Root-max Problems, Hybrid Expansion-Contraction, and Quadratically Convergent Optimization of Passive Systems

Tim Mitchell*   Paul Van Dooren†

September 2, 2021

Abstract

We present quadratically convergent algorithms to compute the extremal value of a real parameter for which a given rational transfer function of a linear time-invariant system is passive. This problem is formulated for both continuous-time and discrete-time systems and is linked to the problem of finding a realization of a rational transfer function such that its passivity radius is maximized. Our new methods make use of the Hybrid Expansion-Contraction algorithm, which we extend and generalize to the setting of what we call root-max problems.

1 Introduction

Robustness measures play an important role in systems and control. They provide margins for the perturbations that one can allow on a given nominal dynamical system such that the perturbed system still performs as desired. A classical example of such a measure is the so-called *distance to instability* [Van85], which measures how much one can perturb a stable matrix before destabilization is a possibility. A generalization of this is the *complex stability radius* (better known by its reciprocal, the $\mathcal{H}_\infty$ norm), which measures how much (complex-valued) uncertainty in a dynamical system with input and output can be tolerated before stability is no longer guaranteed [ZDG96, HP05]. Meanwhile, the *real structured stability radius* and μ-value further restrict the uncertainty to be real-valued or structured in a particular sense [HP90a, HP90b]. Such measures are often the subject of optimization in robust control, since it is natural to desire that the robustness of models to uncertainty/perturbation be maximized. Furthermore, in the area of model order reduction, the $\mathcal{H}_\infty$ norm is one of the main indicators of how well a reduced-order surrogate mimics the behavior of a larger (and often computationally unwieldy) system [Glo84]. Numerical procedures for computing these robustness measures have been developed in the last few decades and have historically been focused on linear time-invariant systems described by their generalized state-space model.

In this paper, we consider a problem that is linked to maximizing the *passivity radius* [OVD05], which measures how much one can can perturb a passive system before

---

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

†Department of Mathematical Engineering, Université catholique de Louvain, Louvain-La-Neuve, Belgium. `paul.vandooren@uclouvain.be`, ORCID: 0000-0002-0115-9932. Visiting the Max Planck Institute of Magdeburg.

it may lose passivity. A linear time-invariant system is said to be *passive* if it is stable and the Hermitian part of its transfer function is positive definite on the imaginary axis. More specifically, we propose new algorithms to compute the extremal value of a real parameter for which a certain rational transfer function is passive, where the transfer function is defined by a parameterized linear time-invariant system in state-space form. The computation of this extremal parameter value is also important, as it allows one to construct certificates for the passivity of the parameterized passive systems. As discussed in [MVD20b, MVD20a], these certificates play a crucial role in the solution of two important problems: (i) finding a realization of a given passive system with optimal passivity radius and (ii) finding the closest passive system to a given non-passive system. The first algorithms to compute this extremal value were recently proposed in [MVD20b] and [MVD20a], respectively, for the continuous- and discrete-time cases. Here we improve upon these techniques by proposing better algorithms that have local quadratic convergence and are also much faster and more reliable in practice. In the process, we also establish that the earlier methods of [MVD20b, MVD20a] converge at least superlinearly. Finding a nearby passive system to a non-passive one is another problem that has already been considered in the literature (see [GS18] and [FGL20]), but it was suggested in [MVD20b, MVD20a] that techniques like those developed in this paper could be applied to address that problem as well.

A core part of our new methods (from which they derive their quadratic convergence) is an iteration called Hybrid Expansion-Contraction (HEC). HEC was first conceived as a way to approximate the $\mathcal{H}_\infty$ norm of large-scale systems [Mit14, MO16] and was subsequently extended to approximating the real structured stability radius [GGMO17]. However, HEC and its convergence properties have only been described for these two specific settings, while the structure of our problem of interest here is quite different. Unlike the $\mathcal{H}_\infty$ norm, which is computed by obtaining a global maximizer of a function in one real variable, the extremal value we consider here for the optimization of passive systems is computed by iterating over two real variables. Consequently, another contribution of this paper is to connect these seemingly disparate things, namely, by (i) identifying that all of these problems are actually specific instances of what we call *root-max problems* and (ii) generalizing HEC and its convergence results to this new class. Besides enabling our new methods here, we hope that this much more accessible and generalized presentation of HEC will both increase awareness for identifying root-max problems and ease facilitation of new HEC-based methods.

The paper is organized as follows. We first establish notation and preliminary material in §2. Then, in §3, we introduce root-max problems and generalize HEC and its convergence results to this problem class. In §4, we describe the continuous-time version of our passivity radius problem and our corresponding new algorithm to solve it, while the discrete-time case is handled in §5. Numerical experiments and concluding remarks are, respectively, given in §6 and §7.

2 Preliminaries

We begin with notation. The set of Hermitian matrices in $\mathbb{C}^{n \times n}$ is denoted by $\mathbb{H}_n$, with $A \succ 0$ ($A \succeq 0$) additionally signifying that $A \in \mathbb{H}_n$ is positive (semi-)definite. $\Lambda(A)$ denotes the spectrum of a matrix A and, when A is Hermitian, we additionally use the shorthand $\lambda_{\min}(A)$ to denote its smallest eigenvalue. $\mathrm{Re}\,(Z)$ and $\mathrm{Im}\,(Z)$, respectively, denote the real and imaginary parts of a complex matrix Z, while the (conjugate) transpose of a vector or matrix V is denoted by V^{T} (V^{H}). We use I_n for the $n \times n$ identity matrix.

The models that we consider here are given by their standard state-space form, which means that their associated transfer functions are proper. In the continuous-time setting,

the transfer function arises from the *Laplace transform* of the system

$$\begin{aligned}
\dot{x}(t) &= Ax(t) + Bu(t),\ x(0) = 0,\\
y(t) &= Cx(t) + Du(t),
\end{aligned} \tag{2.1}$$

where $A \in \mathbb{C}^{n \times n}$, $B \in \mathbb{C}^{n \times m}$, $C \in \mathbb{C}^{p \times n}$, $D \in \mathbb{C}^{p \times m}$, and $x(t)$, $u(t)$, $y(t)$ are vector-valued functions denoting, respectively, the *state*, *input*, and *output* of the system. In the discrete-time setting, the transfer function arises from the *z-transform* applied to

$$\begin{aligned}
x_{k+1} &= Ax_k + Bu_k,\ x_0 = 0,\\
y_k &= Cx_k + Du_k,
\end{aligned} \tag{2.2}$$

where now x_k, u_k, and y_k are vector-valued sequences denoting, respectively, the *state*, *input*, and *output* of the system. In both cases, we denote these systems by four-tuples of matrices $\mathcal{M} \coloneqq \{A, B, C, D\}$ and their associated rational matrices

$$\mathcal{T}(\lambda) \coloneqq C(\lambda I_n - A)^{-1}B + D \quad \text{and} \quad \mathcal{T}^{\mathsf{H}}(\lambda) \coloneqq B^{\mathsf{H}}(\lambda I_n - A^{\mathsf{H}})^{-1}C^{\mathsf{H}} + D^{\mathsf{H}} \tag{2.3}$$

are, respectively, the associated *transfer function* and *para-conjugate transfer function*, where the variable λ stands for the Laplace variable s in the continuous-time setting and the delay operator z in the discrete-time case.

We restrict ourselves to systems which are *minimal*, i.e., the pair (A, B) is *controllable* (for all $\lambda \in \mathbb{C}$, $\operatorname{rank} \begin{bmatrix} \lambda I_n - A & B \end{bmatrix} = n$), and the pair (A, C) is *observable* (i.e., $(A^{\mathsf{H}}, C^{\mathsf{H}})$ is controllable). If the model is not minimal, one can always construct a minimal realization by removing the uncontrollable and unobservable parts, which can be done in a backward stable manner [VD81].

For rates of convergence, we use the notion of Q-quadratic and Q-superlinear convergence, where "Q" stands for "quotient"; see [NW99, p. 619] for more details.

3   Root-max problems and Hybrid Expansion-Contraction

Let $\mathcal{D}_1 \subseteq \mathbb{R}$ be connected, $\mathcal{D}_2 \subset \mathbb{R}^N$ be compact, and $g : \mathcal{D}_1 \times \mathcal{D}_2 \to \mathbb{R}$ be a continuous function. Consider the root-finding problem, which we call a *root-max problem*:

$$f(\varepsilon) \coloneqq \max_{x \in \mathcal{D}_2} g(\varepsilon, x) = 0, \tag{3.1}$$

where $\varepsilon \in \mathcal{D}_1$, and for any fixed $\hat{\varepsilon} \in \mathcal{D}_1$, $g(\hat{\varepsilon}, x)$ is bounded above. If there exists $\varepsilon_{\mathrm{lb}}, \varepsilon_0 \in \mathcal{D}_1$ with $f(\varepsilon_{\mathrm{lb}}) < 0 \leq f(\varepsilon_0)$, then by continuity of g it is clear that (3.1) has at least one root $\varepsilon_\star \in (\varepsilon_{\mathrm{lb}}, \varepsilon_0] \subseteq \mathcal{D}_1$ such that $f(\varepsilon_\star) = 0$. Of course, if $f(\varepsilon_0) = 0$ holds, then we can take $\varepsilon_\star = \varepsilon_0$. Without loss of generality, we assume that $\varepsilon_{\mathrm{lb}} < \varepsilon_0$. While using $f(\varepsilon_{\mathrm{lb}}) < 0$ and $0 \leq f(\varepsilon_0)$ is a convenient convention here, note that each of these can be modified to be (non-)strict as desired.

Remark 3.1. *In addition to* (3.1)*, one can also equivalently consider a* root-min problem*, in which case $g(\hat{\varepsilon}, x)$ must then instead be bounded below for any fixed $\hat{\varepsilon} \in \mathcal{D}_1$. For now, we restrict to the root-max form with $\varepsilon_{\mathrm{lb}} < \varepsilon_0$, since HEC was originally developed using this convention.*

Many well-known distance measures can be written in the form of the root-max problem given by (3.1) (or as a root-min problem). For example, the distance to instability, the $\mathcal{H}_\infty$ norm, and the real stability radius all fall in this problem class. As we explain later, our particular problem of interest, the optimization of the passivity radius of passive systems, is also in this class.

A key component of the our new passivity-optimization methods here is the Hybrid Expansion-Contraction (HEC) algorithm of [Mit14, MO16]. In [MO16, p. 997], HEC is described as a quadratically convergent "adaptively positively or negatively damped Newton

method" for approximating the $\mathcal{H}_\infty$ norm of large-scale systems, where "positively or negatively damped" means that HEC sometimes takes steps larger or smaller than the regular Newton step, which is depicted in [MO16, Fig. 4]. In this large-scale setting, the strategy for approximating the $\mathcal{H}_\infty$ norm is to compute the unique root of a particular monotonically increasing function in one real variable; the reciprocal of this root is the $\mathcal{H}_\infty$ norm. As evaluating this function is also expensive, a scalable subroutine [GGO13] was designed to instead efficiently a compute lower bound to it, which in practice, often coincides with the true function value, and then a Newton-bisection-based outer iteration was proposed by the authors to compute the root. However, as noted by Mitchell in [GGO13, Acknowledgements] and described more fully in [MO16, Section 3.2], this root-finding-based algorithm of [GGO13] can break down and may converge to arbitrarily bad approximations to the $\mathcal{H}_\infty$ norm which are not even locally optimal. Moreover, when this breakdown happens, the convergence rate also degrades to linear. These problems arise precisely because the function whose root is sought is not guaranteed to be computed accurately, and so using a standard root-finding method as the outer iteration is fraught with danger. The HEC algorithm overcomes these deficiencies by employing one-sided convergence enabled by a key observation: when only lower bounds to the function value are guaranteed, if the computed estimate is negative, the direction of the unique root cannot be determined, but if the computed estimate is positive, one does know that the root lies to the left. Under mild assumptions, HEC is guaranteed to converge to locally optimal approximations to the $\mathcal{H}_\infty$ norm by computing a decreasing sequence of upper bounds to the root, and in practice [MO16, Section 8], HEC often converges to the true value of the $\mathcal{H}_\infty$ norm.

In the general context of root-max problems, HEC often converges to roots of (3.1), but it does not actually guarantee this. Instead, HEC generically converges to what we call a *pseudoroot* of (3.1), which will be defined momentarily; in [Mit14, MO16], no name was given for this concept as it was not needed in that context.

3.1 The HEC algorithm and its convergence properties

We now generalize the convergence properties of HEC to our root-max problem given by (3.1). At a very high level, the proofs we give below for the convergence guarantees of HEC often follow the arguments given by Mitchell and Overton for specifically for approximating the $\mathcal{H}_\infty$ norm [MO16, Section 4]. However, our new generalization here makes these convergence results significantly easier to understand and allows them to be applied to other settings. We define $g_x : \mathcal{D}_1 \to \mathbb{R}$ and $g_\varepsilon : \mathcal{D}_2 \to \mathbb{R}$ as follows:

$$g_x(\varepsilon) := g(\varepsilon, x), \qquad \text{where } x \in \mathcal{D}_2 \text{ is fixed}, \tag{3.2a}$$

$$g_\varepsilon(x) := g(\varepsilon, x), \qquad \text{where } \varepsilon \in \mathcal{D}_1 \text{ is fixed}. \tag{3.2b}$$

Remark 3.2. *To avoid this section from becoming significantly more technical, we assume that for all $\varepsilon \in \mathcal{D}_1$ that function g_ε is differentiable. However, the guaranteed convergence of HEC to pseudoroots (Theorem 3.6) does not critically rely on differentiability of $g_\varepsilon(x)$, and this result can be adapted to the case of $g_\varepsilon(x)$ being nonsmooth at stationary points. That said, the quadratic convergence of HEC (Theorem 3.8) does require additional smoothness assumptions.*

Remark 3.3. *The HEC algorithm requires a root-finding method and optimization solver. For simplicity, we assume that such methods are deterministic and converge exactly. Deterministic means that the methods return the same answer for the same initial data. Exact convergence for optimization means that a solver converges to stationary points, i.e., where the norm of the gradient is zero. Finally, HEC requires, in theory and practice, an optimization solver which is monotonic, i.e., it always increases the value of the objective function being maximized at successive iterates. As most unconstrained solvers do this by design, this is a mild requirement.*

4

Algorithm 1 Hybrid Expansion-Contraction (HEC)

Input: $\varepsilon_{\text{lb}}, \varepsilon_0 \in \mathcal{D}_1$ and $x_0 \in \mathcal{D}_2$ such that $f(\varepsilon_{\text{lb}}) < 0 \leq g(\varepsilon_0, x_0) \leq f(\varepsilon_0)$
Output: $(\tilde{\varepsilon}, \tilde{x})$ such that $g(\tilde{\varepsilon}, \tilde{x}) = 0$ and $\tilde{x}$ is a stationary point of $g_{\tilde{\varepsilon}}(x)$

1: **for** $k = 0, 1, 2, \ldots$ **do**
2:     *// Contraction: deterministic root-finding method initialized at ε_k*
3:     $\hat{\varepsilon}_k \leftarrow$ a root of $g_{x_k}(\varepsilon)$ with $\hat{\varepsilon}_k \in (\varepsilon_{\text{lb}}, \varepsilon_k]$
4:     **if** x_k is a stationary point of $g_{\hat{\varepsilon}_k}(x)$ **then**
5:         $(\tilde{\varepsilon}, \tilde{x}) \leftarrow (\hat{\varepsilon}_k, x_k)$
6:         **return**
7:     **end if**
8:     *// Expansion: deterministic optimization method initialized at x_k*
9:     $x_{k+1} \leftarrow$ a stationary point of $g_{\hat{\varepsilon}_k}(x)$ with $g_{\hat{\varepsilon}_k}(x_{k+1}) > g_{\hat{\varepsilon}_k}(x_k)$
10:    $\varepsilon_{k+1} \leftarrow \hat{\varepsilon}_k$
11: **end for**

NOTE: *If the conditional statement in line 4 is never satisfied, then by Theorem 3.6, HEC produces two infinite sequences $\{\varepsilon_k\}$ and $\{x_k\}$, with the former converging to $\tilde{\varepsilon}$ and the latter having at least one cluster point, any of which we denote as $\tilde{x}$. Contraction must use a root-finding method with bracketing, e.g., Newton-bisection, to ensure a root in the given bracket $(\varepsilon_{\text{lb}}, \varepsilon_k]$ is found. The inequality in the expansion phase is guaranteed by simply initializing optimization at x_k and using a monotonic optimization solver. Finally, HEC can begin with either an expansion or contraction phase, and which is more convenient may depend on the particular application.*

Definition 3.4. *Given $\tilde{\varepsilon} \in \mathcal{D}_1$ and $\tilde{x} \in \mathcal{D}_2$, $(\tilde{\varepsilon}, \tilde{x})$ is a* pseudoroot *of (3.1) if $g(\tilde{\varepsilon}, \tilde{x}) = 0$ and $\tilde{x}$ is a stationary point of $g_{\tilde{\varepsilon}}(x)$.*

Defining pseudoroot in terms of a stationary point of $g_{\tilde{\varepsilon}}(x)$, as opposed to a local maximizer, which might seem more intuitive, is intentional. The reason for this is subtle and requires more context to explain, so we defer this discussion to Remark 3.7. As we see in the following simple result (whose proof we omit as it is elementary), pseudoroots are intimately related with roots of (3.1).

Lemma 3.5. *Let $\tilde{\varepsilon} \in \mathcal{D}_1$, $\tilde{x} \in \mathcal{D}_2$, and $(\tilde{\varepsilon}, \tilde{x})$ be a pseudoroot of (3.1). Then $\tilde{\varepsilon}$ is a root of (3.1) if and only if $\tilde{x}$ is a global maximizer of $g_{\tilde{\varepsilon}}(x)$. Otherwise, $0 < f(\tilde{\varepsilon})$.*

We now give an overview of HEC. By construction, HEC generates a monotonically decreasing sequence $\{\varepsilon_k\} \to \tilde{\varepsilon}$. For $x_k \in \mathcal{D}_2$ fixed with $g_{x_k}(\varepsilon_k) \geq 0$, first note that we have

$$g_{x_k}(\varepsilon_{\text{lb}}) \leq f(\varepsilon_{\text{lb}}) < 0 \leq g_{x_k}(\varepsilon_k) \leq f(\varepsilon_k).$$

The *one-parameter contraction phase* reduces ε_k by finding a root $\hat{\varepsilon}_k \in (\varepsilon_{\text{lb}}, \varepsilon_k]$ of $g_{x_k}(\varepsilon)$. By the inequalities above, there must be at least one root in this bracket. If $g_{x_k}(\varepsilon_k) = 0$, the contraction phase simply returns $\hat{\varepsilon}_k = \varepsilon_k$. Otherwise, bisection can be used to find a root in $(\varepsilon_{\text{lb}}, \varepsilon_k)$. But if $g_{x_k}(\varepsilon)$ is sufficiently smooth at $\hat{\varepsilon}_k$, then, e.g., Newton's or Halley's method could find it with far fewer iterations. Of course, these faster root-finding methods are not guaranteed to converge and $g_{x_k}(\varepsilon)$ may not be always be sufficiently smooth, which is why it is instead critical to combine both approaches, e.g., Newton-bisection. By employing bracketing and bisection, convergence to a root of $g_{x_k}(\varepsilon)$ is ensured, but the ability to also take Newton (or Halley) steps, assuming that they fall inside the current bracket, can yield a quadratic (or cubic) rate of convergence when sufficient smoothness holds. Subsequently, for $\hat{\varepsilon}_k \in \mathcal{D}_1$ now fixed and $g_{\hat{\varepsilon}_k}(x_k) = 0$, the *multi-parameter expansion phase* attempts to maximize $g_{\hat{\varepsilon}_k}(x)$ by initializing an optimization solver at x_k. If optimization returns $x_{k+1} = x_k$, there is nothing to do, e.g., when x_k is a stationary point of $g_{\hat{\varepsilon}_k}(x)$. Otherwise, by employing a monotonic solver, optimization will converge to a stationary point x_{k+1}
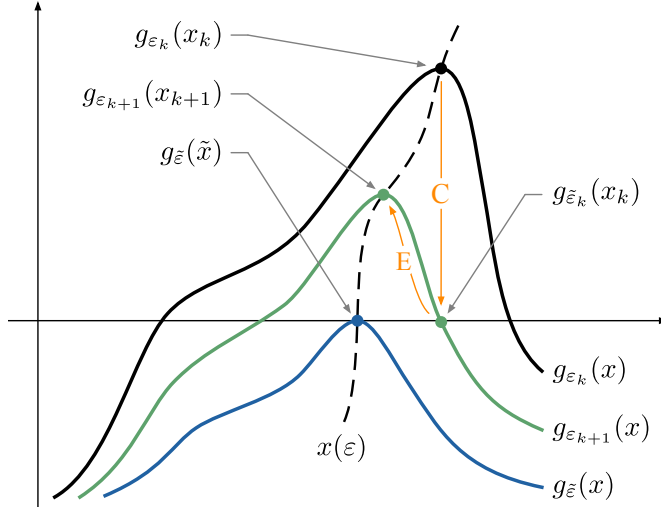
Figure 1: Illustration of HEC of converging to a pseudoroot $(\tilde{\varepsilon}, \tilde{x})$ of (3.1), where $\tilde{\varepsilon}$ is also a root of (3.1). At iteration k, HEC has found the global maximizer x_k of $g_{\varepsilon_k}(x)$. The contraction phase, denoted by 'C' above, computes $\varepsilon_{k+1} = \hat{\varepsilon}_k \leq \varepsilon_k$, which is a root of $g_{x_k}(\varepsilon)$. The subsequent expansion phase, denoted by 'E' above, then computes the global maximizer x_{k+1} of $g_{\varepsilon_{k+1}}(x)$. The dashed curve $x(\varepsilon)$ denotes a continuous path of maximizers of $g_{\varepsilon}(x)$ as ε is varied, where $x(\tilde{\varepsilon})$ corresponds to the pseudoroot $(\tilde{\varepsilon}, \tilde{x})$.

of $g_{\hat{\varepsilon}_k}(x)$ (typically a maximizer) such that $g_{\hat{\varepsilon}_k}(x_{k+1}) > 0$. Beyond monotonicity of the solver, HEC does not specify a specific optimization method, though fast methods should be used when possible. This process of alternating between root finding (contraction) and optimization (expansion) is repeated in a loop and it converges to a pseudoroot of (3.1).

In Algorithm 1, we provide pseudocode for HEC for computing pseudoroots of the root-max problem given by (3.1). Fig. 1 depicts HEC converging to a pseudoroot $(\tilde{\varepsilon}, \tilde{x})$, where $\tilde{\varepsilon}$ is indeed a root of $f(\varepsilon)$ and $\tilde{x}$ is a maximizer of $g_{\tilde{\varepsilon}}(x)$. We now formally state the convergence results of HEC.

Theorem 3.6 (Convergence of HEC). *Given valid initial data, Algorithm 1 generates the sequences $\{\varepsilon_k\}$ converging monotonically to a limit $\tilde{\varepsilon}$ and $\{x_k\}$ with at least one cluster point, where $(\tilde{\varepsilon}, \tilde{x})$ is a pseudoroot of (3.1).*

Proof. We assume that conditional statement in line 4 of Algorithm 1 is never met, as otherwise the theorem clearly holds. Since the algorithm ensures that $\{\varepsilon_k\}$ is a monotonically decreasing sequence which is bounded below by ε_{lb}, it must converge to a limit $\tilde{\varepsilon}$, and so it follows that $\hat{\varepsilon}_k = \varepsilon_{k+1}$ converges to the same limit. By construction, for all $k \geq 1$, the algorithm also ensures that $g_{\varepsilon_k}(x_k) > 0$ with x_k being a stationary point of $g_{\varepsilon_k}(x)$. Now suppose that $g(\varepsilon_k, x_k)$ does not converge to zero. Then there is a subsequence $\{x_{k_i}\}$ for which $g(\varepsilon_{k_i}, x_{k_i})$ is bounded below by some $\gamma > 0$. Thus, by taking a further subsequence if necessary, we may assume without loss of generality that x_{k_i} converges to a limit $\tilde{x}$. By continuity of g, it follows that $g(\varepsilon_{k_i}, x_{k_i})$ converges to $g(\tilde{\varepsilon}, \tilde{x}) \geq \gamma$. However, since $\hat{\varepsilon}_{k_i}$ also converges to $\hat{\varepsilon}$, then $g(\hat{\varepsilon}_{k_i}, x_{k_i})$ must converge to the same limit $g(\tilde{\varepsilon}, \tilde{x})$, which is a contradiction, since by definition of the contraction step, $g_{x_{k_i}}(\hat{\varepsilon}_{k_i}) = 0$ must hold for all i. Thus $g(\varepsilon_k, x_k)$ must converge to zero. Although the sequence $\{x_k\}$ may not converge, it is bounded since $\mathcal{D}_2$ is a compact subset of $\mathbb{R}^N$, and so $\{x_k\}$ must have at least one cluster point. As $\|\nabla g_{\varepsilon_k}(x_k)\| = 0$ holds for all $k \geq 1$, clearly $\|\nabla g_{\tilde{\varepsilon}}(\tilde{x})\| = 0$ also holds, and so $\tilde{x}$ is also a stationary point of $g_{\tilde{\varepsilon}}(x)$, hence $(\tilde{\varepsilon}, \tilde{x})$ is a pseudoroot of (3.1). $\square$
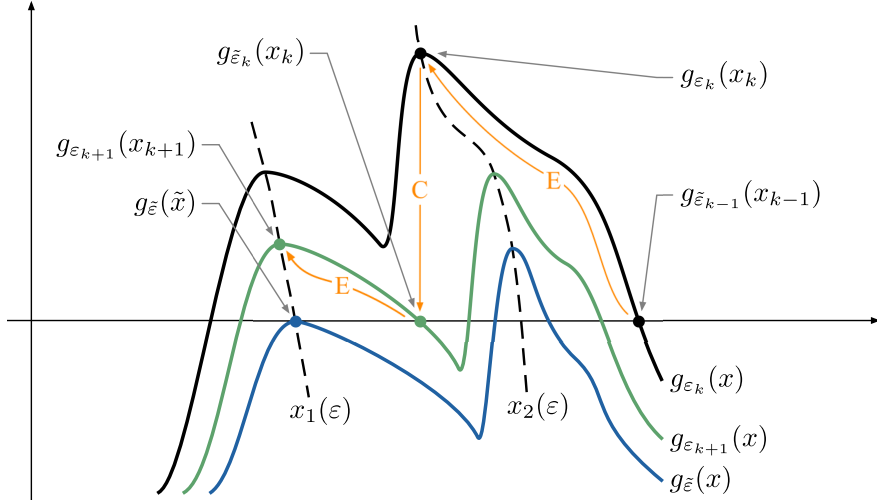
Figure 2: Illustration of HEC of encountering two different paths of stationary points, $x_1(\varepsilon)$ and $x_2(\varepsilon)$, with HEC eventually converging to a pseudoroot $(\tilde{\varepsilon}, \tilde{x})$ of (3.1) on path $x_1(\varepsilon)$, but $\tilde{\varepsilon}$ is not a root of (3.1). Note that paths of stationary points do not necessarily need to contain a pseudoroot (intersect with the x-axis), and although HEC may encounter and/or oscillate between multiple such paths as the algorithm converges, this does not affect the convergence result for HEC described by Theorem 3.6. For more details, see the caption of Fig. 1.

Remark 3.7. *Stationary points of $g_{\hat{\varepsilon}_k}(x)$ computed in the expansion phases will typically be maximizers, and some optimization solvers can guarantee convergence to maximizers (under appropriate assumptions). However, while Theorem 3.6 guarantees that HEC converges to a pseudoroot $(\tilde{\varepsilon}, \tilde{x})$ of (3.1), it does not guarantee that $\tilde{x}$ is a local maximizer of $g_{\tilde{\varepsilon}}(x)$, just that it is a stationary point. Nevertheless, whenever the expansion phases consistently return local maximizers, we do observe in practice that $\tilde{x}$ is also a local maximizer; see [MO16, GGMO17]. While it seems unlikely that $\tilde{x}$ would only be stationary, we do not believe it is impossible; e.g., it is easy to imagine that the functions $g_{\varepsilon_k}(x)$ shown in Fig. 1 could instead converge to a function $g_{\tilde{\varepsilon}}(x)$ that is constant in an interval about $\tilde{x}$.*

Although Fig. 1 depicts HEC only encountering a single continuous path $x(\varepsilon)$ of global maximizers of $g_\varepsilon(x)$, note that, per Theorem 3.6, HEC may actually encounter multiple such paths of stationary points as it progresses, and that these paths can consist of global or local maximizers (or sometimes even both). For example, Fig. 2 shows a depiction where $x_1(\varepsilon)$ and $x_2(\varepsilon)$ are two separate continuous paths of local maximizers of $g_\varepsilon(x)$ and HEC encounters both paths. Again, encountering multiple such paths does not affect the convergence result of Theorem 3.6. However, to show that the sequence $\{\varepsilon_k\}$ generated by HEC converges quadratically to $\tilde{\varepsilon}$, it will be simpler to assume that HEC eventually only encounters a single continuous path of local maximizers, like as is shown in Fig. 1.

Theorem 3.8 (Quadratic convergence of HEC). *Suppose that Algorithm 1 converges as described in Theorem 3.6 but that the sequence $\{x_k\}$ only has a single cluster point $\tilde{x}$ and $\tilde{x}$ lies on an open continuous path $x(\varepsilon)$ of stationary points of $g_\varepsilon(x)$ with $\tilde{x} = x(\tilde{\varepsilon})$. If $g(\varepsilon, x)$ and $x(\varepsilon)$ are both twice continuously differentiable, respectively, at $(\tilde{\varepsilon}, \tilde{x})$ and $\tilde{\varepsilon}$, $\tilde{x}$ is a local maximizer of $g_{\tilde{\varepsilon}}(x)$, and $\frac{\partial}{\partial \varepsilon} g(\varepsilon, \tilde{x}) \neq 0$, then the sequence $\{\varepsilon_k\}$ converges to $\tilde{\varepsilon}$ Q-quadratically.*

Proof. We begin by defining function

$$h(\varepsilon) := g(\varepsilon, x(\varepsilon)), \tag{3.3}$$

noting that since $(\tilde\varepsilon, \tilde x)$ being a pseudoroot, $h(\tilde\varepsilon) = g(\tilde\varepsilon, \tilde x) = 0$. Since the sequence $\{x_k\}$ only has one cluster point, there also exists some K such that for all $k \geq K$, all of the following properties hold:

(i) point x_k lies on $x(\varepsilon)$ with $x_k = x(\varepsilon_k)$, and so $h(\varepsilon_k) = g_{x_k}(\varepsilon_k)$,

(ii) $h'(\varepsilon_k) = g'_{x_k}(\varepsilon_k) \neq 0$,

(iii) $h(\varepsilon)$ and $g_{x_k}(\varepsilon)$ are twice continuously differentiable at ε_k.

By our assumptions, all of these statements also hold at $\tilde\varepsilon$. The agreement of the first derivatives in (ii) follows from the envelope theorem (or more generally, [RW98, Theorem 10.31]), since maximizers of $g_\varepsilon(x)$ do not occur at ∞ as $\mathcal{D}_2$ is compact.

Having established the needed properties above, we now consider the corresponding Newton steps for $h(\varepsilon)$ and $g_{x_k}(\varepsilon)$ at ε_k, which also must coincide, i.e.,

$$\varepsilon_k^{\mathrm{N}} := \varepsilon_k - \frac{h(\varepsilon_k)}{h'(\varepsilon_k)} = \varepsilon_k - \frac{g_{x_k}(\varepsilon_k)}{g'_{x_k}(\varepsilon_k)}. \tag{3.4}$$

However, Algorithm 1 sets $\varepsilon_{k+1} := \hat\varepsilon_k$, where $g_{x_k}(\hat\varepsilon_k) = 0$. Separately applying Taylor's theorem to $h(\varepsilon)$ and $g_{x_k}(\varepsilon)$, we have

$$0 = h(\tilde\varepsilon) = h(\varepsilon_k) + h'(\varepsilon_k)(\tilde\varepsilon - \varepsilon_k) + \tfrac{1}{2}h''(\xi_k)(\tilde\varepsilon - \varepsilon_k)^2$$

for some $\xi_k \in [\tilde\varepsilon, \varepsilon_k]$ and

$$0 = g_{x_k}(\hat\varepsilon_k) = g_{x_k}(\varepsilon_k) + g'_{x_k}(\varepsilon_k)(\hat\varepsilon_k - \varepsilon_k) + \tfrac{1}{2}g''_{x_k}(\eta_k)(\hat\varepsilon_k - \varepsilon_k)^2$$

for some $\eta_k \in [\tilde\varepsilon, \varepsilon_k]$. Dividing these equations by the derivative factors, subtracting the first from the second, and using (3.4) along with $\varepsilon_{k+1} = \hat\varepsilon_k$, we obtain

$$\varepsilon_{k+1} - \tilde\varepsilon = c_k(\tilde\varepsilon - \varepsilon_k)^2 + d_k(\varepsilon_{k+1} - \varepsilon_k)^2 \tag{3.5}$$

where

$$c_k = \frac{h''(\xi_k)}{2h'(\varepsilon_k)} \qquad \text{and} \qquad d_k = -\frac{g''_{x_k}(\eta_k)}{2g'_{x_k}(\varepsilon_k)}.$$

To establish quadratic convergence, we need to bound $\varepsilon_{k+1} - \varepsilon_k$ in terms of $\tilde\varepsilon - \varepsilon_k$. To this do, consider the Taylor expansions of $h(\varepsilon)$ and $g_{x_k}(\varepsilon)$ but with only the first two terms, i.e.,

$$0 = h(\tilde\varepsilon) = h(\varepsilon_k) + h'(\zeta_k)(\tilde\varepsilon - \varepsilon_k)$$

for some $\zeta_k \in [\tilde\varepsilon, \varepsilon_k]$ and

$$0 = g_{x_k}(\hat\varepsilon_k) = g_{x_k}(\varepsilon_k) + g'_{x_k}(\tau_k)(\hat\varepsilon_k - \varepsilon_k)$$

for some $\tau_k \in [\tilde\varepsilon, \varepsilon_k]$. As $h(\varepsilon_k) = g_{x_k}(\varepsilon_k)$ and $\varepsilon_{k+1} = \hat\varepsilon_k$, it follows that

$$\frac{\varepsilon_{k+1} - \varepsilon_k}{\tilde\varepsilon - \varepsilon_k} = \frac{h'(\zeta_k)}{g'_{x_k}(\tau_k)}, \tag{3.6}$$

which converges to 1 as $k \to \infty$, since $h'(\zeta_k)$ and $g'_{x_k}(\tau_k)$ both[1] converge to $h'(\tilde\varepsilon) \neq 0$. Dividing (3.5) by $(\varepsilon_k - \tilde\varepsilon)^2$ and taking the absolute value yields

$$\frac{|\varepsilon_{k+1} - \tilde\varepsilon|}{(\varepsilon_k - \tilde\varepsilon)^2} = \left| c_k + d_k \left(\frac{\varepsilon_{k+1} - \varepsilon_k}{\varepsilon_k - \tilde\varepsilon} \right)^2 \right|.$$

By (3.6), the squared term on the right converges to 1 as $k \to \infty$, while c_k and d_k also converge since their numerators are bounded and their denominators each converge to $h'(\tilde\varepsilon) \neq 0$. Thus, Algorithm 1 converges Q-quadratically. $\qquad\square$

---

[1] Note that in [MO16, p. 1000], there is a typo: in the second to last line of the proof of Theorem 4.4, $g'_{u_k v_k}(\varepsilon_k)$ actually should be $g'_{u_k v_k}(\tau_k)$.

Remark 3.9. *Theorem 3.8 is independent of how fast the contraction and expansion sub-problems are solved, respectively, by some root-finding method and optimization solver. The critical thing for quadratic convergence of HEC is that the expansion phase returns local maximizers of $g_{x_k}(\varepsilon)$, so that the derivatives of $h(\varepsilon)$ and $g_{x_k}(\varepsilon)$ coincide at ε_k. Per [MO16, Section 4.3], if the expansion phases are only solved inexactly, but this inexactness goes to zero in the limit, then HEC still converges at least Q-superlinearly; for more details, see [Mit14, Sections 3.1 and 3.2] and [DES82]. This can be practical, since if the expansion phases are particularly expensive to solve and require many iterations of optimization, it can be significantly faster to run HEC using this early contraction strategy. Trading off quadratic for superlinear convergence generally only results in a slight increase the total number of HEC iterations, which is typically more than compensated by much cheaper expansion phases; see [MO16, Section 8]*

Remark 3.10. *If HEC converges to a finite number of cluster points of $\{x_k\}$, rather than a unique one as supposed in Theorem 3.8, then it is easy to see that if the other conditions of Theorem 3.8 hold for any subsequence $\{x_{k_i}\}$ converging to a particular cluster point, then $\{\varepsilon_{k_i}\}$ must converge at least quadratically. Thus, if these other conditions also hold for any subsequence to any of the finitely many cluster points, then we expect that the rate of convergence of $\{\varepsilon_k\}$ should still be quadratic.*

4 Continuous-time passive systems

Returning to the optimization of passive systems, we first consider the continuous-time case where the finite-dimensional state-space model $\mathcal{M} \coloneqq \{A, B, C, D\}$ is given by (2.1) and its corresponding transfer function $\mathcal{T}(s)$ (2.3) is proper and minimal. Furthermore, for the remainder of the paper, we consider passive transfer functions $\mathcal{T}(s)$ (so $m = p$). We begin with the theoretical background defining the passivity optimization problem we wish to solve, which as we will show, is equivalent to a root-min problem.

4.1 Passivity of continuous-time proper parameterized systems

The material here in this subsection is mostly drawn from [MVD20b] but is recalled here in a concise way so that we can easily refer to it. We also briefly recall definitions and properties following [Wil72] and refer to the literature for proofs and more details.

Given $\mathcal{T}(s)$, consider the following rational function of $s \in \mathbb{C}$:

$$\Phi(s) \coloneqq \mathcal{T}^{\mathsf{H}}(-s) + \mathcal{T}(s),$$

which coincides with twice the Hermitian part of $\mathcal{T}(s)$ on the imaginary axis:

$$\Phi(\mathbf{i}\omega) = [\mathcal{T}(\mathbf{i}\omega)]^{\mathsf{H}} + \mathcal{T}(\mathbf{i}\omega).$$

Definition 4.1. *The transfer function $\mathcal{T}(s)$ is*

1. *passive if $\Phi(\mathbf{i}\omega) \succeq 0$ for all $\omega \in \mathbb{R} \cup \{\infty\}$ and A is stable, i.e., its eigenvalues are in the closed left half-plane, with any occurring on the imaginary axis (infinity included) being semi-simple,*

2. *strictly passive if $\Phi(\mathbf{i}\omega) \succ 0$ for all $\omega \in \mathbb{R} \cup \{\infty\}$ and A is asymptotically stable, i.e., its eigenvalues are in the open left half-plane.*

Using the matrix

$$W_{\mathrm{c}}(X, \mathcal{M}) \coloneqq \begin{bmatrix} -A^{\mathsf{H}}X - XA & C^{\mathsf{H}} - XB \\ C - B^{\mathsf{H}}X & D^{\mathsf{H}} + D \end{bmatrix}, \tag{4.1}$$

we have the following necessary and sufficient conditions for passivity of a finite-dimensional continuous-time system in state-space form; see [Wil72].

9

Theorem 4.2. *Let $\mathcal{M} := \{A, B, C, D\}$ with transfer function $\mathcal{T}(s)$ be minimal and proper. Then $\mathcal{T}(s)$ is (strictly) passive if and only if there exists an $X \in \mathbb{H}_n$ such that $X \succ 0$ and $W_c(X, \mathcal{M}) \succeq 0$ ($W_c(X, \mathcal{M}) \succ 0$).*

In [MVD20b], the following class of systems, parameterized by $\xi \in \mathbb{R}$, was considered:

$$\mathcal{M}_\xi := \{A_\xi, B, C, D_\xi\} = \{A + \tfrac{\xi}{2}I_n, B, C, D - \tfrac{\xi}{2}I_m\}, \tag{4.2a}$$

$$\mathcal{T}_\xi(s) := C(sI - A_\xi)^{-1}B + D_\xi = C((s - \tfrac{\xi}{2})I_n - A)^{-1}B + D - \tfrac{\xi}{2}I_m, \tag{4.2b}$$

$$\Phi_\xi(s) := \mathcal{T}_\xi^{\mathsf{H}}(-s) + \mathcal{T}_\xi(s). \tag{4.2c}$$

An important problem is to compute the values of ξ for which these systems are passive or strictly passive. The following theorem, which is a combination of [MVD20b, Theorem 4.5 and Lemma 6.2], classifies these values of ξ. Note that strict passivity of $\mathcal{T}(s)$ implies regularity of $\Phi(s)$, since $\Phi(\infty)$ is then regular. Thus, $\Phi_\xi(s)$ is also guaranteed to be regular for almost all ξ, since $\Phi_\xi(\infty)$ is regular for almost all ξ.

Theorem 4.3. *Let $\mathcal{M} := \{A, B, C, D\}$ with transfer function $\mathcal{T}(s)$ be minimal and proper. Then, for any $\xi \in \mathbb{R}$, the parameterized system $\mathcal{M}_\xi$ with transfer function $\mathcal{T}_\xi(s)$, as defined in (4.2), is also minimal and*

$$\Xi := \sup_{-\infty < \xi < \infty} \{\xi : \mathcal{T}_\xi(s) \text{ is strictly passive}\} = \max_{-\infty < \xi < \infty} \{\xi : \mathcal{T}_\xi(s) \text{ is passive}\} \tag{4.3}$$

is bounded. Moreover, $\mathcal{T}_\xi(s)$ is strictly passive for $\xi \in (-\infty, \Xi)$, passive but not strictly passive for $\xi = \Xi$, and non-passive for $\xi \in (\Xi, +\infty)$.

Proof. The minimality of the shifted transfer functions is well known. The rest of the proof is based on the fact that passivity of $\mathcal{T}_\xi(s)$ is linked to the inequality $W_c(X, \mathcal{M}) \succeq \xi \operatorname{diag}(X, I_m)$ for some $X \succ 0$, and that strict passivity of $\mathcal{T}_\xi(s)$ is linked to the strict inequality $W_c(X, \mathcal{M}) \succ \xi \operatorname{diag}(X, I_m)$ for some $X \succ 0$. Consequently, for all $\tilde{\xi} < \xi$, $\mathcal{T}_\xi(s)$ being passive implies that $\mathcal{T}_{\tilde{\xi}}(s)$ is strictly passive, and if $\mathcal{T}_\xi(s)$ is strictly passive, then $\mathcal{T}_{\tilde{\xi}}(s)$ is also strictly passive in an open neighborhood about ξ. This proves that the interval for ξ corresponding to strictly passive systems and non-passive systems are both open and connected, and that there is a single boundary point Ξ that must be passive, but not strictly passive. The boundedness of Ξ follows from the minimality of the realization for $\mathcal{T}_\xi(s)$. We refer to [MVD20b] for the details. $\qquad \square$

By computing Ξ, we can ascertain whether $\mathcal{M}_\xi$ is (strictly) passive or non-passive for all $\xi \in \mathbb{R}$. For a given value of ξ, by Definition 4.1, $\mathcal{M}_\xi$ corresponds to a strictly passive system if and only if $\Phi_\xi(s)$ is positive definite over the entire $\mathbf{i}\omega$ axis, with $\omega = \infty$ included, and A_ξ is asymptotically stable. Checking stability is straightforward, while checking the positive definiteness condition is more involved. For $\xi, \omega \in \mathbb{R}$, consider

$$\gamma(\xi, \omega) := \lambda_{\min}(\Phi_\xi(\mathbf{i}\omega)) \quad \text{and} \quad \begin{aligned} \gamma_\xi(\omega) &:= \gamma(\xi, \omega) \text{ where } \xi \in \mathbb{R} \text{ is fixed,} \\ \gamma_\omega(\xi) &:= \gamma(\xi, \omega) \text{ where } \omega \in \mathbb{R} \text{ is fixed.} \end{aligned} \tag{4.4}$$

Clearly, $\Phi_\xi(\mathbf{i}\omega) \succ 0$ if and only if $\gamma_\xi(\omega) > 0$, and at $\omega = \infty$, this is simply equivalent to $D_\xi^{\mathsf{H}} + D_\xi \succ 0$, with $\lim_{\omega \to \infty} \gamma_\xi(\omega) = \lambda_{\min}(D_\xi^{\mathsf{H}} + D_\xi) > 0$. If A_ξ is asymptotically stable, then $\mathcal{T}_\xi(s)$ has no poles on the imaginary axis, and so neither does $\Phi_\xi(s)$; hence, $\gamma_\xi(\omega)$ is a continuous function. Thus, if $D_\xi^{\mathsf{H}} + D_\xi \succ 0$ and A_ξ is asymptotically stable, then $\gamma_\xi(\omega_1) \leq 0$ if and only if $\det \Phi_\xi(\mathbf{i}\omega_2) = 0$ for $\omega_1, \omega_2 \in \mathbb{R}$, with $\omega_1 = \omega_2$ not necessarily holding. Summarizing, we have the following necessary and sufficient algebraic continuous-time conditions for the strict passivity of $\mathcal{T}_\xi(s)$:

(C1) $A_\xi = A + \tfrac{\xi}{2}I_n$ is asymptotically stable,

(C2) $D_\xi^\mathsf{H} + D_\xi = D^\mathsf{H} + D - \xi I_m \succ 0$ (positive definiteness at $\omega = \infty$),

(C3) $\det \Phi_\xi(\mathbf{i}\omega) \neq 0$ for all $\omega \in \mathbb{R}$ (implying positive definiteness for all finite ω provided that (C1) and (C2) also hold).

A bracket containing Ξ can be easily computed. A simple lower bound on Ξ is

$$\Xi_{\mathrm{lb}} := \lambda_{\min}(W_{\mathrm{c}}(I_n, \mathcal{M})), \tag{4.5}$$

as clearly

$$W_{\mathrm{c}}(I_n, \mathcal{M}_{\Xi_{\mathrm{lb}}}) = W_{\mathrm{c}}(I_n, \mathcal{M}) - \Xi_{\mathrm{lb}} I_{n+m} \succeq 0$$

holds, and so by Theorems 4.2 and 4.3, $\mathcal{T}_{\Xi_{\mathrm{lb}}}(s)$ is passive. Meanwhile, (C1) and (C2) will no longer be satisfied if ξ is too large: (C1) holds if and only if $\xi < -2\alpha(A)$, where $\alpha(A)$ is the spectral abscissa of A, and (C2) holds if and only if $\xi < \lambda_{\min}(D^\mathsf{H} + D)$. Thus, a simple upper bound for Ξ is

$$\Xi_{\mathrm{ub}} := \min\left\{-2\alpha(A), \lambda_{\min}(D^\mathsf{H} + D)\right\}. \tag{4.6}$$

Let us now look at $\xi \in [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}})$, where for this half open interval, A_ξ is asymptotically stable and $D_\xi^\mathsf{H} + D_\xi$ is positive definite. Therefore, in order to verify the strict passivity of $T_\xi(s)$, one only needs to verify condition (C3), i.e., that $\det \Phi_\xi(\mathbf{i}\omega) \neq 0$ for all $\omega \in \mathbb{R}$. This condition can be checked via the following result for $\mathcal{M}_\xi$, which is well known in the literature for general systems $\mathcal{M}$.

Theorem 4.4. *Let $\xi \in \mathbb{R}$ and $\mathcal{M}_\xi$ and $\Phi_\xi(s)$ be as defined in (4.2). Then, for any $\omega \in \mathbb{C}$ (not $\mathbb{R}$) such that $\mathbf{i}\omega \notin \Lambda(A_\xi)$, $\det \Phi_\xi(\mathbf{i}\omega) = 0$ if and only if*

(i) *$\det(M_\xi - \omega N) = 0$ and*

(ii) *if $D_\xi^\mathsf{H} + D_\xi$ is nonsingular, $\det(H_\xi - \mathbf{i}\omega I_{2n}) = 0$,*

where the regular Hermitian pencil $M_\xi - \lambda N$ is defined by matrices

$$M_\xi := \begin{bmatrix} 0 & A_\xi & B \\ A_\xi^\mathsf{H} & 0 & C^\mathsf{H} \\ B^\mathsf{H} & C & D_\xi^\mathsf{H}{+}D_\xi \end{bmatrix} \quad and \quad N := \begin{bmatrix} 0 & \mathbf{i}I_n & 0 \\ -\mathbf{i}I_n & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{4.7}$$

while the Hamiltonian matrix

$$H_\xi := \begin{bmatrix} A_\xi & 0 \\ 0 & -A_\xi^\mathsf{H} \end{bmatrix} - \begin{bmatrix} B \\ C^\mathsf{H} \end{bmatrix} (D_\xi^\mathsf{H}{+}D_\xi)^{-1} \begin{bmatrix} C & -B^\mathsf{H} \end{bmatrix}. \tag{4.8}$$

Proof. Writing

$$M_\xi - \omega N := \begin{bmatrix} 0 & A_\xi - \mathbf{i}\omega I_n & B \\ A_\xi^\mathsf{H} + \mathbf{i}\omega I_n & 0 & C^\mathsf{H} \\ B^\mathsf{H} & C & D_\xi^\mathsf{H}{+}D_\xi \end{bmatrix},$$

and using the Schur identity of determinants with respect to the leading $2n \times 2n$ block, which by assumption is nonsingular, we obtain that

$$\det(M_\xi - \omega N) = \det \begin{bmatrix} 0 & A_\xi - \mathbf{i}\omega I_n \\ A_\xi^\mathsf{H} + \mathbf{i}\omega I_n & 0 \end{bmatrix} \det \Phi_\xi(\mathbf{i}\omega).$$

As $\mathbf{i}\omega \notin \Lambda(A_\xi)$, we have (i), and since $\Phi_\xi(\mathbf{i}\omega)$ is regular, $M_\xi - \lambda N$ must be a regular pencil. Meanwhile, for (ii), $(H_\xi - \mathbf{i}\omega I_{2n}) \begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix}$ is the Schur complement of $M_\xi - \omega N$ with respect to the trailing $m \times m$ block, which requires the additional assumption that $D_\xi^\mathsf{H} + D_\xi$ is nonsingular. Via the Schur identity of determinants, we have that

$$\det(M_\xi - \omega N) = \det(D_\xi^\mathsf{H} + D_\xi) \det\left((H_\xi - \mathbf{i}\omega I_{2n}) \begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix}\right).$$

$\square$

Corollary 4.5. *Let $\xi \in [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}})$ and let $\gamma_\xi(\omega)$ be as defined in (4.4), where $\omega \in \mathbb{R}$. Then $\gamma_\xi(\omega)$ has at most $2n$ real zeros, all of which must be finite.*

Proof. If $\gamma_\xi(\omega) = 0$, then $\det \Phi_\xi(\mathbf{i}\omega) = 0$. Since $\xi \in [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}})$, (C1) and (C2) both hold, and so the assumptions of Theorem 4.4 are met. Hence, $\det \Phi_\xi(\mathbf{i}\omega) = 0$ if and only if $\det(H_\xi - \mathbf{i}\omega I_{2n}) = 0$. Finally, as $H_\xi \in \mathbb{C}^{2n \times 2n}$, it has $2n$ (finite) eigenvalues. $\qquad \square$

Given the bracket $[\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}}]$, Theorem 4.4 immediately leads to a bisection method for computing Ξ [MVD20b, p. 144]. For any $\xi \in [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}})$, (C3) can be verified by computing the eigenvalues of either $M_\xi - \lambda N$ or H_ξ (the pencil form is preferred numerically, since it only has a linear dependence on ξ). Via the following result, Mehrmann and Van Dooren also proposed a second improved algorithm for computing Ξ [MVD20b, p. 146].

Theorem 4.6. *Let $\gamma(\xi, \omega)$ and $\gamma_\xi(\omega)$ be the real functions defined in (4.4), where $\xi, \omega \in \mathbb{R}$. Then $\gamma(\xi, \omega)$ is continuous on the domain $[\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}}) \times \mathbb{R}$ and $\gamma_\xi(\omega)$ has the following properties:*

  (i) *if $\xi \in [\Xi_{\mathrm{lb}}, \Xi)$, then $\gamma_\xi(\omega) > 0$ for all ω,*

  (ii) *if $\xi = \Xi$, then $\gamma_\xi(\omega) \geq 0$ with $\gamma_\xi(\omega) = 0$ for at least one value of ω,*

  (iii) *if $\xi \in (\Xi, \Xi_{\mathrm{ub}})$ with $\Xi < \Xi_{\mathrm{ub}}$, then $\gamma_\xi(\omega) < 0$ holds on a subset of $\mathbb{R}$ consisting of non-overlapping open bounded intervals.*

Proof. This was proven in [MVD20b, Theorem 5.1] except for the claim in (iii) that the intervals are bounded, which follows directly from Corollary 4.5. $\qquad \square$

Mehrmann and Van Dooren's improved method computes a monotonically decreasing sequence $\{\xi_k\} \to \Xi$, where the initial estimate is $\xi_0 = \Xi_{\mathrm{ub}} - \tau$ for some small tolerance $\tau > 0$. On the kth iteration, via Theorem 4.4 and computing the eigenvalues of $M_{\xi_k} - \lambda N$, the bounded intervals where $\gamma_{\xi_k}(\omega) < 0$ are obtained. Taking $\hat{\omega}$ to be the midpoint of the largest of these intervals with $\gamma_{\xi_k}(\hat{\omega}) < 0$ holding, ξ_{k+1} is obtained by setting it to the smallest value of ξ such that $\gamma_{\hat{\omega}}(\xi) = 0$; this is done by computing all the eigenvalues of a matrix pencil closely related to $M_{\xi_k} - \lambda N$ (and of the same order). This process is continued in a loop until convergence to Ξ; see [MVD20b, section 5] for more details.

4.2   An HEC-based algorithm for computing the continuous-time Ξ

By Theorem 4.6, Ξ defined in (4.3) can instead be computed via this root-min problem:

$$f(\xi) = \min_{\omega \in \mathcal{D}_2} \gamma(\xi, \omega) = 0, \tag{4.9}$$

where $\xi \in \mathcal{D}_1 = [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}}]$, $\mathcal{D}_2 \subset \mathbb{R}$ is compact, and $\gamma(\xi, \omega)$ is defined in (4.4). If $\xi \neq \Xi_{\mathrm{ub}}$, by Corollary 4.5 and Theorem 4.6, $\mathcal{D}_2$ can be taken to be compact, since for all other values of $\xi \in \mathcal{D}_1$, $\gamma_\xi(\omega)$ has at most $2n$ zeros, which are all finite, and minimizers of $\gamma_\xi(\omega)$ that occur where $\gamma_\xi(\omega)$ is negative clearly must lie between these zeros. Since A_ξ is asymptotically stable for all $\xi < \Xi_{\mathrm{ub}}$, it follows that $\gamma_\xi(\omega)$ is also bounded below (and above) for all $\xi < \Xi_{\mathrm{ub}}$. Thus, (4.9) meets the criteria to be a valid root-min problem, and so we can use HEC to find pseudoroots of it.

Remark 4.7. *For a root-min problem, the initialization requirements for HEC, using the notation of (3.1) and Algorithm 1, are $\varepsilon_{\mathrm{lb}}, \varepsilon_0 \in \mathcal{D}_1$ and $x_0 \in \mathcal{D}_2$ such that $f(\varepsilon_0) \leq g(\varepsilon_0, x_0) \leq 0 < f(\varepsilon_{\mathrm{lb}})$. For computing Ξ, it will also be more convenient to use the convention that $f(\varepsilon_0) \leq g(\varepsilon_0, x_0) < 0 \leq f(\varepsilon_{\mathrm{lb}})$ holds.*

Remark 4.8. *We wish to compute Ξ to a desired relative accuracy, but when $\Xi = 0$, this notion does not make sense. To avoid providing technical implementation details regarding this, for sake of simplicity, we assume that neither Ξ nor any estimate for Ξ that our algorithm encounters is zero.*

First note that $0 \leq f(\Xi_{\mathrm{lb}})$ always holds. If we have a $\xi_0 \in [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}})$ and $\omega_0 \in \mathbb{R}$ such that $\gamma(\xi_0, \omega_0) < 0$, then the initialization conditions of HEC are met, and so HEC can be used to compute a pseudoroot $(\tilde{\xi}, \tilde{\omega})$ of (4.9) with both $\tilde{\xi} \in [\Xi_{\mathrm{lb}}, \xi_0)$ and $\Xi \leq \tilde{\xi}$ holding. To determine whether estimate $\tilde{\xi}$ is sufficiently close to Ξ, we do the following. For some relative tolerance $\tau \in (0,1)$, we set $\xi = \tilde{\xi} - \tau|\tilde{\xi}|$ (we assume that $\Xi_{\mathrm{lb}} < \xi$, as otherwise we are done) and then compute the real eigenvalues of $M_\xi - \lambda N$. If this matrix pencil has no real eigenvalues, then by Theorem 4.4, we know that $\gamma_\xi(\omega)$ has no zeros, and so by Theorem 4.6, $\xi < \Xi$ must hold. Thus, $\Xi \in (\tilde{\xi}, \xi]$, and so ξ must agree with Ξ to the desired number of digits.[2] Otherwise, if $M_\xi - \lambda N$ does have real eigenvalues, then $\gamma_\xi(\omega)$ has zeros, and if $\gamma_\xi(\omega) < 0$ holds on at least one of the intervals derived from these zeros, then $\Xi < \xi$ holds by Theorem 4.6. Updating ω_0 to be the midpoint of one of these intervals where $\gamma_\xi(\omega) < 0$, obviously $\gamma_\xi(\omega_0) < 0$ holds, and so HEC can be restarted to find a new pseudoroot $(\hat{\xi}, \hat{\omega})$ of (4.9) with $\hat{\xi} \in [\Xi, \xi)$. This process of running HEC and computing the real eigenvalues of $M_\xi - \lambda N$ to find regions where $\gamma_\xi(\omega)$ is negative is repeated in a loop until estimate ξ, which is decreasing monotonically, becomes sufficiently close to Ξ.

For initializing our new algorithm, it is required that we always choose $\xi < \Xi_{\mathrm{ub}}$; e.g., evaluating $\mathcal{T}_\xi(\mathbf{i}\omega)$ requires inverting $\mathbf{i}\omega I - A_\xi$, and this matrix may not always be invertible when $\xi = \Xi_{\mathrm{ub}}$. Choosing $\xi_0 = \Xi_{\mathrm{ub}} - \tau|\Xi_{\mathrm{ub}}|$ as our first estimate suffices, as this still allows us to obtain Ξ to the desired accuracy. Again, we assume that $\Xi_{\mathrm{lb}} < \xi_0$, as otherwise we are done. The user provides some $\omega_0 \in \mathbb{R}$ as an initial guess for HEC. If $\gamma_{\xi_0}(\omega_0) < 0$ holds, then our algorithm as described above can begin. Otherwise, we must find another point where $\gamma_{\xi_0}(\omega)$ is negative. This can be done in multiple ways. We can evaluate $\gamma_{\xi_0}(\omega)$ on a grid and/or randomly chosen points. We could also initialize some optimization solver at these points to try to find a minimizer $\tilde{\omega}$ of $\gamma_{\xi_0}(\omega)$ such that $\gamma_{\xi_0}(\tilde{\omega}) < 0$. If, after some reasonable amount of effort, such a point has not been found, we then resort to computing the eigenvalues of $M_{\xi_0} - \lambda N$ in order to obtain all the zeros of $\gamma_{\xi_0}(\omega)$. Then, as described above, we can determine if there exists a point where $\gamma_{\xi_0}(\omega)$ is negative. Since evaluating $\gamma_{\xi_0}(\omega)$ is much cheaper than computing the eigenvalues of $M_{\xi_0} - \lambda N$ (we elaborate on this momentarily), it is generally beneficial in terms of the overall runtime to first try a decent number of points, possibly with optimization. This also increases the chances that the first pseudoroot $(\tilde{\xi}, \tilde{\omega})$ of (4.9) found by HEC also provides its root, i.e., $\tilde{\xi} = \Xi$; in this case, our algorithm only computes the eigenvalues of $M_\xi - \lambda N$ for a single value of ξ. In contrast, recall that the earlier algorithm of Mehrmann and Van Dooren (described at the end of §4.1), on every iteration, requires computing the eigenvalues of $M_\xi - \lambda N$ plus the eigenvalues of a second related matrix pencil with the same order $(2n + m)$.

Pseudocode of our new algorithm for continuous-time Ξ is given in Algorithm 2. In practice, we observe that HEC is only restarted a handful of times, often just one. By construction of the valid data to initialize HEC on every iteration of Algorithm 2, it follows from Theorem 3.6 that HEC is indeed guaranteed to compute a pseudoroot of (4.9) on every iteration. Under mild assumptions that generally hold in practice, we now show that Algorithm 2 has local quadratic convergence to pseudoroots of (4.9).

Theorem 4.9 (Quadratic convergence of Algorithm 2). *Let $(\tilde{\xi}, \tilde{\omega})$ be any pseudoroot of* (4.9) *computed by HEC within Algorithm 2. If*

  (i) *after some point, HEC only generates iterates that lie on a single continuous path of local maximizers $\omega(\xi)$,*

  (ii) *$\tilde{\omega}$ is a local maximizer of $\gamma_{\tilde{\xi}}(\omega)$,*

  (iii) *$\frac{\partial}{\partial \xi}\gamma(\xi, \tilde{\omega}) \neq 0$,*

---

[2]Note that when $\tilde{\xi}$ is only sufficiently close to Ξ (and not equal), it does not necessarily follow that $\tilde{\omega}$ is close to the global minimizer(s) of $\gamma_\Xi(\omega)$.

Algorithm 2 HEC-based algorithm for continuous-time Ξ

Input: $\mathcal{M}$, $\omega_0 \in \mathbb{R}$, $\tau \in (0,1)$, Ξ_{lb} (4.5), and Ξ_{ub} (4.6)
Output: ξ such that $|\Xi - \xi| \leq \tau|\Xi|$ for continuous-time Ξ for $\mathcal{M}$

1: $\xi \leftarrow \Xi_{\mathrm{ub}} - \tau|\Xi_{\mathrm{ub}}|$
2: **if** $\xi \leq \Xi_{\mathrm{lb}}$ **then**
3:    **return**
4: **end if**
5: $\texttt{find\_negative} \leftarrow \gamma_\xi(\omega_0) \geq 0$ *// a boolean variable*
6: **while** true **do**
7:    **if** $\texttt{find\_negative}$ **then**
8:       $\Omega \leftarrow \{\omega \in \mathbb{R} : \det(M_\xi - \omega N) = 0 \text{ and } \gamma_\xi(\omega) = 0\}$
9:       **if** $\exists \omega_1, \omega_2 \in \Omega$ s.t. $\gamma_\xi(w) < 0 \; \forall \omega \in (\omega_1, \omega_2)$ **then**
10:          $\omega_0 \leftarrow 0.5(\omega_1 + \omega_2)$ *// (C3) holds, $\omega_1 < \omega_2$*
11:       **else**
12:          **return** *// $\gamma_\xi(\omega) \geq 0 \; \forall \omega \in \mathbb{R}$ and $\xi \approx \Xi$ to tolerance*
13:       **end if**
14:    **end if**
15:    *// $\gamma_\xi(\omega_0) < 0$ and $\Xi \in [\Xi_{\mathrm{lb}}, \xi)$ so run HEC with this initial data*
16:    $(\tilde{\xi}, \tilde{\omega}) \leftarrow$ a pseudoroot of (4.9) obtained by HEC with $\Xi \leq \tilde{\xi} < \xi$
17:    $\xi \leftarrow \tilde{\xi} - \tau|\tilde{\xi}|$
18:    $\texttt{find\_negative} \leftarrow$ true
19: **end while**

NOTE: *Per Remark 4.8, for simplicity of the pseudocode, we assume that $\Xi \neq 0$ and $\xi \neq 0$ never holds; if not, a second tolerance is needed. When the matrices defining $\mathcal{M}$ are all real, there is symmetry, i.e., $\gamma(\xi, -\omega) = \gamma(\xi, \omega)$, and so the search domain for ω can be reduced from $\mathbb{R}$ to $\omega \in [0, \infty)$. While taking advantage of this symmetry does not affect the asymptotic work complexity, it can nevertheless reduce the constant factors to speed up the overall run time.*

  (iv) $\gamma(\xi, \omega)$ *is twice continuously differentiable at* $(\tilde{\xi}, \tilde{\omega})$, *and*

  (v) $\omega(\xi)$ *is twice continuously differentiable at* $\tilde{\xi}$,

all hold, then Algorithm 2 converges to pseudoroot $(\tilde{\xi}, \tilde{\omega})$ *Q-quadratically. Furthermore, if in addition to* (ii), $\gamma(\tilde{\xi}, \tilde{\omega})$ *corresponds to a simple eigenvalue of* $\Phi_{\tilde{\xi}}(i\tilde{\omega})$ *and* $\nabla^2 \gamma_{\tilde{\xi}}(\tilde{\omega})$, *the Hessian of* $\gamma_{\tilde{\xi}}(\omega)$ *evaluated at* $\tilde{\omega}$, *is invertible, then* (iv) *and* (v) *are automatically satisfied.*

Proof. Conditions (i)–(v) implying the quadratic convergence of HEC is simply a translation of Theorem 3.8 to the setting of (4.9). For the second part of the theorem, if $\gamma(\tilde{\xi}, \tilde{\omega})$ corresponds to a simple eigenvalue, then $\gamma(\xi, \omega)$ is analytic near $(\tilde{\xi}, \tilde{\omega})$. The path of local maximizers of $\gamma_\xi(\omega)$ as ξ varies that contains $\tilde{\omega}$ can be characterized by the equality $\nabla\gamma(\xi, \omega) = 0$ in a neighborhood about $(\tilde{\xi}, \tilde{\omega})$. Thus, by the implicit function theorem, if $\nabla^2\gamma_{\tilde{\xi}}(\tilde{\omega})$ is invertible, $\nabla\gamma(\xi, \omega(\xi)) = 0$, where $\omega(\xi)$ is analytic. $\qquad\square$

    Under the smoothness assumptions of Theorem 4.9, in a neighborhood of the pseudoroot in question, the contraction and expansion phases within HEC can also be solved with fast convergence rates. Moreover, even if these assumptions do not hold, an extension of the analysis of Boyd and Balakrishnan [BB90] shows that near any minimizer, $\gamma_\xi(\omega)$ is twice continuously differentiable with Lipschitz second derivative, even if the minimizer is associated with an eigenvalue of $\Phi_\xi(i\omega)$ of multiplicity greater than one; for more details, see [MO21]. Thus, the expansion phases can always be solved quickly using secant or Newton's method, and any use of the early contraction technique discussed in Remark 3.9 should be limited, e.g., only in initial iterations when one cannot necessarily expect to be sufficiently close to the fast convergence regime.

For the contraction and expansion phases, we now describe how to compute the first and second derivatives of $\gamma_\xi(\omega)$ and $\gamma_\omega(\xi)$ defined in (4.4). Given a Hermitian matrix $H(t)$ depending on $t \in \mathbb{R}$, formulas for the first and second derivatives of a simple eigenvalue of a can be found in, e.g., [Lan64, OW95]. These formulas also need $H'(t)$ and $H''(t)$, so we present the first and second matrix derivatives of $\Phi_\xi(\mathbf{i}\omega)$ here. Letting $Z_k := C((\mathbf{i}\omega - \frac{\xi}{2})I_n - A)^{-k}B$, we have that

$$\frac{\partial}{\partial \xi}\mathcal{T}_\xi(\mathbf{i}\omega) = \tfrac{1}{2}(Z_2 - I_m), \qquad\qquad \frac{\partial^2}{\partial \xi^2}\mathcal{T}_\xi(\mathbf{i}\omega) = \tfrac{1}{2}Z_3, \qquad (4.10\text{a})$$

$$\frac{\partial}{\partial \omega}\mathcal{T}_\xi(\mathbf{i}\omega) = -\mathbf{i}Z_2, \qquad\qquad \frac{\partial^2}{\partial \omega^2}\mathcal{T}_\xi(\mathbf{i}\omega) = -2Z_3, \qquad (4.10\text{b})$$

and so

$$\frac{\partial}{\partial \xi}\Phi_\xi(\mathbf{i}\omega) = \tfrac{1}{2}(Z_2 + Z_2^{\mathsf{H}}) - I_m, \qquad \frac{\partial^2}{\partial \xi^2}\Phi_\xi(\mathbf{i}\omega) = \tfrac{1}{2}(Z_3 + Z_3^{\mathsf{H}}), \qquad (4.11\text{a})$$

$$\frac{\partial}{\partial \omega}\Phi_\xi(\mathbf{i}\omega) = -\mathbf{i}Z_2 + \mathbf{i}Z_2^{\mathsf{H}}, \qquad \frac{\partial^2}{\partial \omega^2}\Phi_\xi(\mathbf{i}\omega) = -2(Z_3 + Z_3^{\mathsf{H}}). \qquad (4.11\text{b})$$

Note that by using a Hessenberg factorization of A, per [Lau81], $\mathcal{T}_\xi(\mathbf{i}\omega)$ and Z_k can be transformed to equivalent systems where the inverses inside of them, for any values of ξ and ω, can be applied to a vector in just $\mathcal{O}(n^2)$ work instead of $\mathcal{O}(n^3)$. Thus, $\Phi_\xi(\mathbf{i}\omega)$ and its matrix derivatives given in (4.11) can be obtained in $\mathcal{O}(mn^2 + m^2n)$ work. Using the convention that computing the eigenvalues and eigenvectors of a matrix is an atomic operation with cubic complexity, the total cost to compute $\gamma(\xi, \omega)$ and its first and second derivatives with respect to ξ and ω is $\mathcal{O}(mn^2 + m^2n + m^3)$ work.

The cost of Algorithm 2 is dominated by computing the of zeros $\gamma_\xi(\omega)$. Since HEC generally converges quickly, as do its expansion and contraction phases, we can consider its number of evaluations of $\gamma(\xi, \omega)$ as a constant. Hence, in Algorithm 2, HEC does $\mathcal{O}(mn^2 + m^2n + m^3)$ work. Meanwhile, finding the zeros of $\gamma_\xi(\omega)$ involves computing all the eigenvalues of $M_\xi - \lambda N$, which itself is $\mathcal{O}((n+m)^3)$ work. Thus, for all but the smallest values of n, the HEC portion of Algorithm 2 should only be a fraction of the cost to compute the eigenvalues of $M_\xi - \lambda N$.

The "improved" algorithm of Mehrmann and Van Dooren has the same asymptotic work complexity as our method, but the hidden constant factor is much larger. This is partly because on each iteration, their algorithm solves two large eigenvalue problems of order $2n + m$. However, it also often requires more iterations than Algorithm 2 does. While Mehrmann and Van Dooren did not analyze the convergence properties of their method, our new framework of root-max problems and HEC also shows that their method converges at least Q-superlinearly under generic conditions. To see this, note that on each iteration, their method computes a single point where $\gamma_\xi(\omega) < 0$ holds (as opposed to finding a minimizer), but in the limit, these single points do converge to a minimizer of $\gamma_\xi(\omega)$ as $\xi \to \Xi$. In other words, their algorithm can also been seen as an HEC iteration with a very aggressive early contraction scheme. Per Remark 3.9, such an iteration converges at least superlinearly.

5 Discrete-time passive systems

We now present the discrete-time analogues of the optimization problem and our new algorithm given in §4.

5.1 Passivity of discrete-time proper parameterized systems

The material in this subsection closely follows [MVD20a]. For $z \in \mathbb{C}$, we now consider

$$\Phi(z) := \mathcal{T}^{\mathsf{H}}(z^{-1}) + \mathcal{T}(z),$$

which coincides with twice the Hermitian part of $\mathcal{T}(z)$ on the unit circle:

$$\Phi(\mathrm{e}^{\mathbf{i}\omega}) = [\mathcal{T}(\mathrm{e}^{\mathbf{i}\omega})]^{\mathsf{H}} + \mathcal{T}(\mathrm{e}^{\mathbf{i}\omega}).$$

Definition 5.1. *The transfer function $\mathcal{T}(z)$ is*

1. *passive if $\Phi(\mathrm{e}^{\mathbf{i}\omega}) \succeq 0$ for all $\omega \in (-\pi, \pi]$ and A is stable, i.e., its eigenvalues are in the closed unit disk, with any occurring on the unit circle being semi-simple,*

2. *strictly passive if $\Phi(\mathrm{e}^{\mathbf{i}\omega}) \succ 0$ for all $\omega \in (-\pi, \pi]$ and A is asymptotically stable, i.e., its eigenvalues are in the open unit disk.*

The necessary and sufficient conditions for passivity in the discrete-time case (see [MVD20a]) now make use of the linear matrix function

$$W_{\mathrm{d}}(X, \mathcal{M}) := \begin{bmatrix} X & XA & XB \\ A^{\mathsf{H}}X & X & C^{\mathsf{H}} \\ B^{\mathsf{H}}X & C & D^{\mathsf{H}} + D \end{bmatrix}. \tag{5.1}$$

Theorem 5.2. *Let $\mathcal{M} := \{A, B, C, D\}$ with transfer function $\mathcal{T}(z)$ be minimal and proper. Then $\mathcal{T}(z)$ is (strictly) passive if and only there exists an $X \in \mathbb{H}_n$ such that $X \succ 0$ and $W_{\mathrm{d}}(X, \mathcal{M}) \succeq 0$ ($W_{\mathrm{d}}(X, \mathcal{M}) \succ 0$).*

In [MVD20a], the following class of parametric systems was considered:

$$\mathcal{M}_\xi := \{A_\xi, B_\xi, C_\xi, D_\xi\} = \left\{ \tfrac{A}{1-\xi}, \tfrac{B}{1-\xi}, \tfrac{C}{1-\xi}, \tfrac{D-\xi I_m}{1-\xi} \right\}, \tag{5.2a}$$

$$\mathcal{T}_\xi(z) := C_\xi(zI_n - A_\xi)^{-1}B_\xi + D_\xi = \tfrac{1}{1-\xi}\left(C((1-\xi)zI_n - A)^{-1}B + D - \xi I_m \right), \tag{5.2b}$$

$$\Phi_\xi(z) := \mathcal{T}_\xi^{\mathsf{H}}(z^{-1}) + \mathcal{T}_\xi(z). \tag{5.2c}$$

where $\xi \in (-\infty, 1)$ and it is again important to compute for which values of ξ these systems are passive or strictly passive. The following theorem was given in [MVD20a], in a slightly modified form; we omit its proof as it is similar to that of Theorem 4.3. Note that strict passivity of $T(z)$ again implies regularity of $\Phi(z)$, since it is regular for any point on the unit circle. Moreover, $\Phi_\xi(z)$ is then also regular for almost all ξ since it is an analytic perturbation of $\Phi(z)$.

Theorem 5.3. *Let $\mathcal{M} := \{A, B, C, D\}$ with transfer function $\mathcal{T}(z)$ be minimal and proper. Then, for $\xi \in (-\infty, 1)$, the parameterized system $\mathcal{M}_\xi$ with transfer function $\mathcal{T}_\xi(z)$, as defined in (5.2), is also minimal and*

$$\Xi := \sup_{-\infty < \xi < 1} \{\xi : \mathcal{T}_\xi(z) \text{ is strictly passive}\} = \max_{-\infty < \xi < 1} \{\xi : \mathcal{T}_\xi(z) \text{ is passive}\} \tag{5.3}$$

is bounded. Moreover, $\mathcal{T}_\xi(z)$ is strictly passive for $\xi \in (-\infty, \Xi)$, passive but not strictly passive for $\xi = \Xi$, and not passive for $\xi \in (\Xi, 1)$.

By Definition 5.1, $\mathcal{M}_\xi$ is strictly passive if and only if $\Phi_\xi(z) \succ 0$ holds over the entire unit circle and A_ξ is asymptotically stable. Stability is again straightforward to check, but checking the positive definiteness condition is a little more subtle in the discrete-time case. For $\xi, \omega \in \mathbb{R}$, now consider

$$\gamma(\xi, \omega) := \lambda_{\min}(\Phi_\xi(\mathrm{e}^{\mathbf{i}\omega})) \quad \text{and} \quad \begin{aligned} \gamma_\xi(\omega) &:= \gamma(\xi, \omega) \ \text{ where } \xi \in \mathbb{R} \text{ is fixed,} \\ \gamma_\omega(\xi) &:= \gamma(\xi, \omega) \ \text{ where } \omega \in \mathbb{R} \text{ is fixed,} \end{aligned} \tag{5.4}$$

where $\Phi_\xi(z)$ is defined in (5.2). Clearly $\Phi_\xi(\mathrm{e}^{\mathbf{i}\omega}) \succ 0$ is equivalent to $\gamma_\xi(\omega) > 0$, and $\gamma_\xi(\omega)$ is continuous if A_ξ is asymptotically stable, as then $\Phi_\xi(z)$ cannot have any poles on the unit circle. Hence, if A_ξ is asymptotically stable and $\Phi_\xi(\mathrm{e}^{\mathbf{i}\tilde{\omega}}) \succ 0$ for some $\tilde{\omega} \in (-\pi, \pi]$, then $\Phi_\xi(\mathrm{e}^{\mathbf{i}\omega}) \succ 0$ for all $\omega \in (-\pi, \pi]$ if and only if $\det \Phi_\xi(\mathrm{e}^{\mathbf{i}\omega})$ has no zeros. Thus, $\mathcal{T}_\xi(z)$ is strictly passive if and only if the following conditions all hold:

(D1) $A_\xi = \frac{A}{1-\xi}$ is asymptotically stable,

16

(D2) $\Phi_\xi(e^{i\tilde\omega}) \succ 0$ (positive definiteness at a unimodular point, say, $e^{i\tilde\omega} = 1$),

(D3) $\det \Phi_\xi(e^{i\omega}) \neq 0$ for all $\omega \in (-\pi, \pi]$ (implying positive definiteness on the entire unit circle provided that (D1) and (D2) also hold).

In contrast to its continuous-time analogue (C2), note that (D2) does not require that $D_\xi^\mathsf{H} + D_\xi$ be positive definite (or even invertible).

A bracket containing the discrete-time Ξ is as follows. Again using the relation between the linear matrix inequalities of $\mathcal{M}$ and $\mathcal{M}_\xi$, with $X = 2I_n$, we can choose

$$\Xi_{\text{lb}} := \tfrac{1}{2}(\lambda_{\min} W_{\text{d}}(2I_n, \mathcal{M})) \tag{5.5}$$

as a lower bound on Ξ, since it follows that

$$(1 - \Xi_{\text{lb}})W_{\text{d}}(2I_n, \mathcal{M}_{\Xi_{\text{lb}}}) = W_{\text{d}}(2I_n, \mathcal{M}) - 2\Xi_{\text{lb}}I_{2n+m} \succeq 0,$$

holds and so by Theorems 5.2 and 5.3, we have that $\mathcal{T}_{\Xi_{\text{lb}}}(z)$ is passive. Meanwhile,

$$\Xi_{\text{ub}} := 1 - \rho(A) \tag{5.6}$$

is an upper bound, where $\rho(A)$ is the spectral radius of A, since obviously A_ξ is asymptotically stable if $\xi < \Xi_{\text{ub}}$.

Given $\xi \in [\Xi_{\text{lb}}, \Xi_{\text{ub}})$, (D1) must always holds, so to verify strict passivity of $T_\xi(z)$ we need to check that (D2) and (D3) also both hold. Checking (D2) is simple. If $\lambda_{\min}(\Phi_\xi(e^{i\tilde\omega})) \leq 0$ for any $\tilde\omega \in \mathbb{R}$, then $T_\xi(z)$ is not strictly passive, and there is no need to check (D3). Otherwise, since A_ξ is asymptotically stable, if $\lambda_{\min}(\Phi_\xi(e^{i\tilde\omega})) > 0$, we have that $\mathcal{T}_\xi(z)$ is strictly passive if and only if (D3) holds, which can be checked via the following result.[3]

Theorem 5.4. *Let $\xi \in (-\infty, 1)$ and $\mathcal{M}_\xi$ and $\Phi_\xi(z)$ be as defined in (5.2). Then, for any $z \neq 0 \in \mathbb{C}$ such that $z \notin \Lambda(A_\xi)$ and $z^{-1} \notin \Lambda(A_\xi^\mathsf{H})$, the latter of which follows automatically if $|z| = 1$, $\det \Phi_\xi(z) = 0$ if and only if*

(i) *$\det(M_\xi - zN_\xi) = 0$ and*

(ii) *if $\widetilde{D}_\xi$ is nonsingular, $\det(S_\xi - zT_\xi) = 0$,*

where $\widetilde{D}_\xi := D^\mathsf{H} + D - 2\xi I_m$, the regular pencil $M_\xi - \lambda N_\xi$ is defined by

$$M_\xi := \begin{bmatrix} 0 & A & B \\ (\xi-1)I_n & 0 & 0 \\ B^\mathsf{H} & C & \widetilde{D}_\xi \end{bmatrix} \quad and \quad N_\xi := \begin{bmatrix} 0 & (1-\xi)I_n & 0 \\ -A^\mathsf{H} & 0 & -C^\mathsf{H} \\ 0 & 0 & 0 \end{bmatrix}, \tag{5.7}$$

while the symplectic pencil $S_\xi - \lambda T_\xi$ is defined by

$$S_\xi := \begin{bmatrix} (\xi-1)I_n & 0 \\ -B\widetilde{D}_\xi^{-1}B^\mathsf{H} & A - B\widetilde{D}_\xi^{-1}C \end{bmatrix} \quad and \quad T_\xi := \begin{bmatrix} (B\widetilde{D}_\xi^{-1}C - A)^\mathsf{H} & C^\mathsf{H}\widetilde{D}_\xi^{-1}C \\ 0 & (1-\xi)I_n \end{bmatrix}. \tag{5.8}$$

Proof. Writing

$$M_\xi - zN_\xi := \begin{bmatrix} 0 & A + (\xi-1)zI_n & B \\ zA^\mathsf{H} + (\xi-1)I_n & 0 & zC^\mathsf{H} \\ B^\mathsf{H} & C & \widetilde{D}_\xi \end{bmatrix},$$

and using the Schur identity of determinants with respect to the leading $2n \times 2n$ block, which by assumption is nonsingular, we obtain that

$$\det(M_\xi - zN_\xi) = \det\begin{bmatrix} 0 & A + (\xi-1)zI_n \\ zA^\mathsf{H} + (\xi-1)I_n & 0 \end{bmatrix} \det((1 - \xi)\Phi_\xi(z)).$$

---

[3]The generalized eigenvalue problem given by the matrices in (5.7) is denoted $\Gamma(\xi, \omega)$ in [MVD20a, p. 1263], but note that its bottom right block, $D^\mathsf{H} + D - \xi I_m$, contains a typo; it should be $D^\mathsf{H} + D - 2\xi I_m$, which we denote $\widetilde{D}_\xi$ in Theorem 5.4.

As $z \notin \Lambda(A_\xi)$ and $z^{-1} \notin \Lambda(A_\xi^{\mathsf{H}})$ we have (i), and since $\Phi_\xi(z)$ is regular, $M_\xi - \lambda N_\xi$ must be a regular pencil. For (ii), we again apply the Schur identity of determinants, now with respect to the trailing $m \times m$ block, which is possible by our additional assumption that $\widetilde{D}_\xi$ is nonsingular:

$$\det(M_\xi - zN_\xi) = \det \widetilde{D}_\xi \det \left(\begin{bmatrix} 0 & A+(\xi-1)zI_n \\ zA^{\mathsf{H}}+(\xi-1)I_n & 0 \end{bmatrix} - \begin{bmatrix} B \\ zC^{\mathsf{H}} \end{bmatrix} \widetilde{D}_\xi^{-1} \begin{bmatrix} B^{\mathsf{H}} & C \end{bmatrix} \right).$$

Clearly $\det(M_\xi - zN_\xi)$ can only be zero if and only if the last determinant above is zero. Multiplying the matrix inside this determinant by $\begin{bmatrix} 0 & I_n \\ I_n & 0 \end{bmatrix}$ from the left and rearranging terms yields $S_\xi - zT_\xi$. This matrix pencil is easily verified as symplectic, i.e., for $J := \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$, $S_\xi^{\mathsf{H}} J S_\xi = T_\xi^{\mathsf{H}} J T_\xi$ holds. $\square$

Corollary 5.5. *Let $\xi \in [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}})$ and let $\gamma_\xi(\omega)$ be as defined in (5.4), where $\omega \in (-\pi, \pi]$. Then $\gamma_\xi(\omega)$ has at most $2n$ real zeros.*

Proof. If ω is a zero of $\gamma_\xi(\omega)$, then $\det \Phi_\xi(\mathrm{e}^{\mathrm{i}\omega}) = 0$. As $\xi \in [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}})$, $\rho(A_\xi) < 1$ holds, and so the assumptions of Theorem 5.4 are met. Hence, $\det \Phi_\xi(\mathrm{e}^{\mathrm{i}\omega}) = 0$ if and only if $\det(M_\xi - \mathrm{e}^{\mathrm{i}\omega} N_\xi)$. The proof is completed by noting that $\mathrm{rank}\, N_\xi \leq 2n$. $\square$

Using Theorem 5.4, Mehrmann and Van Dooren proposed a bisection method to compute discrete-time Ξ, and via the following result, a discrete-time analogue of their improved procedure we described in §4.1; for more details, see [MVD20a, section 7].

Theorem 5.6. *Let $\gamma(\xi, \omega)$ and $\gamma_\xi(\omega)$ be the real functions defined in (5.4), where $\xi, \omega \in \mathbb{R}$. Then $\gamma(\xi, \omega)$ is continuous on the domain $[\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}}) \times (-\pi, \pi]$ and $\gamma_\xi(\omega)$ has the following properties:*

(i) *if $\xi \in [\Xi_{\mathrm{lb}}, \Xi)$, then $\gamma_\xi(\omega) > 0$ for all ω,*

(ii) *if $\xi = \Xi$, then $\gamma_\xi(\omega) \geq 0$ with $\gamma_\xi(\omega) = 0$ for at least one value of ω,*

(iii) *if $\xi \in (\Xi, \Xi_{\mathrm{ub}})$ with $\Xi < \Xi_{\mathrm{ub}}$, then $\gamma_\xi(\omega) < 0$ holds on a subset of $(-\pi, \pi]$ consisting of non-overlapping open intervals or on all of $(-\pi, \pi]$.*

Proof. Statements (i) and (ii) follow from [MVD20a], while (iii) follows from the facts that $\gamma_\xi(\omega)$ is continuous, and by Corollary 5.5, it can have at most $2n$ zeros. $\square$

Remark 5.7. *For any $\xi \in (\Xi, \Xi_{\mathrm{ub}})$, note that $\gamma_\xi(\omega)$ always has at least two zero-crossings in the continuous-case setting, while the discrete-time version of $\gamma_\xi(\omega)$ may not have any, i.e., $\gamma_\xi(\omega) < 0$ can hold for all $\omega \in (-\pi, \pi]$. This is why it is necessary to check that both (D2) and (D3) hold at each estimate ξ encountered when computing discrete-time Ξ, but in the continuous-time case, only (C3) needs to be checked at each estimate. Note that the descriptions of the discrete-time algorithms in [MVD20a, section 7] do not make this important distinction clear.*

5.2 An HEC-based algorithm for computing discrete-time Ξ

By Theorem 5.6, Ξ defined in (5.3) can be computed via this root-min problem:

$$f(\xi) = \min_{\omega \in \mathcal{D}_2} \gamma(\xi, \omega) = 0, \tag{5.9}$$

where now $\xi \in \mathcal{D}_1 = [\Xi_{\mathrm{lb}}, \Xi_{\mathrm{ub}}]$, $\mathcal{D}_2 = (-\pi, \pi]$ is obviously compact, $\gamma(\xi, \omega)$ is defined in (5.4), and $\gamma_\xi(\omega)$ is bounded below. Our continuous-time HEC-based algorithm and results from §4.2 extend to the discrete-time setting and work similarly, so for brevity, we only focus on the key points and differences.

Pseudocode for our new algorithm for discrete-time Ξ is given in Algorithm 3. Per Remark 5.7, the need to check that both (D2) and (D3) hold on each iteration means that the

---

Algorithm 3 HEC-based algorithm for discrete-time Ξ

Input: $\mathcal{M}$, $\omega_0 \in (-\pi, \pi]$, $\tau \in (0, 1)$, Ξ_{lb} (5.5), and Ξ_{ub} (5.6)
Output: ξ such that $|\Xi - \xi| \leq \tau |\Xi|$ for discrete-time Ξ for $\mathcal{M}$

1: $\xi \leftarrow \Xi_{\mathrm{ub}} - \tau |\Xi_{\mathrm{ub}}|$
2: **if** $\xi \leq \Xi_{\mathrm{lb}}$ **then**
3:   **return**
4: **end if**
5: $\texttt{find\_negative} \leftarrow \gamma_\xi(\omega_0) \geq 0$ // *a boolean variable*
6: **while** true **do**
7:   **if** $\texttt{find\_negative}$ **then**
8:     **if** $\gamma_\xi(0) < 0$ **then**
9:       $\omega_0 \leftarrow 0$ // *(D2) does not hold*
10:    **else**
11:      $\Omega \leftarrow \{\omega \in (-\pi, \pi] : \det(M_\xi - \mathrm{e}^{\mathbf{i}\omega} N_\xi) = 0 \text{ and } \gamma_\xi(\omega) = 0\}$
12:      **if** $\exists \omega_1, \omega_2 \in \Omega \cup \{\min \Omega + 2\pi\}$ s.t. $\gamma_\xi(w) < 0 \; \forall \omega \in (\omega_1, \omega_2)$ **then**
13:        $\omega_0 \leftarrow 0.5(\omega_1 + \omega_2)$ // *(D2) and (D3) hold, $\omega_1 < \omega_2$*
14:      **else**
15:        **return** // *$\gamma_\xi(\omega) \geq 0 \; \forall \omega \in (-\pi, \pi]$ and $\xi \approx \Xi$ to tolerance*
16:      **end if**
17:    **end if**
18:   **end if**
19:   // *$\gamma_\xi(\omega_0) < 0$ and $\Xi \in [\Xi_{\mathrm{lb}}, \xi)$ so run HEC with this initial data*
20:   $(\tilde{\xi}, \tilde{\omega}) \leftarrow$ a pseudoroot of (5.9) obtained by HEC with $\Xi \leq \tilde{\xi} < \xi$
21:   $\xi \leftarrow \tilde{\xi} - \tau |\tilde{\xi}|$
22:   $\texttt{find\_negative} \leftarrow \texttt{true}$
23: **end while**

NOTE: *See Remark 4.8 and Algorithm 2 for more details on tolerances and symmetry. In line 12, $\Omega \cup \{\min \Omega + 2\pi\}$ is used so that the "wrap-around" interval, i.e., $[\max \Omega, \min \Omega + 2\pi]$ is not missed.*

pseudocode is a bit more complicated than for continuous-time Ξ. As such, one might conclude that the problem of computing Ξ is trickier in the discrete-time case; however, as we explain in the numerical results, it seems that the exact opposite is true, due to a numerical issue that only arises in the continuous-time case. To implement HEC for Algorithm 3, we make use of the first and second derivatives of $\gamma_\xi(\omega)$ and $\gamma_\omega(\xi)$ defined in (5.4). To that end, we provide the discrete-time analogues of the matrix derivatives given in (4.10), as the remaining computations are readily apparent. Letting $Z_k := C((1 - \xi)\mathrm{e}^{\mathbf{i}\omega} I_n - A)^{-k} B$, we have that

$$\frac{\partial}{\partial \xi} \mathcal{T}_\xi(\mathrm{e}^{\mathbf{i}\omega}) = \frac{\mathcal{T}_\xi(\mathrm{e}^{\mathbf{i}\omega}) + \mathrm{e}^{\mathbf{i}\omega} Z_2 - I_m}{1 - \xi}, \qquad \frac{\partial^2}{\partial \xi^2} \mathcal{T}_\xi(\mathrm{e}^{\mathbf{i}\omega}) = \frac{2}{1 - \xi} \left(\mathrm{e}^{2\mathbf{i}\omega} Z_3 + \frac{\partial \mathcal{T}_\xi(\mathrm{e}^{\mathbf{i}\omega})}{\partial \xi} \right), \qquad (5.10\mathrm{a})$$

$$\frac{\partial}{\partial \omega} \mathcal{T}_\xi(\mathrm{e}^{\mathbf{i}\omega}) = -\mathbf{i} \mathrm{e}^{\mathbf{i}\omega} Z_2, \qquad \frac{\partial^2}{\partial \omega^2} \mathcal{T}_\xi(\mathrm{e}^{\mathbf{i}\omega}) = \mathrm{e}^{\mathbf{i}\omega} Z_2 - 2(1 - \xi)\mathrm{e}^{2\mathbf{i}\omega} Z_3. \qquad (5.10\mathrm{b})$$

In Algorithm 3, the costs to run HEC and compute zeros of $\gamma_\xi(\omega)$ are the same as in the continuous-time setting discussed in §4.2. Theorem 4.9 also extends, i.e., under mild assumptions that generally hold in practice, Algorithm 3 converges quadratically to pseudoroots of (5.9). Relatedly, Mehrmann and Van Dooren's improved algorithm [MVD20a, section 7] for discrete-time Ξ also converges at least superlinearly.

6 Numerical experiments

We implemented the continuous- and discrete-time versions of our new HEC-based method and the improved midpoint-based iteration of Mehrmann and Van Dooren. In this section, for brevity, we use HEC to refer to former (Algorithms 2 and 3) and MP (for midpoint) to refer to the latter. All codes were implemented with relative tolerances and set to compute Ξ to 14 digits. Experiments were done using MATLAB R2021a on a 2020 MacBook Pro with a quad-core Intel Core i5 1038NG7 CPU and 16 GB of RAM running macOS 10.15.7. Code and data to reproduce all experiments is included in the supplementary material.

6.1 Implementation details

We first discuss implementing Algorithm 1. The expansion phase was implemented using `fmincon`, while the contraction phase was implemented using a our own Halley-bisection root-finding code; first and second derivative information is used in both. Due to rounding errors, it may be that contraction phase sometimes computes a root $\hat{\varepsilon}_k$ of $g_{x_k}(\varepsilon)$ such that $g_{x_k}(\hat{\varepsilon}_k) < 0$, instead of $g_{x_k}(\hat{\varepsilon}_k) \geq 0$, which is required at every iteration (for a root-max problem). However, if this occurs, it suffices to just perturb the computed root by a small multiple of the Halley step to correct the sign; a more complicated workaround involving shifting the root problems is suggested in [MO16, section 7] and [GGMO17, Appendix A], but we do not recommend that. Algorithm 1 is terminated at an approximate pseudoroot once both ε_k and x_k are no longer changing significantly with respect to their respective previous values; this condition is checked twice per iteration, after the contraction phase and after the expansion phase. Since in the context of computing Ξ, the expansion phases can be solved quickly, we did not use early contraction.

For our purposes here, we simply used `eig` for all eigenvalues problems, though it is advisable to use structure-preserving solvers for numerical robustness; e.g., see [BBMX99]. To compute zeros of $\gamma_\xi(\omega)$, we used the pencils given by the matrices in (4.7) and (5.7) and respectively identified their real and unimodular eigenvalues using a tolerance.[4] Note that if $\gamma_\xi(\omega)$ has a minimizer or maximizer $\hat{\omega}$ such that $\gamma_\xi(\hat{\omega}) = 0$ (or approximately equal), then this corresponds to a (nearly) multiple eigenvalue (with multiplicity at least two) of the pencil given by (4.7) or (5.7), as appropriate. This always happens as any of the methods approach Ξ, and it is generally also true at computed pseudoroots and at $\omega = 0$ when the problems have symmetry. Due to rounding errors, such eigenvalues, even when computed via a structure-preserving solver, may not be detected as (close to) real or unimodular. If this happens, a zero of $\gamma_\xi(\omega)$ will be missed, which in turn can cause any of the algorithms to stagnate. Fortunately, a robust fix is easy: if $(\tilde{\xi}, \tilde{\omega})$ is the most recent computed pseudoroot, simply explicitly add $\tilde{\omega}$ as a zero of $\gamma_\xi(\omega)$; a similar fix is also necessary for MP. For more details, see [BLO03, pp. 371–373], where this fix was proposed in the context of computing the pseudospectral abscissa.

For continuous-time Ξ, there is an additional numerical difficulty when computing the zeros of $\gamma_\xi(\omega)$ when $\xi \approx \Xi_{\mathrm{ub}}$. Although these zeros must be finite, they still may be arbitrarily far away from the origin, and so there may be large errors in the imaginary parts of the corresponding computed real eigenvalues of $M_\xi - \lambda N$. Mehrmann and Van Dooren recommended using a tolerance so that the first estimate ξ tested was sufficiently far away from Ξ_{ub} to help avoid such problems. However, we have observed that even a relatively large perturbation may still be insufficient to avoid failure of MP. Our MP code uses $\xi_0 = \Xi_{\mathrm{ub}} - |\Xi_{\mathrm{ub}}|10^{-4}$, but only small perturbations are done for subsequent estimates in order to obtain the desired 14-digit accuracy; of course, if $\Xi \approx \Xi_{\mathrm{ub}}$, high accuracy may not be possible with MP. In contrast, our HEC-based method is much less susceptible to

---

[4]If A, B, C, and D are all real, then `eig` returns real eigenvalues of (4.7) without any rounding error in their imaginary parts; otherwise, the imaginary part may be nonzero numerically.

Table 1: MP and HEC compared on two continuous-time (cont.) and one discrete-time (disc.) examples. For Random, MP was tested in two configurations: MP-fail with $\xi_0 = \Xi_{ub} - |\Xi_{ub}|10^{-5}$ and MP with $\xi_0 = \Xi_{ub} - |\Xi_{ub}|10^{-4}$. The number of iterations is shown in the "iters." column; the average number of iterations of Algorithm 1 is also given in parentheses for HEC. The number of eigenvalue problems solved is shown under the "# `eig` (order, type)" columns, separated into the number of order $2n + m$ matrix pencils "$(2n + m, \mathrm{P})$" and the number of order m matrices "(m, M)". The overall running time in seconds is given under "time (sec.)", while the computed estimates for Ξ is given in the rightmost column.

| Alg. | iters. | # `eig` (order, type) | | time (sec.) | Ξ estimate |
		$(2n + m, \mathrm{P})$	(m, M)		
Random ($n = 200$, $m = 10$, cont.) — $[\Xi_{lb}, \Xi_{ub}] = [-25.56407, -10.90965]$					
MP-fail	1	1	1	0.440	-10.9097612001839
MP	14	27	49	8.892	-14.4073741346323
HEC	2(5.0)	2	78	0.909	-14.4073741346323
RLC ($n = 200$, $m = 1$, cont.) — $[\Xi_{lb}, \Xi_{ub}] = [-32.1267, 2.022606]$					
MP	4	8	18	2.360	0.562483988863916
HEC	1(4.0)	2	41	0.767	0.562483988863891
ISS ($n = 228$, $m = 3$, disc.) — $[\Xi_{lb}, \Xi_{ub}] = [-3.007437, 3.117278 \times 10^{-6}]$					
MP	15	29	550	8.490	$-9.37320364701040 \times 10^{-5}$
HEC	2(4.0)	1	97	0.374	$-9.37320364699013 \times 10^{-5}$

this issue, since even if only one root of $\gamma_\xi(\omega)$ is detected, it generally can still be used to start Algorithm 1. Even if this root is a stationary point, a small perturbation to the left or right generally yields a point for starting Algorithm 1. More generally, structure-preserving eigensolvers can be used or one can increase the allowed amount of rounding error in the imaginary part of an eigenvalue in proportion with the magnitude of the eigenvalue.

Finally, in line 8 of Algorithm 3 when checking (D2), instead of always checking the sign of $\gamma_\xi(0)$, after the first pseudoroot has been computed we check $\gamma_\xi(\tilde{\omega} + \frac{1}{2}\pi)$. If this is negative, we set $\omega_0 \leftarrow \tilde{\omega} + \frac{1}{2}\pi$ in line 9. The reason is because if the previous pseudoroot has $\tilde{\omega} = 0$, $\gamma_\xi(\tilde{\omega}) < 0$ almost always holds due to rounding error even though it should be exactly zero. Shifting by, e.g., $\frac{1}{2}\pi$, ensures that (D2) is checked at a new point; note that shifting by π or 2π would not ensure this.

6.2 Experiments

We begin with a randomly generated continuous-time example with complex matrices (denoted Random) to illustrate (i) when our method encounters at least two pseudoroots before converging (see Fig. 3a) and (ii) the aforementioned difficulty of computing zeros of $\gamma_\xi(\omega)$ when $\xi \approx \Xi_{ub}$ (see Fig. 4a). In Table 1, we see that MP is about ten times slower than HEC. Although HEC required more computations of $\lambda_{\min}(\Phi_\xi(e^{i\omega}))$, it only needed to solve two of the large eigenvalue problems involving $M_\xi - \lambda N$. Meanwhile, MP required 27 solves with the pencils and took 14 iterations to converge. HEC converged to Ξ at its second pseudoroot, and Algorithm 1 on average took 5.0 iterations to converge to a pseudoroot.

Our second continuous-time example is the electric RLC circuit model used in [BGVD20]. We refer to Fig. 3b and Table 1 for the complete performance details, but note that HEC was over three times faster than MP for this RLC example, with both methods converging faster and with less work than on the random example.

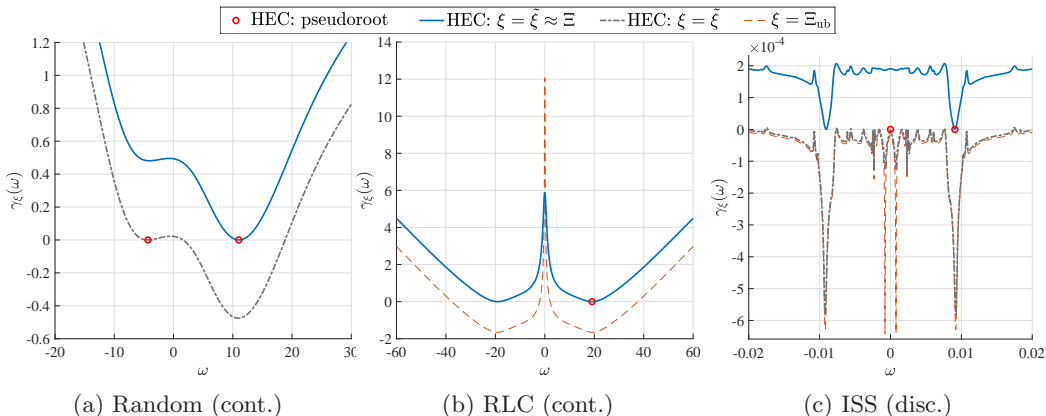(a) Random (cont.)   (b) RLC (cont.)   (c) ISS (disc.)

Figure 3: The pseudoroots $(\tilde{\xi}, \tilde{\omega})$ and corresponding estimates $\tilde{\xi}$ for Ξ computed by HEC until convergence with $\tilde{\xi} \approx \Xi$.

To compare the discrete-time methods, we used the ISS model from the SLICOT benchmark examples.[5] Since ISS is a continuous-time model, we converted it to a minimal discrete-time one by calling `c2d` using a sampling time of 0.001 followed by `minreal`. In Fig. 3b and Table 1, we see that HEC was almost 23 times faster than MP, again due to the great disparity in the number of large generalized eigenvalue problems solved. In fact, for ISS, HEC also solved far fewer smaller standard eigenvalues problems as well. From Table 1 and Fig. 4b, we also see that MP did not quite compute Ξ to the requested 14-digit accuracy, while HEC apparently did. This slight inaccuracy is the result of MP solving root problems via solving eigenvalue problems, but such errors can be larger; see the caption of Fig. 4b for more details.

While we have established that MP converges at least superlinearly, an examination of its iterates (not shown) seems to indicate that it too may converge quadratically like HEC. However, as demonstrated by Random and ISS, where MP respectively required 14 and 15 iterations, MP can incur many iterations before it gets near its faster convergence regime. The key problem on these examples is that MP chooses the largest interval where $\gamma_\xi(\omega) < 0$ holds to determine how to reduce estimate ξ. But this can be a particularly bad strategy if $\gamma_\xi(\omega)$ has a zero very far away from the origin, as is the case for both Random and ISS when $\xi \approx \Xi_{\mathrm{ub}}$. While one could consider altering this strategy to improve performance, such an MP variant would still be slower than HEC and also still have the aforementioned numerical issues.

7 Conclusion

Using Hybrid Expansion-Contraction, we have presented faster and more numerically robust algorithms to compute Ξ, the extremal real value for which a given parametric linear time-invariant system is passive, a problem which is linked to maximizing the passivity radius. Our new methods outperform the existing algorithms of Mehrmann and Van Dooren, and for large-scale problems, we also note that with a sparse eigenvalue solver, Hybrid Expansion-Contraction can be used alone to efficiently estimate Ξ, which the earlier methods cannot do. We hope that our generalization of Hybrid Expansion-Contraction, its convergence guarantees, and identification of root-max problems will help facilitate new fast and robust numerical methods for other quantities, for both small-scale and large-scale problems.
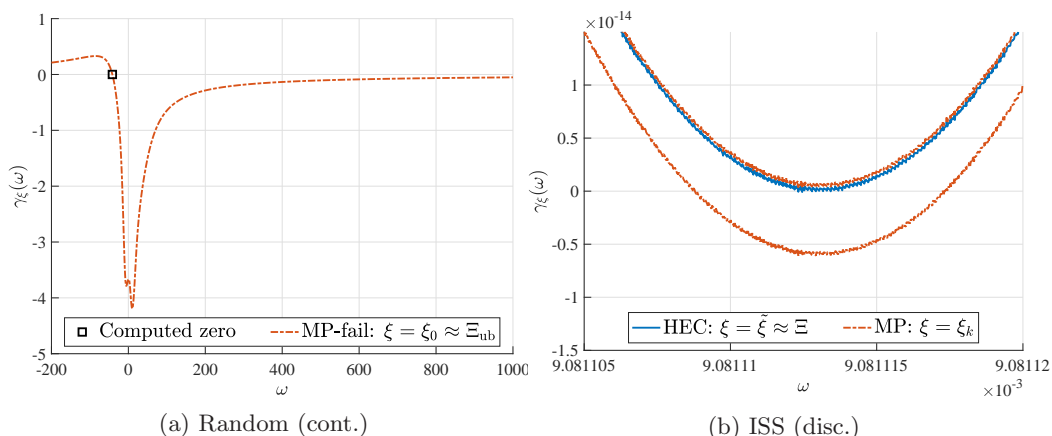
---

[5]Available at `http://slicot.org/20-site/126-benchmark-examples-for-model-reduction`.

(a) Random (cont.)

(b) ISS (disc.)

Figure 4: On the left, MP was initialized with $\xi_0 = \Xi_{\mathrm{ub}} - |\Xi_{\mathrm{ub}}|10^{-5}$. In this case, $\gamma_{\xi_0}(\omega)$ has two zeros at approximately -41.6 and 460600.9, but the latter is not detected due to rounding errors when computing the eigenvalues of $M_{\xi_0} - \lambda N$. Consequently, MP erroneously terminates at ξ_0 with no digits of accuracy but does converge properly when initialized with $\xi_0 = \Xi_{\mathrm{ub}} - |\Xi_{\mathrm{ub}}|10^{-4}$. On the right, we see that eigenvalue computations used in MP to compute the smallest roots of $\gamma_\omega(\xi)$ incurs more rounding errors than our HEC-based approach. See [BM19, section 9.2] for an example where half of the precision can be lost when solving root problems using these eigenvalue techniques.

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.

[BBMX99] P. Benner, R. Byers, V. Mehrmann, and H. Xu. Numerical methods for linear-quadratic and H_∞ control problems. In G. Picci and D. S. Gilliam, editors, *Dynamical Systems, Control, Coding, Computer Vision: New Trends, Interfaces, and Interplay*, volume 25 of *Progress in Systems and Control Theory*, pages 203–222. Birkhäuser, Basel, 1999.

[BGVD20] P. Benner, P. Goyal, and P. Van Dooren. Identification of port-Hamiltonian systems from frequency response data. *Systems Control Lett.*, 143:104741, 2020.

[BLO03] J. V. Burke, A. S. Lewis, and M. L. Overton. Robust stability and a criss-cross algorithm for pseudospectra. *IMA J. Numer. Anal.*, 23(3):359–375, 2003.

[BM19] P. Benner and T. Mitchell. Extended and improved criss-cross algorithms for computing the spectral value set abscissa and radius. *SIAM J. Matrix Anal. Appl.*, 40(4):1325–1352, 2019.

[DES82] R. S. Dembo, S. C. Eisenstat, and T. Steihaug. Inexact Newton methods. *SIAM J. Numer. Anal.*, 19(2):400–408, 1982.

[FGL20] A. Fazzi, N. Guglielmi, and C. Lubich. Finding the nearest passive or non-passive system via hamiltonian eigenvalue optimization. e-print arXiv:2010.15954, arXiv, October 2020. math.NA.

[GGMO17] N. Guglielmi, M. Gürbüzbalaban, T. Mitchell, and M. L. Overton. Approximating the real structured stability radius with Frobenius-norm bounded perturbations. *SIAM J. Matrix Anal. Appl.*, 38(4):1323–1353, 2017.

[GGO13] N. Guglielmi, M. Gürbüzbalaban, and M. L. Overton. Fast approximation of the H_∞ norm via optimization over spectral value sets. *SIAM J. Matrix Anal. Appl.*, 34(2):709–737, 2013.

[Glo84] K. Glover. All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error norms. *Internat. J. Control*, 39(6):1115–1193, 1984.

[GS18] N. Gillis and P. Sharma. Finding the nearest positive-real system. *SIAM J. Numer. Anal.*, 56(2):1022–1047, 2018.

[HP90a] D. Hinrichsen and A. J. Pritchard. A note on some differences between real and complex stability radii. *Systems Control Lett.*, 14(5):401–408, 1990.

[HP90b]   D. Hinrichsen and A. J. Pritchard. Real and complex stability radii: a survey. In *Control of uncertain systems (Bremen, 1989)*, volume 6 of *Progr. Systems Control Theory*, pages 119–162. Birkhäuser Boston, Boston, MA, 1990.

[HP05]   D. Hinrichsen and A. J. Pritchard. *Mathematical Systems Theory I*. Springer-Verlag, Berlin, 2005.

[Lan64]   P. Lancaster. On eigenvalues of matrices dependent on a parameter. *Numer. Math.*, 6:377–387, 1964.

[Lau81]   A. Laub. Efficient multivariable frequency response computations. *IEEE Trans. Autom. Control*, 26(2):407–408, April 1981.

[Mit14]   T. Mitchell. *Robust and efficient methods for approximation and optimization of stability measures*. PhD thesis, New York University, New York, NY 10003, USA, September 2014.

[MO16]   T. Mitchell and M. L. Overton. Hybrid expansion-contraction: a robust scaleable method for approximating the H_∞ norm. *IMA J. Numer. Anal.*, 36(3):985–1014, 2016.

[MO21]   T. Mitchell and M. L. Overton. On properties of univariate max functions at local maximizers. e-print arxiv:2108.07754, arXiv, August 2021.

[MVD20a] V. Mehrmann and P. Van Dooren. Optimal robustness of passive discrete-time systems. *IMA J. Math. Control Inform.*, 37(4):1248–1269, July 2020.

[MVD20b] V. Mehrmann and P. M. Van Dooren. Optimal robustness of port-Hamiltonian systems. *SIAM J. Matrix Anal. Appl.*, 41(1):134–151, 2020.

[NW99]   J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 1999.

[OVD05]   M. L. Overton and P. Van Dooren. On computing the complex passivity radius. In *Proceedings of the 44th IEEE Conference on Decision and Control*, volume 49, pages 7960–7964, December 2005.

[OW95]   M. L. Overton and R. S. Womersley. Second derivatives for optimizing eigenvalues of symmetric matrices. *SIAM J. Matrix Anal. Appl.*, 16(3):697–718, 1995.

[RW98]   R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

[Van85]   C. F. Van Loan. How near is a stable matrix to an unstable matrix? In *Linear algebra and its role in systems theory (Brunswick, Maine, 1984)*, volume 47 of *Contemp. Math.*, pages 465–478. Amer. Math. Soc., Providence, RI, 1985.

[VD81]   P. Van Dooren. The generalized eigenstructure problem in linear system theory. *IEEE Trans. Autom. Control*, 26:111–129, 1981.

[Wil72]   J. C. Willems. Dissipative dynamical systems. II. Linear systems with quadratic supply rates. *Arch. Rational Mech. Anal.*, 45:352–393, 1972.

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