Computing the distance to uncontrollability: the SISO case

Nicola Guglielmi and Ivan Markovsky

Abstract

In this paper, the problem of computing the distance from a given linear time-invariant system to the nearest uncontrollable system is posed and solved in the behavioral setting. In the case of a system with two external variables, the problem is restated as a Sylvester structured distance to singularity problem. The structured distance to singularity problem is then solved by integrating a system of ordinary differential equations which describes the gradient associated to the cost functional. An advantage of the method with respect to other approaches is in its capability to include further constraints. Numerical simulations also show that the method is more robust to the initial approximation than the Newton-type methods.

Index Terms

Sylvester matrix, structured pseudospectrum, structured low-rank approximation, ODEs on matrix manifolds, structured distance to singularity, distance to uncontrollability, behavioral approach.

I. Introduction

Consider a linear time-invariant system $\mathcal{B}$ with a state space representation

$$\begin{aligned} \mathcal{B} &= \mathcal{B}(A, B, C, D) \\ &:= \{\, w = (u, y) \mid \sigma x = Ax + Bu,\ y = Cx + Du \,\}, \end{aligned} \tag{1}$$

Nicola Guglielmi is with Dipartimento di Ingegneria Scienze Informatiche e Matematica, Università degli Studi di L' Aquila and Center of Excellence DEWS, Via Vetoio - Loc. Coppito, I-67010 L' Aquila, Italy. Email: nicola.guglielmi@univaq.it

Ivan Markovsky is with the department ELEC of the Vrije Universiteit Brussel, Building K, Pleinlaan 2, 1050 Brussels, Belgium. Email: imarkovs@vub.ac.be

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times \mathtt{m}}$, $C \in \mathbb{R}^{\mathtt{p}\times n}$, and $D \in \mathbb{R}^{\mathtt{p}\times \mathtt{m}}$ are parameters of $\mathcal{B}$; and σ is the shift operator (in discrete-time) or the derivative operator (in continuous-time)

$$(\sigma x)(t) = x(t+1) \qquad \text{or} \qquad \sigma x = dx/dt. \tag{2}$$

We adopt the behavioral setting [1], *i.e.*, a system is viewed as a set of trajectories (behavior). For a given system $\mathcal{B}$, the parameters A, B, and C of the state space representation (1) of $\mathcal{B}$ are not unique due to a change of basis $x' = Vx$ of the state space. For any nonsingular $n \times n$ matrix V, $\mathcal{B}(VAV^{-1}, VB, CV^{-1}, D)$ is the same model as $\mathcal{B}(A, B, C, D)$, *i.e.*,

$$\mathcal{B}(A, B, C, D) = \mathcal{B}(VAV^{-1}, VB, CV^{-1}, D).$$

In addition, the parameters A, B, and C are not unique due to nonminimality of the state dimension; for example

$$\mathcal{B}(A, B, C, D) = \mathcal{B}\left(\begin{pmatrix} A & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \begin{pmatrix} B \\ 0 \end{pmatrix}, \begin{pmatrix} C & 0 \end{pmatrix}, D\right), \tag{3}$$

for any $A_{12} \in \mathbb{R}^{n\times \Delta n}$, $A_{21} \in \mathbb{R}^{\Delta n\times n}$, and $A_{22} \in \mathbb{R}^{\Delta n\times \Delta n}$.

A state space representation with parameters A and B is *state controllable* if and only if the matrix

$$\mathcal{C}(A, B) := \begin{pmatrix} A & AB & \cdots & A^{n-1}B \end{pmatrix}$$

is full rank. Note that this classical notion of controllability is a property of the pair of matrices (A, B) and is not a property of a system $\mathcal{B} = \mathcal{B}(A, B, C, D)$ due to the nonuniqueness of a state space representation. The question of whether a given state space representation is state controllable is a rank test problem for the structured matrix $\mathcal{C}(A, B)$. A corresponding quantitative measure is the distance of $\mathcal{C}(A, B)$ to rank deficiency, *i.e.*, the smallest $(\Delta A, \Delta B)$, such that

$$\mathcal{C}(\widehat{A}, \widehat{B}) := \mathcal{C}(A, B) + \mathcal{C}(\Delta A, \Delta B)$$

is rank deficient.

Motivated by the issues of computing the numerical rank of a matrix, C. Paige defined in [2] the following measure for distance to uncontrollability

$$\begin{aligned} d_U(A, B) := \text{minimize} \quad & \text{over } \widehat{A},\ \widehat{B} \quad \left\| \begin{pmatrix} A & B \end{pmatrix} - \begin{pmatrix} \widehat{A} & \widehat{B} \end{pmatrix} \right\|_{\mathrm{F}} \\ \text{subject to} \quad & (\widehat{A}, \widehat{B}) \text{ is uncontrollable.} \end{aligned}$$

This problem falls into a broader category of *distance problems* [3], such as distance to instability, distance to positive definiteness, *etc*. There is a big volume of literature devoted on the problem of computing the distance to uncontrollability $d_U(A, B)$, see, *e.g.*, [4], [5], [6], [7], [8], [9], [10]. The measure $d_U(A, B)$, however, is not invariant of the state space representation because it depends on the choice of basis. This issue is resolved in the behavioral setting, where controllability is defined as a property of the system rather then a property of a particular representation.

Definition 1 (Definition V.1 in [11])**.** *A time-invariant dynamical system $\mathcal{B}$ is* controllable *if for any two trajectories $w_p, w_f \in \mathcal{B}$, there is a $\Delta t > 0$ and a trajectory $w_c \in \mathcal{B}$, such that $w_p(t) = w_c(t)$, for all $t < 0$, and $w_f(t) = w_f(t)$, for all $t \geq \Delta t$.*

Checking the controllability property in practice is done by performing a numerical test on the parameters of a specific representation of the system. For example, a linear time-invariant system with a kernel representation

$$\mathcal{B}(R) := \{\, w \mid R_0 w + R_1 \sigma w + \cdots + R_\ell \sigma^\ell w = 0 \,\} \tag{4}$$

is controllable if and only if the polynomial matrix

$$R(z) := R_0 + R_1 z + \cdots + R_\ell z^\ell \tag{5}$$

is left prime, *i.e.*, $R(z)$ is full row rank for all $z \in \mathbb{C}$.

Theorem 1 (Theorem 5.2.10 in [1])**.** *The system $\mathcal{B}$ defined by $R(\sigma) = 0$ is controllable if and only if the rank of the matrix $R(z)$ is the same for all $z \in \mathbb{C}$.*

In the case of two external variables,

$$R = \begin{pmatrix} q & -p \end{pmatrix},$$

with $p, q \in \mathbb{R}[z]$, assuming that $\det(p) \neq 0$ and $\deg(p) \geq \deg(q)$, the system $\mathcal{B} = \ker(R(\sigma))$ is represented by the familiar input/output representation

$$\mathcal{B} = \mathcal{B}(p, q) = \{\, w = (u, y) \;:\; p(\sigma)y = q(\sigma)u \,\}. \tag{6}$$

In this case, by Theorem 1, $\mathcal{B}$ is controllable if and only if p and q have no common factors of degree one or more.

Corollary 1 (Corollary 5.2.11 in [1])**.** *Consider the polynomials $p(z)$ and $q(z)$ and let the degree of p be higher than or equal to the degree of q. The single-input single-output (SISO) system defined by $p(\sigma)y = q(\sigma)u$ is controllable if and only if p and q are co-prime.*

Let $\overline{\mathcal{L}_{\mathrm{c}}}$ be the set of uncontrollable linear time-invariant systems (in the sense of Definition 1)

$$\overline{\mathcal{L}_{\mathrm{c}}} = \{\, \mathcal{B} \mid \mathcal{B} \text{ is LTI and uncontrollable} \,\}$$

and

$$\operatorname{dist}\big(\mathcal{B}(p,q), \mathcal{B}(\hat{p},\hat{q})\big) := \left\| \begin{pmatrix} q \\ p \end{pmatrix} - \begin{pmatrix} \hat{q} \\ \hat{p} \end{pmatrix} \right\|,$$

where $\|\cdot\|$ is the Euclidean norm. The notion of distance to uncontrollability considered in the paper is defined as follows.

Problem 1. *Given a controllable system $\mathcal{B}(p,q)$, find*

$$d_U(\mathcal{B}) := \min_{\widehat{\mathcal{B}} \in \overline{\mathcal{L}_{\mathrm{c}}}} \operatorname{dist}(\mathcal{B}, \widehat{\mathcal{B}}). \tag{7}$$

We refer to $d_U(\mathcal{B})$ as the *uncontrollability radius*.

Alternative approaches and contributions

Problem (7) is a nonconvex optimization problem and can be approached by global optimization, local optimization, and convex relaxation methods. The methods based on global optimization, such as the branch and bound method [12], are too expensive for most real-life problems. In this paper, we consider the local optimization approach [13]. Our main contribution is a new optimization method based on integration of a system of ordinary differential equations, which describes the gradient associated to the cost functional. The method is globally convergent to a locally optimal solution. Simulation results show that it is more robust to the initial approximation than the Newton-type methods. In addition, we incorporate the additional constraint of exactly known coefficients of the polynomials p and q into the method.

Alternative methods for solving problem (7) based on local optimization are developed in the structured low-rank approximation setting [14]. In particular, the method of [15] using kernel representation of the rank constraint and variable projections as well as the method of [16] using image representation of the rank constraint and homotopy can be used to compute locally optimal

solution of problem (7). In all numerical examples shown in Section VI, the proposed method finds a solution with the same or smaller value of the cost function than the one found by the method of [15]. In this sense, the proposed method is more robust to poor initial approximation.

Notation

- σ — shift or derivative operator (2)
- $\mathcal{L}$ — set of linear time-invariant systems
- $\mathcal{L}_{\rm c}/\overline{\mathcal{L}_{\rm c}}$ — controllable/uncontrollable LTI systems
- $\mathcal{B}(A,B,C,D)$ — state space representation (1)
- $\mathcal{B}(R)$ — kernel representation (4)
- $\mathcal{B}(p,q)$ — SISO representation (6)
- $\mathcal{C}(A,B)$ — controllability matrix
- $d_r(A)$ — distance to *unstructured* rank-r matrices
- $d_U(A,B)$ — distance of $\mathcal{B}(A,B,C,D)$ to state uncontrollability
- $d_U\big(\mathcal{B}(p,q)\big)$ — distance of $\mathcal{B}(p,q)$ to uncontrollability
- $\mu(S)$ — inner spectral radius of S
- $\Lambda_\varepsilon^{\mathcal{S}}(S)$ — structured ε-pseudospectrum
- $\mu_\varepsilon(S)$ — inner ε-pseudospectral radius of S
- $\Lambda(S)$ — spectrum of S
- $\mathrm{Syl}(p,q)$ — Sylvester matrix (9)
- $\mathcal{S}$ — set of Sylvester structured matrices
- $\|\cdot\|_F$ — Frobenius norm
- $\mathbf{Id}$ — identity matrix
- $\mathbf{1} = (1\ 1\ \ldots\ 1)^{\rm T}$
- $\langle A,B\rangle = \mathrm{trace}(A^*B)$ — Frobenius inner product

II. Preliminaries

Consider the polynomials

$$\begin{aligned} p(z) &= a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 \\ q(z) &= b_m z^m + b_{m-1}z^{m-1} + \cdots + b_1 z + b_0 \end{aligned} \tag{8}$$

with real coefficients $\{a_i\}$ and $\{b_i\}$ and with $m \leq n$. We often set $a_n = 1$ (a monic). Also, by setting $b_{m+1} = \cdots = b_n = 0$, we can consider the case $m = n$. The polynomials p and q are coprime if and only if the associated Sylvester matrix of dimension $2n \times 2n$,

$$S = \mathrm{Syl}(p,q) := \begin{pmatrix} a_n & a_{n-1} & \dots & a_1 & a_0 & 0 & \dots & 0 \\ 0 & a_n & a_{n-1} & \dots & a_1 & a_0 & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & a_n & a_{n-1} & \dots & a_1 & a_0 \\ 0 & b_m & \dots & b_1 & b_0 & 0 & \dots & 0 \\ 0 & 0 & b_m & \dots & b_1 & b_0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & b_m & \dots & b_1 & b_0 \end{pmatrix} \tag{9}$$

is nonsingular. Then, we have the following result.

Theorem 2. *A SISO system defined by $p(\sigma)y = q(\sigma)u$ is controllable if and only if the Sylvester matrix* $\mathrm{Syl}(p,q)$ *given by* (9) *is nonsingular.*

As a consequence problem (7) can be restated as

$$d_U\big(\mathcal{B}(p,q)\big) = \sup\{\, \varepsilon \;:\; (p+\delta p, q+\delta q) \text{ is controllable} \\ \forall\ \delta p \in \mathbb{R}^s,\ \delta q \in \mathbb{R}^{m+1}, \text{such that } \| \begin{pmatrix} \delta p^{\mathrm{T}} & \delta q^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}} \|_2 < \varepsilon\} \tag{10}$$

where $s = n$ if $p+\delta p$ is constrained to be monic and $s = n+1$ otherwise. We mainly consider two different distances to uncontrollability: one with p monic and one without this constraint. Another interesting case is when only a few coefficients of the polynomials are subject to perturbations; for an extension to this case see Section IV-E.

Remark II.1. Generically, the smallest perturbations which make a SISO system uncontrollable, creates either one real common root or a pair of complex conjugate common roots. It is well-known that in the first case the co-rank of the associated Sylvester matrix is one, while in the second case it is two.

III. STRUCTURED ε-PSEUDOSPECTRUM

The *inner spectral radius* of a matrix A is defined as

$$\mu(A) = \min\{|\lambda| \,|\, \lambda \in \Lambda(A)\},$$

where $\Lambda(A)$ is the spectrum of A. If $\mu(A) = 0$ then A is singular and therefore rank-deficient.

The basic observation used to calculate the distance to uncontrollability $d_U\big(\mathcal{B}(p,q)\big)$ is that (10) is equivalent to the following problem

$$d_U\big(\mathcal{B}(p,q)\big) = \frac{1}{\sqrt{n}} \inf \Big\{ \varepsilon \;:\; \mu\left(\mathrm{Syl}(p,q) + \varepsilon E\right) > 0, \\ \text{for all } E \in \mathcal{S},\ \|E\|_F \le 1 \Big\}, \quad (11)$$

where

$$\mathcal{S} = \{\mathrm{Syl}(p,q) \;:\; p \in \mathbb{R}^{n+1}, q \in \mathbb{R}^{m+1}\} \subset \mathbb{R}^{2n\times 2n} \quad (12)$$

is the set of real Sylvester matrices (see (9)).

Denote by

$$\Lambda_\varepsilon^{\mathcal{S}}(S) = \{\, \lambda \in \Lambda(S + \varepsilon E) \;:\; E \in \mathcal{S},\ \|E\|_F \le 1\}.$$

the structured ε-pseudospectrum (see [17]). Note that $\mathcal{S}$ is a smooth linear manifold which implies that

$$S + \varepsilon E \in \mathcal{S}, \qquad \text{if } E \in \mathcal{S}.$$

Example: Consider the two polynomials of degree 3,

$$\begin{aligned} p(z) &= z^3 + 2z^2 + 2z + 1 \\ q(z) &= 2z^3 + z - 2 \end{aligned}$$

where p is constrained to be monic. The corresponding Sylvester matrix is given by

$$\mathrm{Syl}(p,q) = \begin{pmatrix} 1 & 2 & 2 & 2 & 0 & 0 \\ 0 & 1 & 2 & 2 & 2 & 0 \\ 0 & 0 & 1 & 2 & 2 & 2 \\ 2 & 0 & 1 & -2 & 0 & 0 \\ 0 & 2 & 0 & 1 & -2 & 0 \\ 0 & 0 & 2 & 0 & 1 & -2 \end{pmatrix} \quad (13)$$

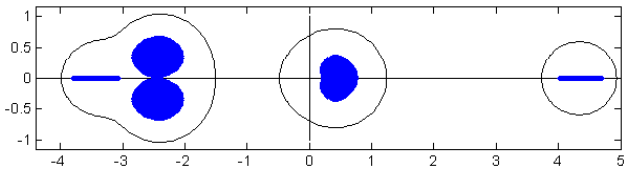

Fig. 1. The approximated structured ε-pseudospectrum for $\varepsilon = \frac{1}{2}$ for Example (13) is filled with blue; the boundary of the unstructured ε-pseudospectrum is plotted in black.

The set $\Lambda_\varepsilon^{\mathcal{S}}(S)$ for $\varepsilon = \frac{1}{2}$ is approximated by dense sampling on the set of admissible perturbations and is plotted in blue in Figure 1. The black curve represents the boundary of the corresponding unstructured ε-pseudospectrum, which means that arbitrary complex perturbations of norm bounded by $\frac{1}{2}$ are considered.

Next, we define $\mu_\varepsilon(S)$, the *inner ε-pseudospectral radius* of S, which is the minimum of the modulus of the elements of the structured ε-pseudospectrum (the case $\varepsilon = 0$, reduces $\mu_\varepsilon(S)$ to the inner spectral radius $\mu(S)$). This gives

$$\mu_\varepsilon(S) = \min\{|\lambda| : \lambda \in \Lambda_\varepsilon^{\mathcal{S}}(S)\}. \tag{14}$$

With this notation we characterize the distance to controllability as

$$d_U\big(\mathcal{B}(p,q)\big) = \frac{1}{\sqrt{n}} \arg\min_{\varepsilon>0}\{\mu_\varepsilon(S) = 0\}.$$

If S is associated to a controllable system, we have that

$$\mu_\varepsilon(S) > 0 \quad \Longleftrightarrow \quad d_U\big(\mathcal{B}(p,q)\big) > \varepsilon.$$

A. A 2-level methodology

In order to find the distance to uncontrollability we have to solve the equation (w.r.t. ε), $\mu_\varepsilon(S) = 0$. We propose a two-level algorithm: at the first level, for any fixed ε we compute a (local) minimizer of

$$\min\{|\lambda| : \lambda \in \Lambda_\varepsilon^{\mathcal{S}}(S)\}, \tag{15}$$

which we denote by $\lambda(\varepsilon) \in \partial\Lambda_\varepsilon^{\mathcal{S}}(S)$. If $\lambda(\varepsilon)$ is a global minimizer then $|\lambda(\varepsilon)| = \mu_\varepsilon(S)$.

The (inner) algorithm we propose finds local optima of problem (15) by determining the stationary point of a system of ODEs. In general there is no assurance that these are global

minimizers, although this seems to be the case in all our experiments of small dimension (where we performed a statistical investigation on a very large number of samples).

At the next level, we consider a continuous branch of minimizers and vary ε by an interpolation based iteration (outer algorithm) which exploits the knowledge of the exact derivative of $\lambda(\varepsilon)$ with respect to ε and exhibits superlinear convergence (a similar methodology has been exploited for different structures, see [18], and also for computing the H_∞ norm of a linear dynamical system [19] and the distance to instability of real matrices, see [20]).

Concerning the outer iteration, we indicate by

$$\lambda(\varepsilon) = \arg \min_{\lambda \in \Lambda_\varepsilon^{\mathcal{S}}(S)} |\lambda|$$

a continuous branch of minimizers (if $\lambda(\varepsilon)$ is not real then also its conjugate is a mimizer and we conventionally select the one with positive real part). Then we aim to compute

$$\varepsilon^\star = \arg \min_{\varepsilon > 0} \{\lambda(\varepsilon) = 0\}.$$

For ε in a left neighbourhood of $\varepsilon^\star$ we expect generically one of the following situations:

(i) There is a unique real minimizer $\lambda(\varepsilon)$. This means that there exists a matrix $E(\varepsilon) \in \mathcal{S}$ of unit norm such that $\lambda(\varepsilon)$ is a real simple eigenvalue of $S + \varepsilon E(\varepsilon)$. This implies that $S + \varepsilon^\star E(\varepsilon^\star)$ has co-rank equal to 1 and the two perturbed polynomials associated to $S + \varepsilon^\star E(\varepsilon^\star)$ have a common root.

(ii) There is a unique pair of complex conjugate minimizers $\lambda(\varepsilon)$ and $\overline{\lambda(\varepsilon)}$. This means that there exists a matrix $E(\varepsilon) \in \mathcal{S}$ of unit norm such that $\lambda(\varepsilon), \overline{\lambda(\varepsilon)}$ is a pair of complex conjugate eigenvalues of $S + \varepsilon E(\varepsilon)$. This implies that $S + \varepsilon^\star E(\varepsilon^\star)$ has co-rank equal to 2 and the two perturbed polynomials associated to $S + \varepsilon^\star E(\varepsilon^\star)$ have two complex conjugate common roots.

This means that—contrarily to the case of unstructured perturbations—we expect that as $\varepsilon \longrightarrow \varepsilon^\star$ we have to expect a *non-defective coalescence* of two complex conjugate eigenvalues in zero.

IV. Computation of the inner ε-pseudospectral radius

In order to approximate $\mu_\varepsilon(S)$, we construct a family of matrices $S + \varepsilon E(t)$ where $E(t) \in \mathcal{S}$ and $\|E(t)\|_F = 1$ such that $\lim_{t \to \infty} E(t) = E_\infty$ and an eigenvalue $\lambda(t)$ of $S + \varepsilon E_\infty$ is a point of $\Lambda_\varepsilon^{\mathcal{S}}(S)$ with locally minimum modulus.

We will use the convention that when an eigenvalue of minimum modulus $\lambda(t)$ is not real, which means it appears pairwise with $\overline{\lambda}(t)$, then we select the eigenvalue with positive imaginary part.

The derivative $\dot{E}(t)$ is chosen in the direction that gives the maximum possible decrease of $|\lambda(t)|$ for the minimum modulus eigenvalue $\lambda(t)$ of $S + \varepsilon E(t)$ along the manifold

$$\mathcal{S}_1 \;=\; \mathcal{S} \cap \{E \in \mathbb{R}^{2n\times 2n} \;:\; \|E\|_F = 1\}. \tag{16}$$

To fulfill the norm constraint $\|E(t)\|_F = 1$, we have to impose $\langle E, \dot{E}\rangle = 0$ where $\langle A, B\rangle = \mathrm{trace}(A^{\mathrm{T}}B)$ denotes the usual Frobenius inner product of the real matrices A and B.

The following result is necessary to the subsequent discussion (for Toeplitz matrices similar results are discussed in [21] and [22]).

Lemma 1. *Let $\mathcal{S} \subset \mathbb{R}^{2n\times 2n}$ be the manifold of Sylvester matrices of dimension $2n$ and $B \in \mathbb{C}^{2n\times 2n}$. The orthogonal projection (with respect to the Frobenius inner product $\langle\cdot,\cdot\rangle$) $P_{\mathcal{S}}(B)$ of B onto $\mathcal{S}$ is given by*

$$B_\perp = P_{\mathcal{S}}(B) \;=\; \mathrm{Syl}(\alpha, \beta) \tag{17}$$

where

$$\begin{aligned}
\alpha_{n-k} &= \frac{1}{n}\sum_{l=1}^{n} \mathrm{Re}\left(B_{l,l+k}\right), \quad k = k_0, \ldots, n \\
\beta_{m-k} &= \frac{1}{n}\sum_{l=1}^{n} \mathrm{Re}\left(B_{n+l,n-m+l+k}\right), \quad k = 0, \ldots, m
\end{aligned}$$

$k_0 = 1$ if p is constrained to be monic, and $k_0 = 0$ otherwise (and $\alpha_n = 0$).

Proof: We have to find $B_\perp \in \mathcal{S}$ such that

$$B_\perp = \arg\min_{S\in\mathcal{S}} \|B - S\|_F.$$

The result follows directly from the property that for a complex vector $x \in \mathbb{C}^n$,

$$\nu_* = \arg\min_{\nu\in\mathbb{R}} \|x - \nu\mathbf{1}\|_F = \frac{1}{n}\sum_{i=1}^{n} \mathrm{Re}(x_i),$$

being $\mathbf{1} = (1\ 1\ \ldots\ 1)^{\mathrm{T}}$. ■

Being $\mathcal{S}$ a manifold of real matrices, Note the obvious property $P_{\mathcal{S}}(B) = P_{\mathcal{S}}\left(\mathrm{Re}(B)\right)$.

A. Steepest descent direction

Our aim is to find an optimal variation Z of E such that $\|E+Z\|_F = 1$ and the eigenvalue of smallest modulus $|\lambda|$ of $S+\varepsilon(E+Z)$ is characterized (locally) by the maximal possible decrease. We follow an approach which extends to structured pseudospectra the ideas developed in [23], [24], [25] and use of the following standard perturbation result for eigenvalues see, e.g., [26, Section II.1.1]. Here and in the following, we denote $\dot{} = d/dt$.

Lemma 2. *Consider the differentiable matrix valued function $C(t)$ for t in a neighborhood of 0. Let $\lambda(t)$ be an eigenvalue of $C(t)$ converging to a simple eigenvalue λ_0 of $C_0 = C(0)$ as $t \to 0$. Let y_0 and x_0 be left and right eigenvectors, respectively, of C_0 corresponding to λ_0, that is, $(C_0 - \lambda_0 I)x_0 = 0$ and $y_0^*(C_0 - \lambda_0 I) = 0$. Then, $y_0^* x_0 \neq 0$ and $\lambda(t)$ is differentiable near $t = 0$ with*

$$\dot{\lambda}(0) = \frac{y_0^* \dot{C}(0) x_0}{y_0^* x_0}.$$

Next observe that for a simple eigenvalue $\lambda(t) = r(t)\mathrm{e}^{\mathbf{i}\theta(t)}$ ($r(t)$ denotes the modulus and $\theta(t)$ the phase) of the matrix-valued function $S + \varepsilon E(t)$, with associated left and right eigenvectors $y(t)$ and $x(t)$ respectively, we have (omitting the dependence on t)

$$\begin{aligned} \frac{d}{dt}|\lambda|^2 &= 2\,\mathrm{Re}(\overline{\lambda}\dot{\lambda}) = 2\,\mathrm{Re}\Big(\overline{\lambda}\,\varepsilon\,\frac{y^*\dot{E}x}{y^*x}\Big) \\ &= 2\varepsilon\,\mathrm{Re}\Big(\frac{(\lambda y)^*\dot{E}x}{y^*x}\Big) = 2\varepsilon r \mathrm{Re}\Big(\frac{y^*\dot{E}x}{\mathrm{e}^{\mathbf{i}\theta}y^*x}\Big). \end{aligned} \tag{18}$$

In the sequel of the paper we shall always impose the following scaling to the eigenvectors y and x,

$$\|y\| = \|x\| = 1, \qquad y^*x = |y^*x|\mathrm{e}^{-\mathbf{i}\theta} \tag{19}$$

which makes the denominator of (18) real and positive ($|y^*x| \neq 0$ since λ is assumed to be simple).

Let $\lambda = r\mathrm{e}^{\mathbf{i}\theta} \neq 0$ be the eigenvalue of minimum modulus of $S+\varepsilon E$. Then the optimal steepest descent direction for $|\lambda|^2$ (see (18) and (19)), with $Z = \dot{E} \in \mathcal{S}$, is given by:

$$\begin{aligned} Z_* = &\arg\min_{Z\in\mathcal{S}} \mathrm{Re}\,(y^*Zx) \\ &\text{subject to } \langle E, Z\rangle = 0 \text{ and } \|Z\|_F = 1. \end{aligned} \tag{20}$$

The solution to (20) is given in the following lemma.

Lemma 3. *Let $E \in \mathcal{S}$ be a $2n \times 2n$ real matrix of unit Frobenius norm, and $y, x \in \mathbb{C}^{2n}$ be non-zero complex vectors. Assume that $P_{\mathcal{S}}(yx^*) \neq 0$. Then the solution of the optimization problem* (20) *is given by*

$$\nu Z_* = -P_{\mathcal{S}}(yx^*) + \langle E, P_{\mathcal{S}}(yx^*)\rangle E \tag{21}$$

where $P_{\mathcal{S}}(B)$ is the orthogonal projection of B onto $\mathcal{S}$ and ν is the Frobenius norm of the matrix on the right hand side.

Proof: By the equality

$$\mathrm{Re}\,(y^* Z x) = \langle Z, \mathrm{Re}\,(yx^*)\rangle$$

and the fact that the inner product with a given vector is minimized over a subspace by orthogonally projecting the vector onto that subspace, the expression in (21) is the orthogonal projection of the complex matrix yx^* to the vector subspace $\mathcal{S}_1 \cap \{Z \ : \ \langle E, Z\rangle = 0\}$. ■

The following result assures that the condition $P_{\mathcal{S}}(yx^*) \neq 0$ considered in Lemma 3 is always fulfilled in the problem we analyze. In order to distinguish the case where p is unconstrained from the case where p is constrained to be monic, we introduce the set $\mathcal{S}^*$ which is the submanifold of Sylvester matrices (9) given by $a_n = 0$.

Lemma 4. *Let $S \in \mathcal{S}$ and either $E \in \mathcal{S}$ or $E \in \mathcal{S}^*$ of unit Frobenius norm, and $\varepsilon > 0$. If $\lambda \neq 0$ is a simple eigenvalue of $S + \varepsilon E$, with left and right eigenvectors y and x scaled according to* (19)*, then*

$$P_{\mathcal{S}}(yx^*) \neq 0. \tag{22}$$

Proof: We analyze first the case where p is not constrained to be monic.

Let y and x be the left and right eigenvectors of $S + \varepsilon E$ associated to $\lambda = r\mathrm{e}^{\mathrm{i}\theta}$. Assume — by contradiction — that $P_{\mathcal{S}}(yx^*) = 0$; this would imply

$$0 = \langle P_{\mathcal{S}}(yx^*), S + \varepsilon E\rangle = \langle yx^*, S + \varepsilon E\rangle$$
$$= \langle \mathrm{Re}(yx^*), S + \varepsilon E\rangle. \tag{23}$$

Observing that

$$\begin{aligned}\left\langle \text{Re}(yx^*), S+\varepsilon E\right\rangle &= \text{Re}\left\langle yx^*, S+\varepsilon E\right\rangle \\ &= \text{Re}\left(y^*\left(S+\varepsilon E\right)x\right) \\ &= \text{Re}\left(re^{\mathbf{i}\theta}y^*x\right),\end{aligned}$$

and exploiting the normalization (19), we obtain

$$\left\langle \text{Re}(yx^*), S+\varepsilon E\right\rangle = r|y^*x| > 0 \tag{24}$$

where positivity follows by the simplicity assumption for λ. This would contradict (23) and consequently (22) holds true.

Second we consider the case where p is constrained to be monic. If we assume that $P_{\mathcal{S}^*}\left(yx^*\right) = 0$, where the projection $P_{\mathcal{S}^*}$ — which is given by (17) by imposing $\alpha_n = 0$ — is relevant to the monic case and is used here to distinguish it from $P_{\mathcal{S}}$, we get

$$P_{\mathcal{S}}\left(yx^*\right) = \begin{pmatrix}\beta I & 0\\ 0 & 0\end{pmatrix} \tag{25}$$

where $P_{\mathcal{S}}$ is the usual projection on the manifold $\mathcal{S}$ to which belongs S (which now contains the submanifold $\mathcal{S}^*$ to which belongs E) and

$$\beta = \frac{1}{n}\sum_{i=1}^{n}\text{Re}\left(y_i\overline{x}_i\right).$$

Now, consider the matrix $C = S + \varepsilon E - \mathbf{Id}$, where $\mathbf{Id}$ is the identity matrix, and define the matrix

$$\widetilde{S} := \mathbf{Id} + \gamma C$$

which preserves the structure of S and also the eigenvectors x and y associated to the shifted eigenvalue λ.

First — by (25) — we obtain (recall that $a_n = 1$ in (9))

$$\left\langle \text{Re}(yx^*), \widetilde{S}\right\rangle = n\beta = \text{Re}\left(\sum_{i=1}^{n} y_i\overline{x}_i\right)$$

which has modulus smaller than 1.

Second, exploiting $\left\langle \text{Re}(yx^*), \widetilde{S}\right\rangle = \text{Re}\left(y^*\widetilde{S}x\right)$, we get

$$\left\langle \text{Re}(yx^*), \widetilde{S}\right\rangle = \widetilde{r}(\gamma) \tag{26}$$

where $|\widetilde{r}(\gamma)|$ can be chosen arbitrarily large if $|\gamma|$ is chosen large enough.

This leads to a contradiction. As a consequence we have that $P_{\mathcal{S}^*}(yx^*) \neq 0$. ■

B. The associated gradient system

Lemma 3 and formula (18) suggest to consider the following differential equation on the manifold $\mathcal{S}_1$ (see (16)),

$$\dot{E} = \Big(-P_{\mathcal{S}}\left(yx^*\right) + \big\langle E, P_{\mathcal{S}}\left(yx^*\right)\big\rangle E\Big)\,|\lambda| \tag{27}$$

where $y(t), x(t)$ are left and right eigenvectors of unit norm respectively to a simple eigenvalue $\lambda(t)$ of $S+\varepsilon E(t)$, and with $y^*x = |y^*x|\mathrm{e}^{-\mathbf{i}\theta}$, where ε is fixed. Observe that the multiplication by $|\lambda|$ in the right-hand side of (27) assures that E is an equilibrium of the ODE in the occurrence $\lambda = 0$.

We are in the position to prove the monotonic decrease of $|\lambda(t)|$ along every solution of (27).

Theorem 3. *Let $E(t)$ of unit Frobenius norm satisfy the differential equation* (27)*. If $\lambda(t)$ is a simple eigenvalue of $S + \varepsilon E(t)$, then*

$$\frac{d}{dt}|\lambda(t)| \;\leq\; 0. \tag{28}$$

Proof: Note that

$$\begin{aligned}
\mathrm{Re}\big(y^*P_{\mathcal{S}}\left(yx^*\right)x\big) &= \mathrm{Re}\big\langle yx^*, P_{\mathcal{S}}\left(yx^*\right)\big\rangle \\
&= \big\langle P_{\mathcal{S}}\left(yx^*\right), P_{\mathcal{S}}\left(yx^*\right)\big\rangle \\
&= \|P_{\mathcal{S}}\left(yx^*\right)\|_F^2,
\end{aligned}$$

and (since $E \in \mathcal{S}$)

$$\mathrm{Re}\left(y^*Ex\right) \;=\; \left\langle E, P_{\mathcal{S}}\left(yx^*\right)\right\rangle.$$

By the Cauchy–Schwarz inequality,

$$\begin{aligned}
|\langle E, P_{\mathcal{S}}\left(yx^*\right)\rangle| &\leq \|E\|_F\|P_{\mathcal{S}}\left(\mathrm{Re}yx^*\right)\|_F \\
&= \|P_{\mathcal{S}}\left(\mathrm{Re}yx^*\right)\|_F.
\end{aligned}$$

Finally, by (27),

$$\mathrm{Re}(y^*\dot{E}x) = \Big(-\|P_{\mathcal{S}}\left(yx^*\right)\|_F^2 + \langle E, P_{\mathcal{S}}\left(yx^*\right)\rangle^2\Big)\,|\lambda| \leq 0, \quad (29)$$

implying (28) by Lemma 2. ■

Remark IV.1. If at some $\bar{t}$ it holds $\lambda(\bar{t}) = 0$ then obviously $E(t)$ is a stationary point of the ODE (27). If $\lambda(t)$ is real for $t < \bar{t}$ then we generically expect it is a simple eigenvalue of $S + \varepsilon E(\bar{t})$. However if $\lambda(t)$ is not real for $t < \bar{t}$ then we would have that $\lambda(t)$ and its conjugate $\overline{\lambda}(t)$ coalesce in $z = 0$ at $t = \bar{t}$ and we expect generically that $\lambda = 0$ is a semi-simple double eigenvalue.

◇

Since we are interested to minimize $|\lambda|$ we address our attention to the stationary points of (27).

C. Stationary points

Since stationary points of (27) are potential minimizers for the computation of $\mu_\varepsilon(S)$, we give the following result for their characterization.

Theorem 4. *Assume that $\lambda \neq 0$. The following are equivalent on solutions of* (27)*:*

(1). $\dfrac{d}{dt}|\lambda| = 0$*;*

(2). $\dot{E} = 0$*;*

(3). E *is a real multiple of* $P_{\mathcal{S}}\left(yx^*\right)$.

Proof: The proof follows directly by applying Theorem 3 and Lemma 4. ■

The following result characterizes the local minimizers.

Theorem 5. *Let $E_* \in \mathcal{S}$ with $\|E_*\|_F = 1$. Let $\lambda_* = r\mathrm{e}^{\mathbf{i}\theta} \neq 0$ be a simple eigenvalue of $S + \varepsilon E_*$ with minimum modulus, with left and right eigenvectors y and x, respectively, both of unit norm and with the normalization $y^*x = |y^*x|\mathrm{e}^{-\mathbf{i}\theta}$. Then the following two statements are equivalent:*

(i) *Every differentiable path* $(E(t), \lambda(t))$ *(for small* $t \geq 0$*) such that* $\|E(t)\|_F \leq 1$ *and* $\lambda(t)$ *is an eigenvalue of* $S + \varepsilon E(t)$*, with* $E(0) = E_*$ *and* $\lambda(0) = \lambda_*$*, has*

$$\frac{d}{dt}|\lambda(t)|\Big|_{t=0} \geq 0.$$

(ii) E_* *is a negative multiple of* $P_{\mathcal{S}}(yx^*)$.

Proof: Assume that (i) does not hold true. Then there is some path $E(t)$ through E_* such that $\frac{d}{dt}|\lambda(t)|\big|_{t=0} < 0$; thus the minimization property established by Lemma 3 together with Lemma 2 shows that also the solution path of (27) passing through E_* is such a path. Hence E_* is not a stationary point of (27), and Theorem 4 then yields that E_* is not a real multiple of $P_{\mathcal{S}}(yx^*)$. This implies that also (ii) does not hold true.

Vice versa, if E_* is not a real multiple of $P_{\mathcal{S}}(yx^*)$, then E_* is not a stationary point of (27), and Theorems 4 and 3 yield that $\frac{d}{dt}|\lambda(t)|\big|_{t=0} < 0$ along the solution path of (27). Moreover, using a similar argument to [24, Theorem 2.2], if

$$E_* = \gamma P_{\mathcal{S}}(yx^*), \qquad \text{with} \quad \gamma > 0,$$

then along the path $E(t) = (1-t)E_*$, $t \in [0, 2]$, we have that

$$\mathrm{Re}(y^*\dot{E}(0)x) = -\gamma\|P_{\mathcal{S}}(yx^*)\|_F^2 < 0$$

and hence, by exploiting Lemma 2, $\frac{d}{dt}|\lambda(t)|\big|_{t=0} < 0$, which contradicts (i). ■

As a consequence, if in Theorem 4 $\lambda \neq 0$ is locally minimal (in modulus),

$$E = E_* = -P_{\mathcal{S}}(yx^*)/\|P_{\mathcal{S}}(yx^*)\|_F$$

that is the projection onto $\mathcal{S}$ of a real matrix of either rank 1 (if λ, x and y are real) or rank 2 (if λ is non real, and consequently also y and x).

D. The system of ODEs

We can write (27) in a compact form for the coefficients $\{\delta a_i\}$ and $\{\delta b_i\}$ of E,

$$E = \mathrm{Syl}(\delta a, \delta b) \tag{30}$$

that is

$$\begin{aligned} \dot{\delta a}_k &= (\alpha_k - \eta\,\delta a_k)\,|\lambda|, \qquad k = k_0, \ldots, n \\ \dot{\delta b}_k &= (\beta_k - \eta\,\delta b_k)\,|\lambda|, \qquad k = 0, \ldots, m, \end{aligned} \tag{31}$$

where $k_0 = 1$ if p is constrained to be monic, $k_0 = 0$ otherwise, α_k and β_k are the elements of $P_{\mathcal{S}}\,(yx^*)$ (see (17)) and

$$\eta \;\; = \;\; \big\langle E, P_{\mathcal{S}}\,(yx^*)\,\big\rangle.$$

This means we have to solve a system of $(m+1)(n+1-k_0)$ ordinary differential equations.

E. A natural extension to constrained systems

Assume that only certain subsets of the coefficients $\{a_i\}, \{b_j\}$ are allowed to be perturbed in order to find a close-by uncontrollable pair. Then the method has the same structure and only the projection changes. In fact, if $\{a_i\}$ does not vary for $i \notin \mathcal{I}$ and the same holds for $\{b_j\}$ for $j \notin \mathcal{J}$, where $\mathcal{I} \subseteq \{0, 1, \ldots, n\}$ and $\mathcal{J} \subseteq \{0, 1, \ldots, m\}$ are the sets of indeces corresponding to the coefficients of the polynomials which are allowed to be perturbed, we have simply to consider in (27) the new projection for $B \in \mathbb{C}^{2n\times 2n}$ is $P_{\mathcal{S}(\mathcal{I},\mathcal{J})}(B)$ given by (17) with

$$\alpha_{n-k} \;\; = \;\; \begin{cases} \dfrac{1}{n}\displaystyle\sum_{l=1}^{n} \mathrm{Re}\,(B_{l,l+k}) & k \in \mathcal{I} \\ 0 & k \notin \mathcal{I} \end{cases}$$

$$\beta_{m-k} \;\; = \;\; \begin{cases} \dfrac{1}{n}\displaystyle\sum_{l=1}^{n} \mathrm{Re}\,(B_{n+l,n-m+l+k}) & k \in \mathcal{J} \\ 0 & k \notin \mathcal{J}. \end{cases}$$

Note that the proof that the $P_{\mathcal{S}(\mathcal{I},\mathcal{J})}\,(yx^*) \neq 0$ is not obtained as a direct extension of Lemma 4.

The system of ODEs we have to solve is still (31) but now the number of ordinary differential equations is $|\mathcal{I}| \cdot |\mathcal{J}|$.

F. Numerical integration

Given $E_\ell \approx E(t_\ell)$ of unit Frobenius norm, and given y_ℓ and x_ℓ left and right eigenvectors of $S + \varepsilon E_\ell$ associated with its eigenvalue λ_ℓ of minimum modulus (if λ_ℓ is not real we choose $\lambda_\ell = r_\ell \mathrm{e}^{\mathbf{i}\theta_\ell}$ with positive imaginary part), with $y_\ell^* x_\ell = |y_\ell^* x_\ell| \mathrm{e}^{-\mathbf{i}\theta_\ell}$,

$$\alpha^{(\ell)} = \{\alpha_k^{(\ell)}\}_{k=k_0}^{n}, \qquad \beta^{(\ell)} = \{\beta_k^{(\ell)}\}_{k=0}^{m}$$

and

$$\delta a^{(\ell)} = \{\delta a_k^{(\ell)}\}_{k=k_0}^{n}, \qquad \delta b^{(\ell)} = \{\delta b_k^{(\ell)}\}_{k=0}^{m},$$

we determine all numerical approximations (see Algorithm 1) at time $t_{\ell+1} = t_\ell + h_\ell$ by applying a step of the Euler method with step-size h_ℓ to (27).

In order to control the step size we simply require that the monotonicity property of the exact flow, that is $|\lambda(t_{\ell+1})| < |\lambda(t_\ell)|$ is preserved by the numerical solution $|\lambda_{\ell+1}| < |\lambda_\ell|$. Since we are only interested in stationary points we can neglect the classical error control estimate on the solution, that is we do not estimate $\|E(t_{\ell+1}) - E_{\ell+1}\|$.

V. A SUPERLINEARLY CONVERGENT OUTER ITERATION FOR APPROXIMATING $d_U\big(\mathcal{B}(p,q)\big)$

In this section, we discuss the outer algorithm. In order to compute the distance to uncontrollability we should consider equation $\mu_\varepsilon(S) = 0$ and minimize its solution.

As a surrogate of this problem, which is of global optimization, we try to compute a value, say $\varepsilon^\star$, such that the boundary of the corresponding ε-pseudospectrum, $\partial\Lambda_{\varepsilon^\star}^{\mathcal{S}}(S)$, crosses the origin.

This would provide an upper bound for the distance; repeating such a search over different regions of the ε-pseudospectrum would increase the probability of computing the exact distance, and hence the robustness of the method.

In order to proceed we indicate by

$$\lambda(\varepsilon) = \arg \min_{\lambda \in \Lambda_\varepsilon^{\mathcal{S}}(S)} |\lambda|$$

a branch of (local) minimizers computed by determining the stationary point of the system of ODEs (27) (or equivalently (31)) which we denote by $E(\varepsilon)$. We make the following generic assumption.

Assumption V.1. *Let $\lambda(\varepsilon) \neq 0$ be a point of locally minimum modulus of $\Lambda_\varepsilon^{\mathcal{S}}(S)$ (with ε fixed), that is an eigenvalue with minimum modulus of the matrix $S + \varepsilon E(\varepsilon)$ (where $E(\varepsilon)$ denotes the corresponding (local) minimizer). Then $\lambda(\varepsilon)$ is simple.*

Moreover we assume that $E(\varepsilon)$ and $\lambda(\varepsilon)$ are smooth with respect to ε.

$\diamond$

Algorithm 1: Euler step applied to the ODEs (31) with step-size control

Data: $\alpha^{(\ell)}, \beta^{(\ell)}, \lambda_\ell, y_\ell, x_\ell$ and $\tilde{h}_\ell$ (step size predicted by the previous step).

Result: $E_{\ell+1}, y_{\ell+1}, x_{\ell+1}, \lambda_{\ell+1}$ and $\tilde{h}_{\ell+1}$.

begin

1 Set $h = \tilde{h}_\ell$.

2 Compute $Z_\ell = P_{\mathcal{S}}\left(y_\ell x_\ell^*\right) := S\left(\delta a^{(\ell)}, \delta b^{(\ell)}\right)$ and $\eta_\ell = \langle E_\ell, Z_\ell \rangle$.

3 Compute

$$\alpha_k^{(\ell+1)} = (1+h)\,\alpha_k^{(\ell)} - h\,\eta_\ell\,\delta a_k^{(\ell)}, \quad k = k_0, \ldots, n$$

$$\beta_k^{(\ell+1)} = (1+h)\,\beta_k^{(\ell)} - h\,\eta_\ell\,\delta b_k^{(\ell)}, \quad k = 0, \ldots, m.$$

4 Compute $\sigma_{\ell+1} = n \cdot \sqrt{\sum_{k=k_0}^{n} \left(\alpha_k^{(\ell+1)}\right)^2 + \sum_{k=0}^{m} \left(\beta_k^{(\ell+1)}\right)^2}$.

5 Normalize as

$$\alpha_k^{(\ell+1)} = \alpha_k^{(\ell+1)}/\sigma_{\ell+1}, \quad \beta_k^{(\ell+1)} = \beta_k^{(\ell+1)}/\sigma_{\ell+1}.$$

6 Set $E_{\ell+1} = S\left(\alpha^{(\ell+1)}, \beta^{(\ell+1)}\right)$.

7 Compute the eigenvalue of minimum modulus $\widehat{\lambda}$ of $S + \varepsilon E_{\ell+1}$, and the left and right eigenvectors $\widehat{y}$, $\widehat{x}$.

8 **if** $|\widehat{\lambda}| \geq |\lambda_\ell|$ **then**

reject the step, reduce the step size as $h := h/\gamma$ and repeat from 3;

else

accept the step: set $h_{\ell+1} = h$, $\lambda_{\ell+1} = \widehat{\lambda}$, $y_{\ell+1} = \widehat{y}$ and $x_{\ell+1} = \widehat{x}$.

9 **if** $h_{\ell+1} = \tilde{h}_\ell$ **then**

increase the step-size as $\tilde{h}_{\ell+1} := \gamma \tilde{h}_\ell$;

else

set $\tilde{h}_{\ell+1} = \tilde{h}_\ell$.

10 Proceed to next step

Since $\lambda(\varepsilon)$ can be computed by the inner algorithm, we may think to approach from the left-side the value $\varepsilon^\star$ such that $\lambda(\varepsilon^\star)=0$, by following a branch $\lambda(\varepsilon)\neq 0$ such that $\lim_{\varepsilon\nearrow\varepsilon^\star}\lambda(\varepsilon)=0$.

Assumption V.1 states that the eigenvalue $\lambda(\varepsilon)$ of minimum modulus of $S+\varepsilon E(\varepsilon)$ is a smooth function of ε in a left neighbourhood of $\varepsilon^\star$.

This property is exploited by the following result, which provides us an explicit and easily computable expression for the derivative of $|\lambda(\varepsilon)|$ (and thus also $\mu_\varepsilon(S)$) w.r.t. ε.

Theorem 6. *Assume the following:*

1) $\varepsilon\in(0,\varepsilon^\star)$ *such that* $\mu_\varepsilon(S)>0$,

2) $\lambda(\varepsilon)$ *be a branch of points of (locally) minimum modulus of* $\Lambda_\varepsilon^{\mathcal{S}}(S)$,

3) *Assumption V.1 holds, i.e.* $\lambda(\varepsilon)$ *and* $E(\varepsilon)$ *are smooth w.r.t.* ε,

and let $y(\varepsilon)$ *and* $x(\varepsilon)$ *be corresponding left and right eigenvectors of* $S+\varepsilon E(\varepsilon)$ *(where* $E(\varepsilon)$ *is a local minimizer), scaled according to* (19)*, with* $\|E(\varepsilon)\|_F=1$ *for all* ε*. Then*

$$\frac{d|\lambda(\varepsilon)|}{d\varepsilon} = -\frac{\|P_{\mathcal{S}}\left(y(\varepsilon)x(\varepsilon)^*\right)\|_F}{|y(\varepsilon)^*x(\varepsilon)|}<0, \quad \textit{for all } \varepsilon.$$

Proof: Note that

$$\begin{aligned}\frac{d}{d\varepsilon}|\lambda(\varepsilon)| &= \frac{1}{2|\lambda(\varepsilon)|}\frac{d}{d\varepsilon}|\lambda(\varepsilon)|^2\\ &= \frac{1}{|\lambda(\varepsilon)|}\mathrm{Re}\left(\overline{\lambda}(\varepsilon)\frac{d}{d\varepsilon}\lambda(\varepsilon)\right).\end{aligned} \tag{32}$$

Now use the derivative formula

$$\frac{d}{d\varepsilon}\lambda(\varepsilon)=\frac{y(\varepsilon)^*(E(\varepsilon)+\varepsilon E'(\varepsilon))x(\varepsilon)}{y(\varepsilon)^*x(\varepsilon)},$$

where $'=d/d\varepsilon$. Formula (32) yields

$$\frac{d}{d\varepsilon}|\lambda(\varepsilon)| = \frac{1}{|\lambda(\varepsilon)|}\mathrm{Re}\left(\overline{\lambda}(\varepsilon)\,\frac{y(\varepsilon)^*(E(\varepsilon)+\varepsilon E'(\varepsilon))x(\varepsilon)}{y(\varepsilon)^*x(\varepsilon)}\right)$$

$$= \frac{1}{|\lambda(\varepsilon)|}\mathrm{Re}\left(\frac{y(\varepsilon)^*(E(\varepsilon)+\varepsilon E'(\varepsilon))x(\varepsilon)}{|y(\varepsilon)^*x(\varepsilon)|\mathrm{e}^{-i\theta(\varepsilon)}}|\lambda(\varepsilon)|\mathrm{e}^{-i\theta(\varepsilon)}\right)$$

$$= \mathrm{Re}\left(\frac{\langle y(\varepsilon)x(\varepsilon)^*, E(\varepsilon)+\varepsilon E'(\varepsilon)\rangle}{|y(\varepsilon)^*x(\varepsilon)|}\right).$$

The main point is to prove that

$$\mathrm{Re}\left(y(\varepsilon)^*E'(\varepsilon)x(\varepsilon)\right) = 0. \tag{33}$$

The minimality property of the modulus of the eigenvalue $\lambda(\varepsilon)$ of $S+\varepsilon E(\varepsilon)$ yields Re $(y(\varepsilon)^* E'(\varepsilon)x(\varepsilon)) \geq 0$. Now suppose that for some ε_0, this inequality would actually be a strict inequality. Consider $\widetilde{E}(\varepsilon)$ of unit norm such that $\widetilde{E}(\varepsilon_0) = E(\varepsilon_0)$ and $\widetilde{E}'(\varepsilon_0) = -E'(\varepsilon_0)$. Then, for all ε sufficiently close to ε_0, we would have that the corresponding eigenvalue $\widetilde{\lambda}(\varepsilon)$ of $S + \varepsilon\widetilde{E}(\varepsilon)$ satisfies $|\widetilde{\lambda}(\varepsilon)| < |\lambda(\varepsilon)|$. This, however, contradicts the extremality of $E(\varepsilon)$ and hence (33) holds true.

Finally note that

$$\text{Re}\left(\frac{\langle y(\varepsilon)x(\varepsilon)^*, E(\varepsilon)\rangle}{|y(\varepsilon)^*x(\varepsilon)|}\right) = -\frac{\|P_{\mathcal{S}}\left(y(\varepsilon)x(\varepsilon)^*\right)\|_F}{|y(\varepsilon)^*x(\varepsilon)|}$$

which concludes the proof. ■

A. The numerical method

As a consequence the function $\varepsilon \to |\lambda(\varepsilon)|$ is smooth for $\varepsilon < \varepsilon^\star$ (where $|\lambda(\varepsilon)| > 0$); applying a Newton's iterate yields, for $\varepsilon_k < \varepsilon^\star$:

$$\varepsilon_{k+1} = \varepsilon_k - \left(\frac{\|P_{\mathcal{S}}\left(y(\varepsilon_k)x(\varepsilon_k)^*\right)\|_F}{|y(\varepsilon_k)^*x(\varepsilon_k)|}\right)^{-1} |\lambda(\varepsilon_k)| \tag{34}$$

where $\lambda(\varepsilon_k)$ is the rightmost eigenvalue of $S + \varepsilon_k E(\varepsilon_k)$, $E(\varepsilon_k)$ being the minimizer computed by the inner method, which integrates numerically the ODE (27). Likely the value ε_{k+1} will be closer to $\varepsilon^\star$ than ε_k but might lie on the right of $\varepsilon^\star$, where the function $|\lambda(\varepsilon)|$ is identically zero; hence it needs a correction to provide a lower bound to $\varepsilon^\star$. This would certainly occur when the function $\lambda(\varepsilon)$ is concave for $\varepsilon < \varepsilon^\star$ (as in the example illustrated in the following Figure 5).

An alternative, which allows to obtain a sequence of lower bounds which is more rapidly convergent to $\varepsilon^\star$, is that of interpolating pairs $(\varepsilon_{k-1}, |\lambda(\varepsilon_{k-1})|), (\varepsilon_k, |\lambda(\varepsilon_k)|)$ for values $\varepsilon_{k-1}, \varepsilon_k < \varepsilon^\star$, implying $|\lambda(\varepsilon_{k-1})|, |\lambda(\varepsilon_k)| > \text{tol}$, tol being a suitable tolerance. Setting $d_k(\varepsilon)$ the cubic Hermite polynomial, such that

$$\begin{aligned} d_k(\varepsilon_\ell) &= |\lambda(\varepsilon_\ell)|, & \ell = k-1, k \\ d_k'(\varepsilon_\ell) &= \frac{\|P_{\mathcal{S}}\left(y(\varepsilon_\ell)x(\varepsilon_\ell)^*\right)\|_F}{|y(\varepsilon_\ell)^*x(\varepsilon_\ell)|}, & \ell = k-1, k \end{aligned} \tag{35}$$

we define $\widehat{\varepsilon}_{k+1}$ as the solution of $d_k(\varepsilon) = 0$. Then, if $|\lambda(\widehat{\varepsilon}_{k+1})| > \text{tol}$ we set $\varepsilon_{k+1} = \widehat{\varepsilon}_{k+1}$, otherwise a bisection technique defines ε_{k+1}.

Algorithm 2: Basic algorithm for computing $\varepsilon^\star$

Data: $\mathrm{tol} > 0$ and $\varepsilon_0, \varepsilon_1, \varepsilon_u$ (such that $|\lambda(\varepsilon_0)| > |\lambda(\varepsilon_1)| > \mathrm{tol}$, and $|\lambda(\varepsilon_u)| < \mathrm{tol}$).

Result: ε_f (approximation of $\varepsilon^\star$).

begin

Set Reject = False and $k = 1$.

while $|\varepsilon_k - \varepsilon_u| \geq \mathrm{tol}$ **do**

  if $\mathrm{Reject} = \mathrm{False}$. **then**

    Store ε_k and $\lambda(\varepsilon_k)$ into the memory.

    Compute the polynomial $d_k(\varepsilon)$ (see (35)).

    Compute $\tilde{\varepsilon}_{k+1}$ the real root of $d_k(\varepsilon)$ closest to ε_k.

    if $\tilde{\varepsilon}_{k+1} > \varepsilon_u$ **then**

      Set $\tilde{\varepsilon}_{k+1} = (\varepsilon_u + \varepsilon_k)/2$.

  else

    Set $\tilde{\varepsilon}_{k+1} = (\varepsilon_u + \varepsilon_k)/2$.

  Compute $\lambda(\tilde{\varepsilon}_{k+1})$ by integrating (31) (equivalently (27)) with initial datum $E(\varepsilon_k)$ (i.e., the previously computed minimizer).

  if $|\lambda(\tilde{\varepsilon}_{k+1})| < \mathrm{tol}$ **then**

    Set $\mathrm{Reject} = \mathrm{True}$.

    Set $\varepsilon_u = \tilde{\varepsilon}_{k+1}$.

  else

    Set $\mathrm{Reject} = \mathrm{False}$.

    Set $\varepsilon_{k+1} = \tilde{\varepsilon}_{k+1}$.

    Order the array $\{\varepsilon_j\}_{j=0}^{k+1}$ in ascending order, $\varepsilon_{j+1} > \varepsilon_j$.

    Set $k = k + 1$.

Set $\varepsilon_f = \varepsilon_k$.

B. Approximation of $\varepsilon^\star$

We present now the algorithm for approximating $\varepsilon^\star$.

We make use of an upper bound ε_u which is located in the region where the function $\lambda(\varepsilon)$ is identically zero and construct a sequence $\{\varepsilon_k\}$ in the region where $|\lambda(\varepsilon_k)|$ is strictly monotonically decreasing, by successively finding zeros of the polynomials $d_k(\varepsilon)$, $k = 1, 2, \ldots$

A natural upper bound is $\varepsilon_u = \sqrt{n}\|p - q\|_F$, a lower bound $\varepsilon_0 = \sigma_{\min}(\mathrm{Syl}(p, q))$, where $\sigma_{\min}(\cdot)$ indicates the smallest singular value, i.e. the unconstrained distance to singularity of the Sylvester matrix $\mathrm{Syl}(p, q)$.

VI. Illustrative examples

In this section, we illustrate the performance of the proposed method on synthetic and real-life examples. Section VI-A illustrates the case when a simple eigenvalue vanishes. Section VI-B illustrates the case when a complex conjugated pair of eigenvalues coalescence at $z = 0$. Finally Section VI-C presents an application of the method on a real-life example of a weakly controllable mechanical system.

In all examples we made intense sampling of the parameter space in order to accurately approximate the structured ε-pseudospectrum so that we can state that what we compute is indeed the distance and not only an upper-bound. Hence the figures illustrate the effective behavior of $\mu_\varepsilon(S)$ as a function of ε and the first intersection to the horizontal axis provides the value $\varepsilon^\star$ which determines the distance $d_U(\mathcal{B}(p, q))$.

A. The real case: a simple eigenvalue vanishes

In the case where a real eigenvalue determines $\mu_\varepsilon(S)$ for $\varepsilon \to \varepsilon^\star$ we expect generically that $S + \varepsilon^\star E$ has a simple zero eigenvalue and hence has rank $2n - 1$. This is illustrated by the following examples.

1) Example: Consider the polynomials (8) of degree 5 with coefficients

$$\begin{array}{llllll} a_5 = 1 & a_4 = 0 & a_3 = 1 & a_2 = 0 & a_1 = 2 & a_0 = 1 \\ b_5 = -2 & b_4 = 1 & b_3 = 1 & b_2 = -1 & b_1 = 0 & b_0 = 1. \end{array}$$

Note that the p polynomial is monic. This property will be preserved in the approximation $\hat{p}$.

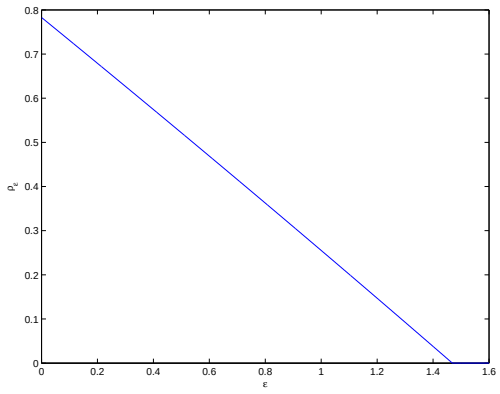

Fig. 2. The function $\varepsilon \to \mu_\varepsilon(S)$ for Example VI-A1.

The computed matrix $S + \varepsilon^\star E(\varepsilon^\star)$ has rank-$2n-1$ due to a simple zero eigenvalue. The perturbed polynomials' $\hat{p} = p + \delta p$, $\hat{q} = q + \delta q$ coefficients are shown (with five digit accuracy) in Table I. The common zero of $\hat{p}$, $\hat{q}$ is

TABLE I

COEFFICIENTS OF THE PERTURBED POLYNOMIALS $\hat{p} = p + \delta p$, $\hat{q} = q + \delta q$ IN THE EXAMPLE OF VI-A1.

$\hat{a}_5 = 1$	$\hat{a}_4 = 0.0144$	$\hat{a}_3 = 0.9729$	$\hat{a}_2 = 0.0510$	$\hat{a}_1 = 1.9039$	$\hat{a}_0 = 1.1811$
$\hat{b}_5 = -1.9778$	$\hat{b}_4 = 0.9583$	$\hat{b}_3 = 1.0787$	$\hat{b}_2 = -1.1483$	$\hat{b}_1 = 0.2795$	$\hat{b}_0 = 0.4732$

$$z_1 = -0.530278660.$$

The value $\varepsilon^\star$ and the estimated distance to uncontrollability are

$$\begin{aligned} \varepsilon^\star &= 1.468981057767730 \\ d_U\big(\mathcal{B}(p,q)\big) &= 0.656948300565638 \end{aligned}$$

The function $\varepsilon \mapsto \mu_\varepsilon(S)$ is shown in Figure 2.

2) Example: Consider again Example VI-A1 but now assume that the only coefficients that can be perturbed are a_0, a_2, a_4 and b_0, b_2 and b_4; this corresponds to setting $\mathcal{I} = \{0, 2, 4\}$ and $\mathcal{J} = \{0, 2, 4\}$ in the projection $P_{\mathcal{S}(\mathcal{I},\mathcal{J})}$ considered in Section IV-E.

The computed matrix $S+\varepsilon^\star E(\varepsilon^\star)$ turns out to have rank-$2n-1$ due to a simple zero eigenvalue. The coefficients of the perturbed polynomials $\hat{p} = p+\delta p$, $\hat{q} = q+\delta q$ are given in Table II. Their

TABLE II

COEFFICIENTS OF THE PERTURBED POLYNOMIALS $\hat{p} = p + \delta p$, $\hat{q} = q + \delta q$ IN THE EXAMPLE OF VI-A2.

$\hat{a}_5 = 0.9175$	$\hat{a}_4 = 0$	$\hat{a}_3 = 0.7629$	$\hat{a}_2 = 0$	$\hat{a}_1 = 1.3186$	$\hat{a}_0 = 1$
$\hat{b}_5 = -1.8715$	$\hat{b}_4 = 1$	$\hat{b}_3 = 1.3691$	$\hat{b}_2 = -1$	$\hat{b}_1 = 1.0607$	$\hat{b}_0 = 1$

common zero is

$$z = -0.5899110938.$$

The value $\varepsilon^\star$ and the estimated distance to uncontrollability are

$$\begin{aligned} \varepsilon^\star &= 3.004405111510952 \\ d_U\big(\mathcal{B}(p,q)\big) &= 1.343610812257265. \end{aligned}$$

3) Example: We consider the two polynomials of degree 9 with coefficients given in Table III. First, we consider the case where $\hat{p}$ is constrained to be monic.

TABLE III

COEFFICIENTS OF THE POLYNOMIALS p AND q IN THE EXAMPLE OF VI-A3.

$a_9 = 1$	$a_8 = 0$	$a_7 = 1$	$a_6 = 0$	$a_5 = 2$	$a_4 = 1$	$a_3 = 2$	$a_2 = 1$	$a_1 = 2$	$a_0 = 0$
$b_9 = 0$	$b_8 = 0$	$b_7 = 1$	$b_6 = 0$	$b_5 = 1$	$b_4 = 4$	$b_3 = 1$	$b_2 = 0$	$b_1 = 1$	$b_0 = -1.$

The computed matrix $S + \varepsilon^\star E(\varepsilon^\star)$ has rank-$2n-1$ due to a simple zero eigenvalue. The coefficients of the perturbed polynomials $\hat{p} = p + \delta p$, $\hat{q} = q + \delta q$ are given in Table IV. They

TABLE IV

COEFFICIENTS OF THE PERTURBED POLYNOMIALS $\hat{p}$ AND $\hat{q}$ IN THE EXAMPLE OF VI-A3.

$\hat{a}_9 = 1$	$\hat{a}_8 = 0.0000$	$\hat{a}_7 = 1.0000$	$\hat{a}_6 = 0.0000$	$\hat{a}_5 = 1.9999$
$\hat{a}_4 = 0.9995$	$\hat{a}_3 = 1.9971$	$\hat{a}_2 = 0.9850$	$\hat{a}_1 = 1.9209$	$\hat{a}_0 = -0.4163$
$\hat{b}_9 = 0.0000$	$\hat{b}_8 = 0.0000$	$\hat{b}_7 = 1.0000$	$\hat{b}_6 = 0.0000$	$\hat{b}_5 = 1.0002$
$\hat{b}_4 = 4.0010$	$\hat{b}_3 = 1.0053$	$\hat{b}_2 = 0.0278$	$\hat{b}_1 = 1.1461$	$\hat{b}_0 = -0.2313$

are not anymore coprime; their common zero equals to

$$z_1 = 0.0001901146.$$

The value $\varepsilon^\star$ and the estimated distance to uncontrollability are

$$\begin{aligned} \varepsilon^\star &= 2.671323516044883 \\ d_U\big(\mathcal{B}(p,q)\big) &= 0.890441086796000. \end{aligned}$$

B. The complex case: coalescence of two eigenvalues in $z = 0$

We consider here the case of coalescence of two complex conjugate eigenvalues in zero. As we expect the coalescence gives rise to a semi-simple double eigenvalue that determines a Sylvester matrix of co-rank equal to two, i.e. $\dim\ker\left(S + \varepsilon^\star E(\varepsilon^\star)\right) = 2$.

1) Example: We consider the two polynomials of degree 3 with coefficients

$$\begin{array}{llll} a_3 = 1 & a_2 = 2 & a_1 = 2 & a_0 = 2 \\ b_3 = 2 & b_2 = 0 & b_1 = 1 & b_0 = -2 \end{array}$$

where p is constrained to be monic.

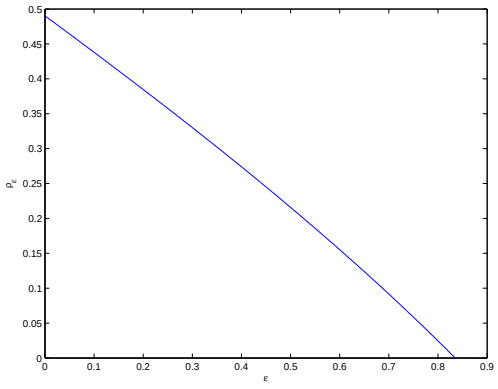

Fig. 3. The function $\varepsilon \to \mu_\varepsilon(S)$ for Example VI-B1.

The perturbed polynomials $\hat{p} = p + \delta p$, $\hat{q} = q + \delta q$ have coefficients

$$\begin{array}{llll} \hat{a}_3 = 1 & \hat{a}_2 = 2.1680 & \hat{a}_1 = 2.2569 & \hat{a}_0 = 1.6991 \\ \hat{b}_3 = 1.9637 & \hat{b}_2 = -0.1619 & \hat{b}_1 = 1.1315 & \hat{b}_0 = -1.9469. \end{array}$$

They are not anymore coprime; in fact they both have the common zeros equal to

$$z_{1,2} = -0.373421293 \pm 1.0276668040\mathbf{i}.$$

The value $\varepsilon^\star$ and the estimated distance to uncontrollability are

$$\begin{aligned} \varepsilon^\star &= 0.835047606282059 \\ d_U\big(\mathcal{B}(p,q)\big) &= 0.482114960273099. \end{aligned}$$

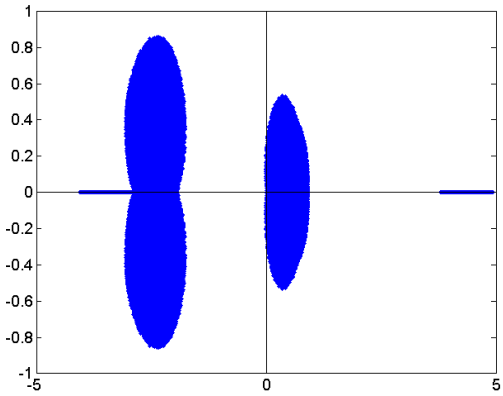

Fig. 4. Approximated structured ε-pseudospectrum for $\varepsilon = \varepsilon^\star$ for Example VI-B1. The origin lies on the boundary of $\Lambda_\varepsilon^{\mathcal{S}}(S)$

Figure 4 illustrates that the structured ε-pseudospectrum has the origin on its boundary, which implies that the compute value $\varepsilon^\star$ truly determines the distance to uncontrollability $d_U(\mathcal{B}(p,q))$.

The computed matrix $S + \varepsilon^\star E(\varepsilon^\star)$ turns out to have rank-$2n-2$. Its generalized null space decomposition (the generalized null space decomposition has been computed by the recent method presented in [27]) is

$$S + \varepsilon^\star E(\varepsilon^\star) = \mathrm{Syl}(\hat{p}, \hat{q}) = V\, Z\, V^{-1} \tag{36}$$

with Z and V given in Table V. This implies that the perturbed polynomials $\hat{p} = p + \delta p$, $\hat{q} = q + \delta q$ have a common pair of complex conjugate roots.

2) Example: Consider again Example VI-B1. However, now we assume that the only coefficients that can be perturbed are a_1, a_2, b_0 and b_2; this corresponds to setting $\mathcal{I} = \{1, 2\}$ and $\mathcal{J} = \{0, 2\}$ in the projection $P_{\mathcal{S}(\mathcal{I},\mathcal{J})}$ considered in Section IV-E.

We obtain the perturbed polynomials $\hat{p} = p + \delta p$, $\hat{q} = q + \delta q$ with coefficients

$$\begin{array}{llll} \hat{a}_3 = 1 & \hat{a}_2 = 2.4826 & \hat{a}_1 = 2.4336 & \hat{a}_0 = 2 \\ \hat{b}_3 = 2 & \hat{b}_2 = -0.2763 & \hat{b}_1 = 1 & \hat{b}_0 = -2.0081 \end{array}$$

TABLE V

THE Z AND V MATRICES IN (36).

$$Z = \begin{pmatrix} 0 & 0 & -0.6805 & -2.9805 & 0.4357 & 0.6448 \\ 0 & 0 & 1.9959 & -0.6704 & -0.9545 & -1.1616 \\ 0 & 0 & 0.0518 & 2.4565 & 3.2313 & 2.4130 \\ 0 & 0 & 1.2700 & -0.7316 & 2.5831 & 1.7297 \\ 0 & 0 & 0.3544 & 1.4105 & -0.7703 & 2.5334 \\ 0 & 0 & 1.4969 & 0.7212 & 1.8223 & -1.3906 \end{pmatrix}, \quad V = \begin{pmatrix} -0.4206 & 0.4970 & 0.7590 & 0 & 0 & 0 \\ -0.3831 & -0.5799 & 0.1674 & 0.6993 & 0 & 0 \\ 0.5911 & -0.0534 & 0.3625 & 0.1927 & 0.6922 & 0 \\ -0.0489 & 0.5184 & -0.3665 & 0.4909 & 0.1370 & 0.5786 \\ -0.4639 & -0.2792 & -0.0743 & -0.4679 & 0.5438 & 0.4321 \\ 0.3307 & -0.2592 & 0.3530 & -0.1183 & -0.4543 & 0.6917 \end{pmatrix}$$

having common zeros equal to

$$z_{1,2} = -0.3688968610 \pm 1.00507209\mathbf{97i}.$$

Naturally the distance to uncontrollability, which is computed as the solution of an optimization problem on a submanifold of $\mathcal{S}$ is now increased with respect to the previous case. We obtain

$$\begin{aligned} \varepsilon^\star &= 1.221408473917948 \\ d_U\big(\mathcal{B}(p,q)\big) &= 0.705180511207017 \end{aligned}$$

Interestingly, in this case, the code described in [28] fails to compute an approximation of $d_U\big(\mathcal{B}(p,q)\big)$.

3) Example: We consider the two polynomials of degree 9 with coefficients given in Table VI. p is constrained to be monic.

TABLE VI

COEFFICIENTS OF THE POLYNOMIALS p AND q IN THE EXAMPLE OF VI-B3.

$a_9 = 1$	$a_8 = 0$	$a_7 = 1$	$a_6 = 1$	$a_5 = 0$	$a_4 = 0$	$a_3 = 1$	$a_2 = 2$	$a_1 = 1$	$a_0 = 2$
$b_9 = 0$	$b_8 = 0$	$b_7 = -1$	$b_6 = 1$	$b_5 = -1$	$b_4 = 1$	$b_3 = 1$	$b_2 = 0$	$b_1 = 2$	$b_0 = 0$

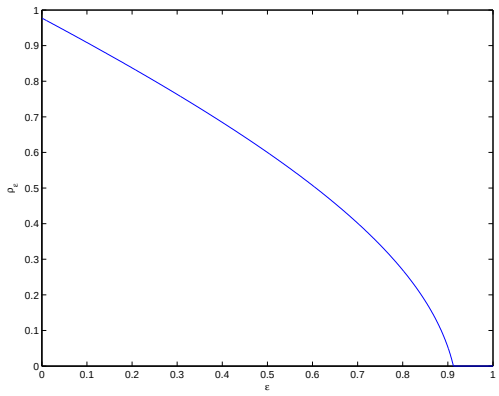

Fig. 5. The function $\varepsilon \to \mu_\varepsilon(S)$ for Example VI-B3. It appears that the left derivative of $\mu_\varepsilon(S)$ does not vanish at the (local) minimizer $\varepsilon = \varepsilon^\star = 0.91\ldots$ such that $\mu_\varepsilon(S) = 0$

The computed matrix $S + \varepsilon^\star E(\varepsilon^\star)$ turns out to have rank-$2n-2$ due to a semi-simple double zero eigenvalue.

This gives the perturbed polynomials $\hat{p} = p + \delta p$, $\hat{q} = q + \delta q$ with coefficients given in Table VII, having a common pair of complex conjugate roots. The common zeros of $\hat{p}$ and $\hat{q}$ equals

TABLE VII

COEFFICIENTS OF THE PERTURBED POLYNOMIALS $\hat{p}$ AND $\hat{q}$ IN THE EXAMPLE OF VI-B3.

$\hat{a}_9 = 1$	$\hat{a}_8 = -0.0166$	$\hat{a}_7 = 0.9775$	$\hat{a}_6 = 1.0007$	$\hat{a}_5 = 0.0131$
$\hat{a}_4 = 0.0048$	$\hat{a}_3 = 0.9943$	$\hat{a}_2 = 1.9951$	$\hat{a}_1 = 1.0013$	$\hat{a}_0 = 2.0033$
$\hat{b}_9 = -0.1182$	$\hat{b}_8 = 0.1949$	$\hat{b}_7 = -0.8561$	$\hat{b}_6 = 0.9449$	$\hat{b}_5 = -1.1039$
$\hat{b}_4 = 0.9908$	$\hat{b}_3 = 1.0558$	$\hat{b}_2 = 0.0271$	$\hat{b}_1 = 1.9787$	$\hat{b}_0 = -0.0239$

$$z_{1,2} = 0.338366068607 \pm 1.27830048225\mathbf{i}.$$

The value $\varepsilon^\star$ and the estimated distance to uncontrollability are

$$\begin{aligned} \varepsilon^\star &= 0.912373286589228 \\ d_U\big(\mathcal{B}(p,q)\big) &= 0.304124428863076. \end{aligned}$$

C. A real-life example

In this section, we apply the proposed method on the model-reduction benchmark of [29, Section 5.1] — a mechanical system consisting of N point masses connected in a chain by ideal

springs and ideal dampers. The input is the force applied on the first mass and the output is the position of the Nth mass. For fixed values of the other parameters, the longer the chain of masses is, *i.e.*, the bigger N is, the "more difficult" to control the system is. This physical intuition is verified by computing the distance to uncontrollability for increasing values of N, see Figure 6.

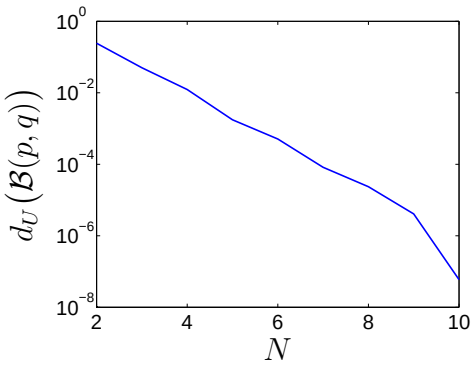

Fig. 6. Distance to uncontrollability as a function of the number of masses N.

VII. Discussion and outlook

We defined the distance to uncontrollability $d_U(\mathcal{B})$ for a linear time-invariant system $\mathcal{B}$ with two external variables in the behavioral setting. The main advantage of the "behavioral" formulation over the classical one $d_U(A, B)$, where A and B are state space parameters of the system, is that $d_U(\mathcal{B})$ is invariant of the model representation while $d_U(A, B)$ is not. Using the input output representation $\mathcal{B}(p, q)$ of the system $\mathcal{B}$, the problem of computing $d_U(\mathcal{B}(p, q))$ is equivalently expressed as a computation of an approximate common divisor of the polynomials p and q.

We presented a local optimization method for computing $d_U(\mathcal{B}(p, q))$ based on integration of a system of ordinary differential equations, which describes the gradient associated to the cost functional. The method allows specification of exactly known coefficients of the polynomials p and q. The presented numerical examples show the robustness of the method to the initial approximation.

The generalization of the results presented in the paper to the multi-input multi-output (MIMO) case, *i.e.*, considering the distance to uncontrollability problem (7) for the class of multivariable

linear time-invariant dynamical system, is a topic of future research. According to Theorem 1, the (exact) controllability property can be checked in the MIMO case by testing the left primness property of the polynomial matrix $R(z)$ (see (5)). This can be done by polynomial algebraic methods, see [30, Chapter 6]. The problem of computing the distance of $R(z)$ to the set of left prime matrices, however, is an open problem.

Another interesting modification of the problem considered in the paper is to replace the distance measure $\left\| \left[\begin{smallmatrix} q \\ p \end{smallmatrix} \right] - \left[\begin{smallmatrix} \hat{q} \\ \hat{p} \end{smallmatrix} \right] \right\|$ with $\angle(\mathcal{B}, \hat{\mathcal{B}})$, where $\angle$ is the gap metric [31].

ACKNOWLEDGMENTS

The first author thanks the Italian M.I.U.R. and the INdAM GNCS for financial support. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement number 258581 "Structured low-rank approximation: Theory, algorithms, and applications".

REFERENCES

[1] J. Polderman and J. C. Willems, *Introduction to Mathematical Systems Theory*. New York: Springer-Verlag, 1998.

[2] C. C. Paige, "Properties of numerical algorithms related to computing controllability," *IEEE Trans. Automat. Control*, vol. 26, pp. 130–138, 1981.

[3] N. Higham, "Matrix nearness problems and applications," in *Applications of Matrix Theory*, M. Gover and S. Barnett, Eds. Oxford University Press, 1989, pp. 1–27.

[4] D. Boley and W.-S. Lu, "Measuring how far a controllable system is from an uncontrollable one," *IEEE Trans. Automat. Control*, vol. 31, pp. 249–251, 1986.

[5] L. Eisner and C. He, "An algorithm for computing the distance to uncontrollability," *Control Lett.*, vol. 17, pp. 453–464, 1991.

[6] S. Khare, H. Pillai, and M. Belur, "Computing the radius of controllability for state space systems," *Control Lett.*, vol. 61, pp. 327–333, 2012.

[7] R. Eising, "Distance between controllable and uncontrollable," *Control Lett.*, vol. 4, pp. 263–264, 1984.

[8] M. Karow and D. Kressner, "On the structured distance to uncontrollability," *Control Lett.*, vol. 58, pp. 128–132, 2009.

[9] G. Hu and E. Davison, "Real controllablity/stabilizability radius of lti systems," *IEEE Trans. Automat. Control*, vol. 49, pp. 254–258, 2004.

[10] M. Gu, E. Mengi, M. Overton, J. Xia, and J. Zhu, "Fast methods for estimating the distance to uncontrollability," *SIAM J. Matrix Anal. Appl.*, vol. 28, pp. 477–502, 2006.

[11] J. C. Willems, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Trans. Automat. Control*, vol. 36, no. 3, pp. 259–294, 1991.

[12] V. Balakrishnan, S. Boyd, and S. Balemi, "Branch and bound algorithm for computing the minimum stability degree of parameter-dependent linear systems," *Int. J. of Robust and Nonlinear Control*, vol. 1, no. 4, pp. 295–317, October–December 1991.

[13] J. Nocedal and S. Wright, *Numerical optimization*. Springer-Verlag, 1999.

[14] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[15] I. Markovsky and K. Usevich, "Structured low-rank approximation with missing data," *SIAM J. Matrix Anal. Appl.*, pp. 814–830, 2013.

[16] M. Ishteva, K. Usevich, and I. Markovsky, "Factorization approach to structured low-rank approximation with applications," *SIAM J. Matrix Anal. Appl.*, 2014.

[17] L. N. Trefethen and M. Embree, *Spectra and pseudospectra: The behavior of nonnormal matrices and operators*. Princeton University Press, 1999.

[18] M. Manetta, "Computation of pseudospectral matrix distances and stability measures," *PhD thesis*, University of L'Aquila, 2014.

[19] N. Guglielmi, M. Gürbüzbalaban and M. Overton, "Fast approximation of the H_∞ norm via optimization over spectral value sets," *SIAM J. Matrix Anal. Appl.*, vol. 34, pp. 709–737, 2013.

[20] N. Guglielmi and M. Manetta, "Approximating real stability radii," *IMA J. Numer. Anal.*, pp. 1–24, 2014. [Online]. Available: http://imajna.oxfordjournals.org/content/early/2014/08/24/imanum. dru038.short?rss=1

[21] S. Noschese and L. Pasquini, "Eigenvalue patterned condition numbers: Toeplitz and hankel cases," *J. Comput. Appl. Math.*, vol. 206, pp. 615–624, 2007.

[22] P. Buttà, N. Guglielmi, and S. Noschese, "Computing the structured pseudospectrum of a toeplitz matrix and its extreme points," *SIAM J. Matrix Anal. Appl.*, vol. 33, pp. 300–319, 2012.

[23] N. Guglielmi and C. Lubich, "Differential equations for roaming pseudospectra: paths to extremal points and boundary tracking," *SIAM J. Numer. Anal.*, vol. 49, pp. 1194–1209, 2011.

[24] ——, "Low-rank dynamics for computing extremal points of real pseudospectra," *SIAM J. Matrix Anal. Appl.*, vol. 34, pp. 40–66, 2013.

[25] N. Guglielmi and M. Overton, "Fast algorithms for the approximation of the pseudospectral abscissa and pseudospectral radius of a matrix," *SIAM J. Matrix Anal. Appl.*, vol. 32, pp. 1166–1192, 2011.

[26] T. Kato, *Perturbation theory for linear operators*. Springer-Verlag, 1995.

[27] N. Guglielmi, M. Overton, and G. Stewart, "An efficient algorithm for computing the generalized null space decomposition," *SIAM J. Matrix Anal. Appl.*, vol. Submitted, 2014.

[28] J. Burke, M. Overton, and A. Lewis, "Pseudospectral components and distance to controllability," *SIAM J. Matrix Anal. Appl.*, vol. 26, pp. 350–361, 2004.

[29] I. Markovsky, "A software package for system identification in the behavioral setting," *Control Engineering Practice*, vol. 21, pp. 1422–1436, 2013.

[30] T. Kailath, *Linear Systems*. Prentice Hall, 1981.

[31] T. Georgiou and M. Smith, "Optimal robustness in the gap metric," *IEEE Transactions on Automatic Control*, vol. 35, no. 6, pp. 673–686, 1990.