcal{C}^\infty$ case.*

4 Smoothness of eigenvalue extrema and applications

We will need the following well-known theorem:

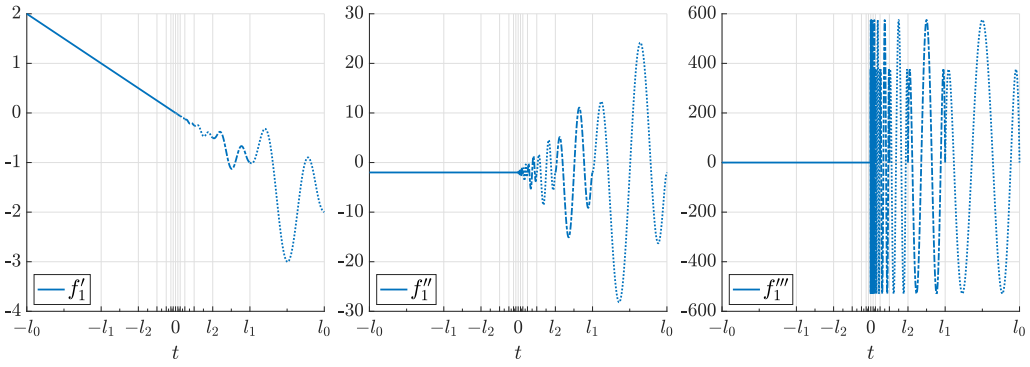
Theorem 4.1 (Rellich)**.** *Let $H : \mathcal{D} \to \mathbb{H}^n$ be an analytic Hermitian matrix family in one real variable. Let $x \in \mathcal{D}$ be given, and let $H(x)$ have eigenvalues $\tilde{\mu}_j \in \mathbb{R}$, $j = 1, \ldots, n$, not necessarily distinct. Then, for sufficiently small $|\varepsilon|$, the eigenvalues of $H(x + \varepsilon)$ can be expressed as convergent power series*

$$\mu_j(\varepsilon) = \tilde{\mu}_j + \tilde{\mu}_j^{(1)}\varepsilon + \tilde{\mu}_j^{(2)}\varepsilon^2 + \cdots, \quad j = 1, \ldots, n. \tag{4.1}$$

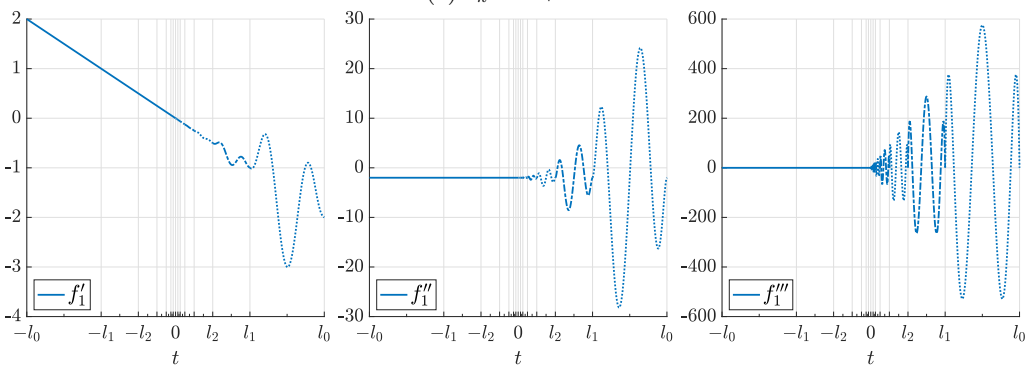
Theorem 4.1 is often stated as part of a deeper theorem of Rellich regarding power series expansion of the eigenvectors; in comparison, the proof of (4.1) is significantly easier, using the theory of algebraic functions to express the eigenvalues as fractional powers of ε and then arguing that, because H is Hermitian, non-integral fractional powers vanish [Kat82, pp. XIX–XX].

We now apply Theorems 4.1 and 2.5 to obtain smoothness results for eigenvalue extrema of univariate real analytic Hermitian matrix families, as well as analogous results for singular value extrema. Subsequently, we discuss how these results are useful in several important applications.

Theorem 4.2. *Let $H : \mathcal{D} \to \mathbb{H}^n$ be an analytic Hermitian matrix family in one real variable on an open domain $\mathcal{D} \subseteq \mathbb{R}$, and let $\lambda_{\max} : \mathbb{H}^n \to \mathbb{R}$ denote algebraically largest eigenvalue. Then $\lambda_{\max} \circ H$ is $\mathcal{C}^2$ with Lipschitz second derivative near all of its local maximizers.*



(a) $s_k = 1 + 2^{-k}$



(b) $s_k = 1 + 2^{-2k}$

Figure 2: Plots of the first three derivatives of f_1 for two different sequences $\{s_k\}$; their $-t^2$ parts are shown in solid, while their p_k parts are shown in dotted for k even and dash-dot for k odd.

Proof. Let $x \in \mathcal{D}$ be any local maximizer of $\lambda_{\max} \circ H$, with $H(x)$ having eigenvalues $\tilde{\mu}_j$. By Theorem 4.1, in a neighborhood of x, the eigenvalues of $H(x+\varepsilon)$ can be expressed as $\mu_j(\varepsilon)$, $j = 1, \ldots, n$, where the $\mu_j(\varepsilon)$ are locally given by the power series (4.1). Since $\lambda_{\max}(H(x+\varepsilon)) = \max_{j\in\{1,\ldots,n\}} \mu_j(\varepsilon)$ with all the μ_j analytic, we can apply Theorem 2.5 to these functions, completing the proof. □

Remark 4.3. *The proof of Theorem 4.2 is essentially the same as the proof given by Boyd and Balakrishnan [BB90], presented differently and in a more general context.*

Corollary 4.4. *Let $H : \mathcal{D} \to \mathbb{H}^n$ be an analytic Hermitian matrix family in one real variable on an open domain $\mathcal{D} \subseteq \mathbb{R}$. Then:*

(i) *$\lambda_{\min} \circ H$ is $\mathcal{C}^2$ near all of its local minimizers, where $\lambda_{\min}$ denotes algebraically smallest eigenvalue;*

(ii) *$\rho \circ H$ is $\mathcal{C}^2$ near all of its local maximizers, where ρ denotes spectral radius* $(\max(\lambda_{\max}, -\lambda_{\min}))$*;*

(iii) *$\rho_{\text{in}} \circ H$ is $\mathcal{C}^2$ near all of its local minimizers at which the minimal value is nonzero, where ρ_{in} denotes inner spectral radius (0 if H is singular, $\rho(H^{-1})^{-1}$ otherwise).*

Furthermore, in each case the second derivative is Lipschitz near the relevant maximizers/minimizers.

Proof. Statements (i) and (ii) follow from applying Theorem 4.2 to $-H$ and $\operatorname{diag}(H, -H)$, respectively. For (iii), apply (ii) to $\rho \circ H^{-1}$ and take the reciprocal. □

Corollary 4.5. *Let $A : \mathcal{D} \to \mathbb{C}^{m\times n}$ be an analytic matrix family in one real variable on an open domain $\mathcal{D} \subseteq \mathbb{R}$, let $\sigma_{\max}$ denote largest singular value, and let $\sigma_{\min}$ denote smallest singular value, noting that the latter is nonzero if and only if the matrix has full rank. Then:*

(i) *$\sigma_{\max} \circ A$ is $\mathcal{C}^2$ near all of its local maximizers, and*

(ii) *$\sigma_{\min} \circ A$ is $\mathcal{C}^2$ near all of its local minimizers at which the minimal value is nonzero.*

Furthermore, in each case the second derivative is Lipschitz near the relevant maximizers/minimizers.

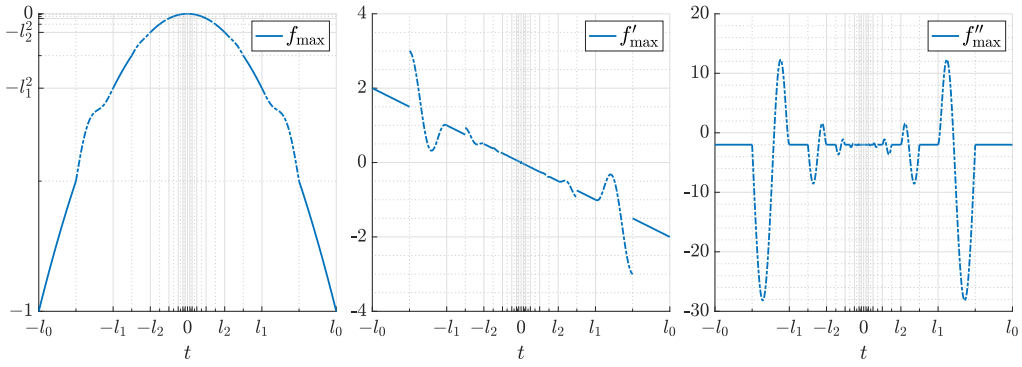


Figure 3: Plots of $f_{\max}$ and its first and second derivatives; their $-t^2$ parts are shown in solid, while their p_k parts are shown in dash-dot.

Proof. If $m \geq n$, consider the real analytic Hermitian matrix family $H : \mathcal{D} \to \mathbb{H}^n$ defined by

$$H(t) = A(t)^* A(t) = (\operatorname{Re} A(t) - \mathrm{i} \operatorname{Im} A(t))^{\mathsf{T}} (\operatorname{Re} A(t) + \mathrm{i} \operatorname{Im} A(t)),$$

whose eigenvalues are the squares of the singular values of $A(t)$. Then (i) and (ii), respectively, follow from applying Corollary 4.4 (ii) and (iii), respectively, to $H(t)$, and then taking the square root. If $n > m$, set $H(t) = A(t)A(t)^*$ instead. □

Corollary 4.5 (i) is the regularity result that Boyd and Balakrishnan established in [BB90]. For Corollary 4.5 (ii), note that the assumption that the minimal value of $\sigma_{\min} \circ A$ is nonzero is necessary; e.g., $\sigma_{\min}(t)$ is nonsmooth at its minimizer $t = 0$.

4.1 The $\mathcal{H}_\infty$ norm

This application was the original motivation for Boyd and Balakrishnan's work. Let $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, $C \in \mathbb{C}^{p\times n}$, and $D \in \mathbb{C}^{p\times m}$ and consider the linear time-invariant system with input and output:

$$\dot{x} = Ax + Bu, \tag{4.2a}$$
$$y = Cx + Du. \tag{4.2b}$$

Assume that A is asymptotically stable, i.e., its eigenvalues are all in the open left half-plane. An important quantity in control systems engineering and model-order reduction is the $\mathcal{H}_\infty$ norm of (4.2), which measures the sensitivity of the system to perturbation and can be computed by solving the following optimization problem:

$$\max_{\omega\in\mathbb{R}} \ \sigma_{\max}(G(\mathrm{i}\omega)), \tag{4.3}$$

where $G(\lambda) = C(\lambda I - A)^{-1}B + D$ is the transfer matrix associated with (4.2). Even though there is only one real variable, finding the global maximum of this function is nontrivial.

By extending Byer's breakthrough result on computing the distance to instability [Bye88], Boyd et al. [BBK89] developed a globally convergent bisection method to solve (4.3) to arbitrary accuracy. Shortly thereafter, a much faster algorithm, based on computing level sets of $\sigma_{\max}(G(\mathrm{i}\omega))$, was independently proposed in [BB90] and [BS90], with Boyd and Balakrishnan showing that this iteration converges quadratically [BB90, Theorem 5.1]. As part of their work, they showed that, with respect to the real variable ω, $\sigma_{\max}(G(\mathrm{i}\omega))$ is $\mathcal{C}^2$ with Lipschitz second derivative near any of its local maximizers [BB90, pp. 2–3]. Subsequently, this smoothness property has been leveraged to further accelerate computation of the $\mathcal{H}_\infty$ norm [GVDV98, BM18].

4.2 The numerical radius

Now consider the numerical radius of a matrix $A \in \mathbb{C}^{n\times n}$:

$$r(A) = \max\{|z| : z \in W(A)\}, \tag{4.4}$$

where $W(A) = \{v^*Av : v \in \mathbb{C}^n, \|v\|_2 = 1\}$ is the field of values (numerical range) of A. Following [HJ91, Ch. 1], the numerical radius can be computed by solving either

$$r(A) = \max_{\theta \in [0,2\pi)} \lambda_{\max}(H(\theta)) \qquad \text{or} \qquad r(A) = \max_{\theta \in [0,\pi)} \rho(H(\theta)), \tag{4.5}$$

where $H(\theta) = \frac{1}{2}\left(\mathrm{e}^{\mathrm{i}\theta}A + \mathrm{e}^{-\mathrm{i}\theta}A^*\right)$.

In [MO05], Mengi and the second author proposed the first globally convergent method guaranteed to compute $r(A)$ to arbitrary accuracy. This was done by employing a level-set technique that converges to a global maximizer of $\lambda_{\max} \circ H$, similar to the aforementioned method of [BB90, BS90] for the $\mathcal{H}_\infty$ norm, and observing, but not proving, quadratic convergence of the method. Quadratic convergence was later proved by Gürbüzbalaban in his PhD thesis [Gür12, Lemma 3.4.2], following the proof used in [BB90], showing that $\lambda_{\max} \circ H$ is $\mathcal{C}^2$ near maximizers.

4.3 Optimization of passive systems

Let $\mathcal{M} = \{A, B, C, D\}$ denote the system (4.2), but now with $m = p$ and the associated transfer function G being minimal and proper [ZDG96]. Mehrmann and Van Dooren [MVD20] have recently shown that another important problem is to compute the maximal value $\Xi \in \mathbb{R}$ such that for all $\xi < \Xi$, the related system $\mathcal{M}_\xi = \{A_\xi, B, C, D_\xi\}$ is strictly passive[1], where $A_\xi = A + \frac{\xi}{2}I_n$ and $D_\xi = D - \frac{\xi}{2}I_m$. Letting G_ξ be the transfer matrix associated with $\mathcal{M}_\xi$, by [MVD20, Theorem 5.1], the quantity Ξ is the unique root of

$$\gamma(\xi) \coloneqq \min_{\omega \in \mathbb{R}} \lambda_{\min}\left(G_\xi(\mathrm{i}\omega)^* + G_\xi(\mathrm{i}\omega)\right) = 0. \tag{4.6}$$

Note that in contrast to the univariate optimization problems discussed previously, computing Ξ is a problem in two real parameters, namely, ξ and ω. In [MVD20, section 5], Mehrmann and Van Dooren introduced both a bisection algorithm to compute Ξ, and an apparently faster "improved iteration" whose exact convergence properties were not established. However, using the fact that $\lambda_{\min}$ in (4.6) is $\mathcal{C}^2$ with Lipschitz second derivative near all its minimizers, as well as some other tools, the first author and Van Dooren have since established a rate-of-convergence result for this "improved iteration" and also presented a much faster and more numerically reliable algorithm to compute Ξ with quadratic convergence [MVD21].

5 Concluding remarks

We have shown that the maximum eigenvalue of a univariate real analytic Hermitian matrix family is unconditionally $\mathcal{C}^2$ near all its maximizers, with Lipschitz second derivative. Although the result is well known in the context of the maximum singular value of a transfer function, its generality and simplicity have apparently not been fully appreciated. We believe that this result and its corollaries may be useful in many applications, some of which were summarized in this paper. We also investigated whether this smoothness property extends to max functions more generally, showing that the pointwise maximum of a finite set of q-times continuously differentiable univariate functions must have zero derivative at a maximizer for $q = 1$, but arbitrarily close to the maximizer, the derivative may not be defined, even when $q = 3$ and the maximizer is isolated.

All figures and the symbolically computed coefficients of p_k given in Appendices A and B can be generated by MATLAB codes that are available upon request.

A Proofs of Theorems 3.1 and 3.2

Lemma A.1. *For f_1 defined in (3.1), if $s_k = 1 + 2^{-2k}$, then the coefficients of the polynomial $p_k(t) = \sum_{j=0}^{9} c_j t^j$ are:*

$$c_j = \begin{cases} z_j 2^{(j-4)k} - 1 & \text{if } j = 2 \\ z_j 2^{(j-4)k} & \text{otherwise} \end{cases} \qquad \text{with} \qquad \begin{aligned} z_9 &= -98304, & z_5 &= -3631104, & z_1 &= -61440, \\ z_8 &= 663552, & z_4 &= 2585088, & z_0 &= 4608. \\ z_7 &= -1966080, & z_3 &= -1210368, & & \\ z_6 &= 3354624, & z_2 &= 359424, & & \end{aligned}$$

[1] A strictly passive system is one whose stored energy is decreasing; for more a formal treatment, see [MVD20].

Proof. The coefficients were computed symbolically in MATLAB by solving the linear system defined by the generalized Vandermonde matrix and right-hand side determining each p_k in (3.1). These formulas were also verified by comparing with numerical computations. □

Proof of Theorem 3.1. Function f_1 defined in (3.1) is clearly $\mathcal{C}^3$ near any nonzero t, since our construction ensures that the first three derivatives of p_k and p_{k+1} match where they meet. We must show that it is also $\mathcal{C}^3$ at $t = 0$. First note that for the coefficients given in Lemma A.1, we can replace their dependency on k with a dependency on t by using $k = -\lceil \log_2 t \rceil$. Thus, f_1 can be written as follows:

$$f_1(t) = \begin{cases} \sum_{j=0}^{9} \tilde{c}_j t^j & \text{if } t > 0 \\ -t^2 & \text{if } t \in [-1, 0] \end{cases} \tag{A.1}$$

where $\tilde{c}_j$ is obtained by replacing k in c_j with $-\lceil \log_2 t \rceil$.

We begin by looking at the first derivative. For f_1' to exist and be continuous at $t = 0$,

$$f_1'(0) = \lim_{\varepsilon \to 0^+} \frac{f_1(0+\varepsilon) - f_1(0)}{\varepsilon} = \lim_{\varepsilon \to 0^+} \frac{f_1(\varepsilon)}{\varepsilon} = 0 \tag{A.2}$$

must hold, i.e., the derivative from the right (over the p_k pieces) must match the derivative from the left (over the $-t^2$ piece). To show that (A.2) holds, we show that each term in the sum in (A.1) divided by t goes to zero as $t \to 0^+$, i.e., that $\lim_{t\to 0^+} \tilde{c}_j t^{j-1} = 0$ for $j \in \{0, 1, \ldots, 9\}$. It is obvious that this holds for $j = 4$ since $c_4 = \tilde{c}_4 = z_4$ is a fixed number. To show the highest-order term ($j = 9$) vanishes as $t \to 0^+$, we can make use of the fact that $0 < 2^{-\lceil \log_2 t \rceil} \leq 2^{-\log_2 t} = t^{-1}$ holds for all $t > 0$, i.e.,

$$\lim_{t\to 0^+} \left| z_9 2^{-5\lceil \log_2 t \rceil} t^8 \right| \leq \lim_{t\to 0^+} \left| z_9 (t^{-1})^5 t^8 \right| = \lim_{t\to 0^+} \left| z_9 t^3 \right| = 0.$$

Similar arguments show that $\lim_{t\to 0^+} \tilde{c}_j t^{j-1} = 0$ holds for $j \in \{5, 6, 7, 8\}$. Using the fact that $0 < 2^{\lceil \log_2 t \rceil} \leq 2^{1+\log_2 t} = 2t$ for all $t > 0$, for $j = 3$, we have that

$$\lim_{t\to 0^+} \left| z_3 2^{\lceil \log_2 t \rceil} t^2 \right| \leq \lim_{t\to 0^+} \left| z_3 2 t^3 \right| = 0,$$

while for $j = 2$ and $j = 1$ we respectively have that

$$\lim_{t\to 0^+} \left| z_2 2^{2\lceil \log_2 t \rceil} - 1 \right| t \leq \lim_{t\to 0^+} \left| z_2 (2t)^2 - 1 \right| t = \lim_{t\to 0^+} \left| z_2 (4t^3 - t) \right| = 0.$$

and

$$\lim_{t\to 0^+} \left| z_1 2^{3\lceil \log_2 t \rceil} \right| \leq \lim_{t\to 0^+} \left| z_1 (2t)^3 \right| = 0.$$

Finally, for $j = 0$, we have that

$$\lim_{t\to 0^+} \frac{z_0 2^{4\lceil \log_2 t \rceil}}{t} \leq \lim_{t\to 0^+} \frac{z_0 (2t)^4}{t} = 0.$$

Hence, we have shown that f_1 is at least $\mathcal{C}^1$ on its domain.

Analogously, for f_1'' to exist and be continuous at $t = 0$,

$$f_1''(0) = \lim_{\varepsilon \to 0^+} \frac{f_1'(0+\varepsilon) - f_1'(0)}{\varepsilon} = \lim_{\varepsilon \to 0^+} \frac{f_1'(\varepsilon)}{\varepsilon} = -2 \tag{A.3}$$

must hold. We have that

$$f_1'(t) = \begin{cases} \sum_{j=1}^{9} j \tilde{c}_j t^{j-1} & \text{if } t > 0 \\ -2t & \text{if } t \in [-1, 0] \end{cases} \tag{A.4}$$

and so we consider $\lim_{t\to 0^+} j\tilde{c}_j t^{j-2}$ for $j \in \{1, \ldots, 9\}$, i.e., the limit of each term in the sum in (A.4) divided by t. We show that for all but $j = 2$, these values goes to zero, while the $j = 2$ value goes to -2 as $t \to 0^+$. For $j = 9$, we have that

$$\lim_{t\to 0^+} \left| 9 z_9 2^{-5\lceil \log_2 t \rceil} t^7 \right| \leq \lim_{t\to 0^+} \left| 9 z_9 (t^{-1})^5 t^7 \right| = 0,$$

with similar arguments showing that $j \in \{5,6,7,8\}$ values also diminish to zero. For $j = 4$, we simply have $\lim_{t\to 0^+} 4z_4 t^2 = 0$. For $j = 3$,

$$\lim_{t\to 0^+} \left|3z_3 2^{\lceil \log_2 t\rceil} t\right| \le \lim_{t\to 0^+} \left|3z_3(2t)t\right| = 0.$$

For $j = 2$, we have that

$$\lim_{t\to 0^+} 2(z_2 2^{2\lceil \log_2 t\rceil} - 1) = \lim_{t\to 0^+} 2z_2(2^{\lceil \log_2 t\rceil})^2 - 2 = -2.$$

Lastly, for $j = 1$, we have that

$$\lim_{t\to 0^+} \frac{\left|z_1 2^{3\lceil \log_2 t\rceil}\right|}{t} \le \lim_{t\to 0^+} \frac{\left|z_1(2t)^3\right|}{t} = 0,$$

and so we have now shown that f_2 is at least $\mathcal{C}^2$ on its domain.

Finally, for f_1''' to exist and be continuous at $t = 0$,

$$f_1'''(0) = \lim_{\varepsilon\to 0^+} \frac{f_1''(0+\varepsilon) - f_1''(0)}{\varepsilon} = \lim_{\varepsilon\to 0^+} \frac{f_1''(\varepsilon) + 2}{\varepsilon} = 0 \tag{A.5}$$

must hold. We have that

$$f_1''(t) = \begin{cases} \sum_{j=2}^{9} j(j-1)\tilde{c}_j t^{j-2} & \text{if } t > 0 \\ -2 & \text{if } t \in [-1,0] \end{cases} \tag{A.6}$$

and so we consider $\lim_{t\to 0^+} j(j-1)\tilde{c}_j t^{j-3}$ for $j \in \{2,\dots,9\}$, i.e., the limit of each term in the sum in (A.6) divided by t. For $j \in \{5,6,7,8,9\}$, we again have similar arguments showing that the corresponding values vanish, so we just show the $j = 9$ case, which follows because

$$\lim_{t\to 0^+} \left|72z_9 2^{-5\lceil \log_2 t\rceil} t^6\right| \le \lim_{t\to 0^+} \left|72z_9(t^{-1})^5 t^6\right| = 0.$$

Again, it is clear that the value for $j = 4$ vanishes. For $j = 3$, we have that

$$\lim_{t\to 0^+} \left|6z_3 2^{\lceil \log_2 t\rceil}\right| \le \lim_{t\to 0^+} \left|6z_3(2t)\right| = 0.$$

Finally, for $j = 2$, we can rewrite (A.5) as follows, making use of these aforementioned limits which vanish and replacing ε by t, to obtain a limit only involving the $j = 2$ term:

$$\begin{aligned} f_1'''(0) = \lim_{t\to 0^+} \frac{f_1''(t) + 2}{t} &= \lim_{t\to 0^+} \frac{2(z_2 2^{2\lceil \log_2 t\rceil} - 1) + 2}{t} \\ &= \lim_{t\to 0^+} \frac{2z_2(2^{\lceil \log_2 t\rceil})^2}{t} \le \lim_{t\to 0^+} \frac{2z_2(2t)^2}{t} = 0. \end{aligned}$$

Thus, f_1 is indeed $\mathcal{C}^3$ on its domain. □

Proof of Theorem 3.2. Since l_k is a power of two, we can rewrite the derivative of p_k, i.e., $p_k'(t) = \sum_{j=1}^{9} j c_j t^{j-1}$, as a function of $\zeta \in [1,2]$:

$$\begin{aligned} \tilde{p}_k'(\zeta) = \sum_{j=1}^{9} j c_j (l_{k+1}\zeta)^{j-1} = \sum_{j=1}^{9} \frac{j c_j}{2^{(k+1)(j-1)}} \zeta^{j-1} &= \frac{2(z_2 2^{-2k} - 1)}{2^{k+1}}\zeta + \sum_{\substack{j=1 \\ j\neq 2}}^{9} \frac{j z_j 2^{(j-4)k}}{2^{(k+1)(j-1)}} \zeta^{j-1} \\ &= \frac{z_2 - 2^{2k}}{2^{3k}}\zeta + \sum_{\substack{j=1 \\ j\neq 2}}^{9} \frac{j z_j 2^{1-j}}{2^{3k}} \zeta^{j-1} = \frac{1}{2^{3k}} \left((z_2 - 2^{2k})\zeta + \sum_{\substack{j=1 \\ j\neq 2}}^{9} \tilde{z}_j \zeta^{j-1} \right), \end{aligned}$$

where $\tilde{z}_j = j z_j 2^{1-j}$. From Lemma A.1, we see that $z_2 - 2^{2k} < 0$ for all $k \ge 10$, while for any k, we have that $\tilde{z}_j < 0$ for $j \in \{1,3,5,7,9\}$ and $\tilde{z}_j > 0$ for $j \in \{4,6,8\}$. Since $\zeta \in [1,2]$, an upper bound for $\tilde{p}_k'$ can be obtained by evaluating its negative terms at $\zeta = 1$ and its positive terms at $\zeta = 2$, i.e., for all $k \ge 10$ and any $\zeta \in [1,2]$, we have that

$$\tilde{p}_k'(\zeta) \le \frac{1}{2^{3k}} \left((z_2 - 2^{2k}) + \sum_{j\in\{1,3,5,7,9\}} \tilde{z}_j + \sum_{j\in\{4,6,8\}} \tilde{z}_j 2^{j-1} \right).$$

For $k \ge 13$, the upper bound on the derivative is negative. Thus, for $k \ge 13$, $\tilde{p}_k'(\zeta) < 0$ for any $\zeta \in [1,2]$, so p_k must be decreasing. Consequently, the $t = 0$ maximizer of f_1 is isolated. Finally, it immediately follows that the $t = 0$ maximizer of $f_{\max} = \max(f_1, f_2)$ is also isolated. □

B Why $s_k = 1 + 2^{-k}$ is insufficient to make (3.1) a $\mathcal{C}^3$ function

For $s_k = 1 + 2^{-k}$, symbolic computation shows that the coefficients of $p_k(t) = \sum_{j=0}^{9} c_j t^j$ are:

$$c_j = \begin{cases} z_j 2^{(j-3)k} - 1 & \text{if } j = 2 \\ z_j 2^{(j-3)k} & \text{otherwise} \end{cases}$$

where the integers z_j remain the same as given in Lemma A.1. To see if (A.5) still holds for this new choice of s_k we look at $\lim_{t\to 0^+} j(j-1)\tilde{c}_j t^{j-3}$ for $j \in \{2, \ldots, 9\}$. However, now none of the individual limits vanish. For example, for $j = 9$, we have that

$$\lim_{t\to 0^+} \left|72 z_9 2^{-6\lceil \log_2 t \rceil} t^6\right| \geq \lim_{t\to 0^+} \left|72 z_9 (2^{-1}t^{-1})^6 t^6\right| = \tfrac{9}{8}\left|z_9\right| \neq 0,$$

where we have used the fact that $0 < 2^{-1}t^{-1} = 2^{-1-\log_2 t} \leq 2^{-\lceil \log_2 t \rceil}$; similarly, the limits for $j \in \{4, 5, 6, 7, 8\}$ do not vanish either. For $j = 3$, we simply have that $\lim_{t\to 0^+} 6z_3 = 6z_3 \neq 0$. Finally, even if all of the terms considered above were to vanish and we substitute in the value for $j = 2$ into (A.5), we nevertheless would end up attaining another limit that does not vanish:

$$\lim_{t\to 0^+} \frac{2z_2 2^{\lceil \log_2 t \rceil}}{t} \geq \lim_{t\to 0^+} \frac{2z_2(t)}{t} = 2z_2 \neq 0.$$

The only remaining way that (A.5) could hold is if all of these non-vanishing terms cancel, but from our experiments (see Figure 2a), we know this is not the case.

References

[BB90] S. Boyd and V. Balakrishnan. A regularity result for the singular values of a transfer matrix and a quadratically convergent algorithm for computing its L_∞-norm. *Systems Control Lett.*, 15(1):1–7, 1990.

[BBK89] S. Boyd, V. Balakrishnan, and P. Kabamba. A bisection method for computing the $\mathcal{H}_\infty$ norm of a transfer matrix and related problems. *Math. Control Signals Systems*, 2:207–219, 1989.

[BM18] P. Benner and T. Mitchell. Faster and more accurate computation of the $\mathcal{H}_\infty$ norm via optimization. *SIAM J. Sci. Comput.*, 40(5):A3609–A3635, October 2018.

[BS90] N. A. Bruinsma and M. Steinbuch. A fast algorithm to compute the H_∞-norm of a transfer function matrix. *Systems Control Lett.*, 14(4):287–293, 1990.

[Bye88] R. Byers. A bisection method for measuring the distance of a stable matrix to unstable matrices. *SIAM J. Sci. Statist. Comput.*, 9:875–881, 1988.

[Gür12] M. Gürbüzbalaban. *Theory and methods for problems arising in robust stability, optimization and quantization.* PhD thesis, New York University, New York, NY 10003, USA, May 2012.

[GVDV98] Y. Genin, P. Van Dooren, and V. Vermaut. Convergence of the calculation of $\mathcal{H}_\infty$-norms and related questions. In A. Beghi, L. Finesso, and G. Picci, editors, *Mathematical Theory of Networks and Systems, 13 ed.*, Proceedings of the MTNS-98 Symposium, Padova, pages 629–632, July 1998.

[HJ91] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, 1991.

[Kat82] T. Kato. *A Short Introduction to Perturbation Theory for Linear Operators.* Springer-Verlag, New York-Berlin, 1982.

[KP02] S. G. Krantz and H. R. Parks. *A Primer of Real Analytic Functions.* Birkhäuser Advanced Texts. Birkhäuser, Boston, MA, second edition, 2002.

[Lax07] P. D. Lax. *Linear Algebra and its Applications.* Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2nd edition, 2007.

[MO05] E. Mengi and M. L. Overton. Algorithms for the computation of the pseudospectral radius and the numerical radius of a matrix. *IMA J. Numer. Anal.*, 25(4):648–669, 2005.

[MVD20] V. Mehrmann and P. M. Van Dooren. Optimal robustness of port-Hamiltonian systems. *SIAM J. Matrix Anal. Appl.*, 41(1):134–151, 2020.

[MVD21] T. Mitchell and P. Van Dooren. Root-max problems, hybrid expansion-contraction, and quadratically convergent optimization of passive systems, 2021. In preparation.

[vNW29] J. von Neumann and E. P. Wigner. Über merkwürdige diskrete Eigenwerte. *Physikalische Zeitschrift*, 40:467–470, 1929.

[ZDG96] K. Zhou, J. C. Doyle, and K. Glover. *Robust and Optimal Control.* Prentice-Hall, Upper Saddle River, NJ, 1996.