SIAM J. MATRIX ANAL. APPL.
Vol. 30, No. 1, pp. 154–172

ON THE ESTIMATION OF THE DISTANCE TO UNCONTROLLABILITY FOR HIGHER ORDER SYSTEMS*

EMRE MENGI†

Abstract. A higher order dynamical system of order k is called controllable if the trajectory of the system as well as its first $k-1$ derivatives can be adjusted to pass through any given point at a finite time by choosing the input appropriately. The distance to uncontrollability is the norm of the smallest perturbation yielding an uncontrollable system. We derive a singular value minimization characterization for the distance to uncontrollability and present a trisection algorithm exploiting the singular value characterization. The algorithm is devised for low accuracy and depends on the extraction of the imaginary eigenvalues of even-odd matrix polynomials of degree $2k$ and size $2n$ with n denoting the size of the system. The well-studied first order distance to uncontrollability can be recovered as a special case.

Key words. matrix polynomials, dynamical systems, distance to uncontrollability, even-odd polynomials

AMS subject classifications. 65F15, 65K05, 93B05, 93B18

DOI. 10.1137/060658588

1. Introduction. A fundamental question concerning the kth order continuous time-invariant dynamical system

$$K_k x^{(k)}(t) + \cdots + K_1 x'(t) + K_0 x(t) = Bu(t), \quad x(0) = x'(0) = \cdots = x^{(k-1)} = 0 \tag{1.1}$$

is the dimension of the subspace of reachable configurations at a given time t' where $B \in \mathbb{C}^{n\times m}$, $K_0, K_1, \ldots K_k \in \mathbb{C}^{n\times n}$, $x(t) \in \mathbb{C}^n$, and $u(t) \in \mathbb{C}^m$. Here $x(t)$ denotes the state vector, $u(t)$ denotes the control input, and $c_0, c_1, \ldots, c_{k-1} \in \mathbb{C}^n$ are given initial conditions. By a configuration at time t' we mean the vector consisting of $x(t')$ as well as its first $k-1$ time derivatives at time t'. We define the space of reachable configurations at time t' as

$$\mathcal{R}_{t'} = \{[\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{k-1}] \ : \quad \exists u(t) \text{ such that (1.1) satisfies } \varepsilon_0 = x(t'),\ \varepsilon_1 = x'(t'), \ldots,\ \varepsilon_{k-1} = x^{(k-1)}(t')\}.$$

We have full control over the system (1.1) if all of the configurations can be attained by choosing $u(t)$ appropriately, that is

$$\dim(\mathcal{R}_{t'}) = nk. \tag{1.2}$$

In this case the system (1.1) is called controllable. Otherwise, the system is called uncontrollable. For convenience we will frequently refer to the tuple of matrices $(K_k, \ldots, K_1, K_0, B)$ as controllable whenever the system (1.1) is controllable.

Controllability of a first order system, specifically with $k = 1$, $K_1 = I$ (the identity matrix) and $K_0 = -A$ (an arbitrary matrix), is well known [9] to be equivalent to

*Received by the editors May 1, 2006; accepted for publication (in revised form) by D. Boley, August 13, 2007; published electronically February 20, 2008. This work was supported in part by the National Science Foundation grant DMS-0412049.
http://www.siam.org/journals/simax/30-1/65858.html
†Department of Mathematics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093 (emengi@math.ucsd.edu).

either of the conditions

$$\operatorname{rank}([B \;\; AB \;\; A^2B \;\; \cdots \;\; A^{n-1}B]) = n$$

or

$$\operatorname{rank}([A - \lambda I \;\; B]) = n \text{ for all } \lambda \in \mathbb{C}. \tag{1.3}$$

A similar characterization for the controllability of a descriptor system with $k = 1$, $K_1 = E$, and $K_0 = -A$ exists [7, 8]. In particular when E is nonsingular the controllability reduces to the condition

$$\operatorname{rank}([A - \lambda E \;\; B]) = n \text{ for all } \lambda \in \mathbb{C}. \tag{1.4}$$

When E is singular, the above condition needs to be accompanied by an additional rank condition that involves the null space of E. Throughout this paper we will assume that the leading coefficient is nonsingular and additionally, when perturbations to the leading coefficient are allowed, the leading coefficient remains nonsingular under all perturbations under consideration. (This condition is stated formally in Lemma 2.2.) Under this nonsingularity assumption, the rank characterizations (1.3) and (1.4) can be generalized to the higher order system and the nearby systems as follows. First observe that (1.1) can be embedded into the first order system

$$\tilde{x}'(t) = \mathcal{A}\tilde{x}(t) + \mathcal{B}u(t), \qquad \tilde{x}(0) = \begin{bmatrix} c_{k-1} \\ c_{k-2} \\ \vdots \\ c_0 \end{bmatrix}, \tag{1.5}$$

where

$$\tilde{x}(t) = \begin{bmatrix} x^{(k-1)}(t) \\ x^{(k-2)}(t) \\ x^{(k-3)}(t) \\ \vdots \\ x(t) \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} K_k^{-1}B \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \text{and}$$

$$\mathcal{A} = \begin{bmatrix} -K_k^{-1}K_{k-1} & -K_k^{-1}K_{k-2} & \dots & -K_k^{-1}K_1 & -K_k^{-1}K_0 \\ I & 0 & & 0 & 0 \\ 0 & I & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & I & 0 \end{bmatrix}.$$

Now the higher order system is controllable if and only if the matrix $[\mathcal{A} - \lambda I \;\; \mathcal{B}]$ has full rank for all λ. Furthermore, for a given λ suppose

$$[\mathcal{A} - \lambda I \;\; \mathcal{B}] \begin{bmatrix} x_{k-1} \\ \vdots \\ x_0 \\ y_0 \end{bmatrix} = 0.$$

Using the definitions of $\mathcal{A}$ and $\mathcal{B}$, it is straightforward to deduce that $x_j = \lambda^j x_0$ and

$$[P(\lambda) \;\; B] \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = 0,$$

where

$$P(\lambda) = \sum_{j=0}^{k} \lambda^j K_j. \tag{1.6}$$

Therefore the null spaces of $[\mathcal{A} - \lambda I \;\; \mathcal{B}]$ and $[P(\lambda) \;\; B]$ have the same dimension, say $l \geq m$, which means $\text{rank}([\mathcal{A} - \lambda I \;\; \mathcal{B}]) = nk + m - l$ and $\text{rank}([P(\lambda) \;\; B]) = n + m - l$. We conclude that the controllability of the higher order system is equivalent to

$$\text{rank}([P(\lambda) \;\; B]) = n \text{ for all } \lambda \in \mathbb{C} \tag{1.7}$$

which was already mentioned in [18] without derivation.

Controllability is thus a rank determination problem, which cannot be performed reliably in the presence of rounding errors. A controllable system may still have nearby uncontrollable systems which potentially is an indicator of a problem with the model. Therefore in [22] for the first order system the distance to uncontrollability was defined as

$$\tau(A, B) = \inf\{\|[\Delta A \;\; \Delta B]\| : \text{the pair } (A + \Delta A, B + \Delta B) \text{ is uncontrollable}\} \tag{1.8}$$

with $\|\cdot\|$ denoting either the spectral norm or the Frobenius norm. Later Eising [10] proved that, in both cases, the distance to uncontrollability is equivalent to a minimization problem involving complex vectors of size n

$$\tau(A, B) = \inf_{q \in \mathbb{C}^n, \|q\|=1} \sqrt{q^* B B^* q + q^* A (I - qq^*) A^* q} \tag{1.9}$$

and a singular value minimization problem, i.e.,

$$\tau(A, B) = \inf_{\lambda \in \mathbb{C}} \sigma_{\min}([A - \lambda I \;\; B]), \tag{1.10}$$

where $\sigma_{\min}$ denotes the smallest singular value. The most efficient computational techniques for the distance to uncontrollability exploit the definition (1.10), though there are hybrid-algorithms [24] developed following Eising's characterizations that make use of both (1.9) and (1.10). Boley observed the connection between the sensitivity of the Kronecker structure of a matrix pencil and distance to uncontrollability and based on (1.10) suggested a practical but an imprecise way to approximate the distance by solving a standard eigenvalue problem [1]. Byers introduced classes of algorithms working on one dimensional or two dimensional grids [5] to minimize $\sigma_{\min}([A - \lambda I \;\; B])$. Later Gao and Neumann [11] and He [16] modified Byers' idea for more efficient computation. Byers' grid-based algorithms and its successors are well-suited for the computation of the distance to uncontrollability with a few digits of precision but are too costly for high accuracy. Gu's bisection algorithm [14] is the first technique that retrieves the global minimum for the problem (1.10) within a factor of two without depending on a grid. Gu's algorithm later was improved by Burke, Lewis, and Overton [3] who suggested a trisection algorithm that computes $\tau(A, B)$ to arbitrary precision. With $O(n^6)$ complexity[1] these algorithms are applicable only to small systems. In [15], it is described how we can benefit from inverse iteration and shift-and-invert preconditioned Arnoldi to reduce the average running time to $O(n^4)$

[1] When we refer to operation counts, we assume eigenvalue computations are atomic operations with cubic complexity.

making the computation of the distance to uncontrollability for medium size systems feasible. For descriptor systems the distance to uncontrollability is discussed and a generalization of the characterization (1.10) is provided in [6].

In this work we extend the definition (1.8) for the first order system to the higher order system (1.1) as

(1.11)
$$\tau(P,B,\alpha) = \inf\{\|[\Delta K_k \ \cdots \ \Delta K_1 \ \Delta K_0 \ \Delta B]\| : \text{ the tuple } (K_k + \alpha_k \Delta K_k, \ldots, K_0 + \alpha_0 \Delta K_0, B + \Delta B) \text{ is uncontrollable}\},$$

where the vector $\alpha = [\alpha_k \ \cdots \ \alpha_1 \ \alpha_0]$ consists of nonnegative real numbers. Notice that with $k = 1$, $K_1 = I$, $K_0 = -A$, and $\alpha = [0\ 1]$ we recover the definition (1.8) for the first order system. Our motivation in introducing the scaling α is mainly to restrict the perturbations to some of the coefficient matrices, by choosing the scaling corresponding to other coefficients to be zero. It also serves the purpose of weighting the perturbations to the coefficients. For instance one may be interested in perturbations in a relative sense with respect to the norm of the coefficients in which case it is desirable to set $\alpha = [\|K_k\| \ \cdots \ \|K_1\| \ \|K_0\|]$.

The distance to uncontrollability of the higher order system defined by (1.11) and the embedded system (1.5) are related yet different quantities. The closest uncontrollable descriptor system to the embedded system would usually be obtained by perturbing the block rows of $\mathcal{A}$ other than the first one, so the resulting uncontrollable system does not correspond to an embedding of a higher order system. For instance if one of the coefficient matrices, say K_j, is considerably larger than the other coefficients as well as B in norm and $K_k^{-1}K_j$ is close to a multiple of the identity matrix, then small perturbations to the $(j+1)$th block row of $[\mathcal{A} \ \mathcal{B}]$ makes it rank deficient and the embedding uncontrollable. Typically we expect that $\tau(\mathcal{A},\mathcal{B}) < \tau(P,B,\alpha)$, since in the definition of $\tau(\mathcal{A},\mathcal{B})$ we have more degrees of freedom when choosing perturbations. Such an example where these two distances differ significantly is given in section 4.2. It is not clear how the existing algorithms to compute $\tau(\mathcal{A},\mathcal{B})$ can be modified to impose the constraints on perturbations to $\mathcal{A}$ and $\mathcal{B}$ so that perturbed systems correspond to the embeddings of higher order systems.

In the next section we provide a singular value minimization characterization for the definition (1.11). We will see that the definition (1.11) in the spectral norm and the Frobenius norm are equivalent just as in the first order case and the characterization we derive reduces to (1.10) for the first order system. The derivation of the singular value characterization uses the rank definition of the controllability (1.7) for the higher order system and all nearby systems which holds only if the leading coefficient is nonsingular and sufficiently away from the closest singular matrix. The equivalent singular value characterization is typically nonconvex. A standard optimization technique such as BFGS will converge only to a local minimum. Applying BFGS repeatedly with various starting points might occasionally fail to return a global minimum. Therefore in section 3 we describe a trisection algorithm locating the global minimum of the equivalent optimization problem. This algorithm is not a generalization of the algorithm of [3], because such an approach is too expensive. The first few steps of the new algorithm are comparatively cheap, but as we require more accuracy the algorithm becomes computationally intensive. With a complexity of $O\left(\frac{1}{\arccos(1-(\frac{tol}{k})^2)} n^3 k^4\right)$ with tol denoting the accuracy required, it is devised for a few digits of precision. Section 4 is devoted to numerical examples illustrating the efficiency of the algorithm.

2. Properties of the higher order distance to uncontrollability and a singular value characterization. The set of controllable tuples is clearly a dense subset of the whole space of matrix tuples. But this does not mean that the uncontrollable tuples are isolated points. On the contrary there are uncontrollable subspaces. For instance the system (1.1) with $K_0 = 0$ and $\text{rank}(B) < n$ is uncontrollable for all $K_k, \ldots, K_1$. Therefore we shall first see that $\tau(P, B, \alpha)$ is indeed attained at some $(\Delta K_k, \ldots, \Delta K_0, \Delta B)$. Note that throughout this work we usually use $\|\cdot\|$ for either the spectral or the Frobenius norm interchangeably when the results hold for both of the norms or when the type of the norm is clear from the context. At other times we clarify the choice of norm using the notation $\|\cdot\|_2, \|\cdot\|_F$ for the spectral and the Frobenius norm, respectively.

LEMMA 2.1. *There exists an uncontrollable tuple* $(K_k+\alpha_k\Delta K_k, \ldots, K_0+\alpha_0\Delta K_0, B + \Delta B)$ *such that* $\tau(P, B, \alpha) = \|[\Delta K_k \ \cdots \ \Delta K_0 \ \Delta B]\|$ *and* $\|\Delta K_j\| \leq \|B\|$ *for all* j, $\|\Delta B\| \leq \|B\|$.

Proof. The matrix $[P(\lambda) \ 0]$ is rank deficient at the eigenvalues of P. Therefore $\tau(P, B, \alpha) \leq \|B\|$ meaning we can restrict the perturbations to the ones satisfying $\|\Delta K_j\| \leq \|B\|$ and $\|\Delta B\| \leq \|B\|$.

Furthermore the set of uncontrollable tuples is closed. To see this, consider any sequence $\{(K'_k, \ldots, K'_0, B')\}$ of uncontrollable tuples. Now for any tuple in the sequence define the associated polynomial as $P'(\lambda) = \sum_{j=0}^{k} \lambda^j K'_j$. The matrix $[P'(\lambda) \ B']$ is rank deficient for some λ, so all combinations of n columns of this matrix are linearly dependent. Let us denote the $l = \binom{m+n}{n}$ polynomials associated with the determinants of the combinations of n columns by $p_1(\lambda), p_2(\lambda), \ldots, p_l(\lambda)$ in any order. These polynomials must share a common root; otherwise $[P'(\lambda) \ B']$ would not be rank deficient for some λ. The common roots $r_1, r_2, \ldots, r_l$ are continuous functions of the tuple $\{(K'_k, \ldots, K'_0, B')\}$ which means at any cluster point of the sequence $r_1 = r_2 = r_3 = \cdots = r_l$. This shows that the set is closed.

Since we are minimizing the spectral or the Frobenius norm over a compact set, $\tau(P, B, \alpha)$ must be attained at some $\|[\Delta K_k \ \cdots \ \Delta K_0 \ \Delta B]\|$. ∎

The main result of this section establishes the equivalence of $\tau(P, B, \alpha)$ to the solution of the singular value minimization problem

$$\xi(P, B, \alpha) = \inf_{\lambda \in \mathbb{C}} \sigma_{\min} \left(\begin{bmatrix} P(\lambda) \\ \hline \sqrt{s_\alpha(|\lambda|)} \end{bmatrix} \ B \right) \tag{2.1}$$

when $\alpha_0 \neq 0$, where

$$s_\alpha(|\lambda|) = \sum_{j=0}^{k} \alpha_j^2 |\lambda|^{2j}.$$

When establishing this equivalence, we seek the perturbations ΔP and ΔB yielding a matrix function $[(P + \Delta P)(\lambda) \ B + \Delta B]$ that is rank deficient at some λ. A relevant problem is the distance to instability of a matrix polynomial which can be posed as

$$\beta(P, \alpha)$$
$$= \inf \left\{ \|[\Delta K_k \ \Delta K_{k-1} \ldots \Delta K_0]\| : (P + \Delta P)(\lambda) = 0, \ \exists \lambda \in \mathbb{C}_b, \ \Delta P = \sum_{j=0}^{k} \alpha_j \lambda^j \Delta K_j \right\}$$

where $\mathbb{C}_b$ is a closed subset of the complex plane corresponding to the unstable region and $\|\cdot\|$ is the spectral norm. A simplified version of this problem with α equal to

the vector of ones was studied in [13]. Let $\partial\mathbb{C}_b$ denote the boundary of the unstable region. It is straightforward to modify Lemma 8 in [13] to deduce the equivalence of $\beta(P,\alpha)$ with the minimization problem

$$\inf_{\lambda\in\partial\mathbb{C}_b} \sigma_{\min}\left(\left[\frac{P(\lambda)}{\sqrt{s_\alpha(|\lambda|)}}\right]\right).$$

Another similar problem is the pseudospectrum of a matrix polynomial which consists of the set of eigenvalues of nearby matrix polynomials. Let us formally define the ϵ-pseudospectrum as

$$\Lambda_\epsilon(P,\alpha)$$
$$= \left\{\lambda\in\mathbb{C} : (P+\Delta P)(\lambda)=0,\ \Delta P=\sum_{j=0}^{k}\alpha_j\lambda^j\Delta K_j,\ \|[\Delta K_k\ \Delta K_{k-1}\ \cdots\ \Delta K_0]\|\le\epsilon\right\}$$

where $\|\cdot\|$ denotes the spectral norm. Here we slightly depart from the original definition suggested by Tisseur and Higham in [23] in the way the nearness to a matrix polynomial is measured. (In [23] the norm of each of the perturbations ΔK_j is constrained to be less than ϵ.) The technique in [23] leads us to the singular value characterization

$$\Lambda_\epsilon(P,\alpha)=\left\{\lambda\in\mathbb{C} : \sigma_{\min}\left(\left[\frac{P(\lambda)}{\sqrt{s_\alpha(|\lambda|)}}\right]\right)\le\epsilon\right\}.$$

The condition $\alpha_0\neq 0$, that is assumed throughout the derivations below, means that the perturbations to K_0 cannot be blocked and avoids the indeterminate case, when $s_\alpha(|\lambda|)=0$. At the end of this section we will present a more general equivalence result that holds no matter what value is assigned to α as long as all of its components are nonnegative. With this restriction on α_0, $\xi(P,B,\alpha)$ must be attained either at a finite λ or at ∞. The latter case is eliminated by the next lemma.

LEMMA 2.2. *Under the assumption that the leading coefficient of* (1.1) *is nonsingular and remains nonsingular under perturbations with norm less than or equal to* $\alpha_k\xi(P,B,\alpha)$ *and* $\alpha_0\neq 0$, *the inequality*

$$\xi(P,B,\alpha) < \lim_{\lambda\to\infty}\sigma_{\min}\left(\left[\frac{P(\lambda)}{\sqrt{s_\alpha(|\lambda|)}}\quad B\right]\right)$$

holds.

Proof. When $\alpha_k=0$, the result immediately follows. When $\alpha_k>0$, we have

$$\sigma_{\min}\left(\left[\frac{K_k}{\alpha_k}\quad B\right]\right)=\lim_{\lambda\to\infty}\sigma_{\min}\left(\left[\frac{P(\lambda)}{\sqrt{s_\alpha(|\lambda|)}}\quad B\right]\right).$$

Suppose $\xi(P,B,\alpha)$ is attained at ∞ and therefore there exist $u_1,v\in\mathbb{C}^n$ and $u_2\in\mathbb{C}^m$ such that

$$\begin{bmatrix}\left(\frac{K_k}{\alpha_k}\right)^*\\ B^*\end{bmatrix}v=\xi(P,B,\alpha)\begin{bmatrix}u_1\\ u_2\end{bmatrix},$$

where $[u_1^T \ u_2^T]^T$ and v have unit length. Multiplying the upper blocks by α_k, the right-hand side by v^*v and collecting all terms on the left yields

$$\begin{bmatrix} K_k^* - \alpha_k \xi(P,B,\alpha) u_1 v^* \\ B^* - \xi(P,B,\alpha) u_2 v^* \end{bmatrix} v = 0.$$

Consequently a perturbation to the leading coefficient with norm at most $\alpha_k \xi(P,B,\alpha)$ yields the singular matrix $K_k - \alpha_k \xi(P,B,\alpha) v u_1^*$, which contradicts the nonsingularity assumption. ∎

THEOREM 2.3. *With the assumptions of Lemma* 2.2 *for the system* (1.1) *the equality* $\tau(P,B,\alpha) = \xi(P,B,\alpha)$ *holds for* τ *defined in* (1.11) *both in the spectral norm and in the Frobenius norm.*

Proof. First we assume that $\tau(P,B,\alpha)$ in (1.11) is defined in the spectral norm and show that $\xi(P,B,\alpha) \leq \tau(P,B,\alpha)$. From Lemma 2.1, there exists $\Delta P(\lambda) = \sum_{j=0}^{k} \alpha_j \lambda^j \Delta K_j$ such that

$$\tau(P,B,\alpha) = \|[\Delta K_k \ \cdots \ \Delta K_0 \ \Delta B]\|$$

and for some $\tilde{\lambda}$ the matrix $[(P+\Delta P)(\tilde{\lambda}) \ \ B + \Delta B]$ is rank deficient, that is

$$\begin{bmatrix} ((P+\Delta P)(\tilde{\lambda}))^* \\ B^* + \Delta B^* \end{bmatrix} v = 0$$

for some unit $v \in \mathbb{C}^n$. We collect the perturbations on the right and divide the upper blocks by $\sqrt{s_\alpha(|\tilde{\lambda}|)}$ to obtain

$$\begin{bmatrix} \left(\frac{P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}}\right)^* \\ B^* \end{bmatrix} v = \begin{bmatrix} \left(-\frac{\Delta P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}}\right)^* \\ -\Delta B^* \end{bmatrix} v.$$

Therefore

$$\begin{aligned} \xi(P,B,\alpha) &\leq \sigma_{\min}\left(\begin{bmatrix} \frac{P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}} & B \end{bmatrix}\right) \\ &= \sigma_{\min}\left(\begin{bmatrix} \left(\frac{P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}}\right)^* \\ B^* \end{bmatrix}\right) \leq \left\| \begin{bmatrix} \left(\frac{P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}}\right)^* \\ B^* \end{bmatrix} v \right\| \\ &= \left\| \begin{bmatrix} \left(\frac{\Delta P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}}\right)^* \\ \Delta B^* \end{bmatrix} v \right\| \leq \left\| \begin{bmatrix} \left(\frac{\Delta P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}}\right)^* \\ \Delta B^* \end{bmatrix} \right\| = \left\| \begin{bmatrix} \frac{\Delta P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}} & \Delta B \end{bmatrix} \right\|. \end{aligned}$$

Moreover,

$$\begin{bmatrix} \frac{\Delta P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}} & \Delta B \end{bmatrix} = [\Delta K_k \ \cdots \ \Delta K_0 \ \Delta B] \begin{bmatrix} \frac{\alpha_k \tilde{\lambda}^k I}{\sqrt{s_\alpha(|\tilde{\lambda}|)}} & 0 \\ \vdots & \vdots \\ \frac{\alpha_1 \tilde{\lambda} I}{\sqrt{s_\alpha(|\tilde{\lambda}|)}} & 0 \\ \frac{\alpha_0 I}{\sqrt{s_\alpha(|\tilde{\lambda}|)}} & 0 \\ 0 & I \end{bmatrix},$$

where the spectral norm of the rightmost matrix is one. It follows from the Cauchy–Schwarz inequality that

$$\xi(P,B,\alpha) \le \left\| \begin{bmatrix} \dfrac{\Delta P(\tilde{\lambda})}{\sqrt{s_\alpha(|\tilde{\lambda}|)}} & \Delta B \end{bmatrix} \right\| \le \|[\Delta K_k \ \cdots \ \Delta K_0 \ \ \Delta B]\| = \tau(P,B,\alpha).$$

For the reverse inequality, still using the spectral norm, we have from Lemma 2.2 that for some φ,

$$\xi(P,B,\alpha) = \sigma_{\min}\left(\begin{bmatrix} \dfrac{P(\varphi)}{\sqrt{s_\alpha(|\varphi|)}} & B \end{bmatrix} \right) = \sigma_{\min}\left(\begin{bmatrix} \left(\dfrac{P(\varphi)}{\sqrt{s_\alpha(|\varphi|)}}\right)^* \\ B^* \end{bmatrix} \right)$$

or equivalently

$$\begin{bmatrix} \frac{(P(\varphi))^*}{\sqrt{s_\alpha(|\varphi|)}} \\ B^* \end{bmatrix} v = \xi(P,B,\alpha) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

where $v, u_1 \in \mathbb{C}^n$, $u_2 \in \mathbb{C}^m$, and the vectors v and $[u_1^T \ u_2^T]^T$ have unit length. We multiply the right-hand side by v^*v, the upper blocks by $\sqrt{s_\alpha(|\varphi|)}$ and collect all terms on the left to obtain

$$\begin{bmatrix} (P(\varphi))^* - \sqrt{s_\alpha(|\varphi|)}\xi(P,B,\alpha)u_1v^* \\ B^* - \xi(P,B,\alpha)u_2v^* \end{bmatrix} v = 0.$$

In other words, the matrix

$$\begin{bmatrix} P(\varphi) - \sqrt{s_\alpha(|\varphi|)}\xi(P,B,\alpha)vu_1^* & B - \xi(P,B,\alpha)vu_2^* \end{bmatrix}$$

is rank deficient. If we set $\Delta K_j = \frac{-\alpha_j \bar{\varphi}^j \xi(P,B,\alpha)vu_1^*}{\sqrt{s_\alpha(|\varphi|)}}$ and $\Delta B = -\xi(P,B,\alpha)vu_2^*$ and define $\Delta P(\lambda) = \sum_{j=0}^m \alpha_j \lambda^j \Delta K_j$, then by noting

$$\Delta P(\varphi) = \sum_{j=0}^m \alpha_j \varphi^j \Delta K_j = -\sqrt{s_\alpha(|\varphi|)}\xi(P,B,\alpha)vu_1^*$$

we see that

$$[(P+\Delta P)(\lambda) \ \ B + \Delta B]$$

is rank deficient at $\lambda = \varphi$. The norm of the perturbations satisfies

$$\begin{aligned} &\|[\Delta K_k \ \cdots \ \Delta K_0 \ \ \Delta B]\| \\ &= \xi(P,B,\alpha) \left\| \begin{bmatrix} \alpha_k \bar{\varphi}^k \dfrac{vu_1^*}{\sqrt{s_\alpha(|\varphi|)}} & \ldots & \alpha_0 \dfrac{vu_1^*}{\sqrt{s_\alpha(|\varphi|)}} & vu_2^* \end{bmatrix} \right\| \le \xi(P,B,\alpha). \end{aligned}$$

Therefore $\tau(P,B,\alpha) \le \|[\Delta K_k \ \cdots \ \Delta K_0 \ \ \Delta B]\| \le \xi(P,B,\alpha)$ as desired.

For the claim about the equality when $\tau(P,B,\alpha)$ is defined in the Frobenius norm, to show $\xi(P,B,\alpha) \le \tau(P,B,\alpha)$ the proof in the first part applies noting that

$$\xi(P,B,\alpha) \le \|[\Delta K_k \ \cdots \ \Delta K_0 \ \ \Delta B]\|_2 \le \|[\Delta K_k \ \cdots \ \Delta K_0 \ \ \Delta B]\|_F = \tau(P,B,\alpha).$$

The second part to show $\tau(P,B,\alpha) \leq \xi(P,B,\alpha)$ applies without modification. ∎

The second part of Theorem 2.3 explicitly constructed the closest uncontrollable system which we state in the next corollary.

COROLLARY 2.4. *Suppose the assumptions of Lemma 2.2 hold. Let $\xi(P,B,\alpha)$ be attained at λ_*, and let the vectors $[u_1^T \ u_2^T]^T$ and v be the unit right and left singular vectors corresponding to*

$$\sigma_{\min}\left(\begin{bmatrix} \dfrac{P(\lambda_*)}{s_\alpha(|\lambda_*|)} & B \end{bmatrix}\right),$$

respectively, where $u_1, v \in \mathbb{C}^n$ and $u_2 \in \mathbb{C}^m$. A closest uncontrollable tuple is $(K_k + \alpha_k \Delta K_k, \ldots, K_0 + \alpha_0 \Delta K_0, B + \Delta B)$, where

$$\Delta K_j = \frac{-\alpha_j \bar{\lambda}_*^j \xi(P,B,\alpha) v u_1^*}{\sqrt{s_\alpha(|\lambda_*|)}}, \quad j = 0, \ldots, k$$

and

$$\Delta B = -\xi(P,B,\alpha) v u_2^*.$$

Finally to remove the condition that $\alpha_0 \neq 0$, clearly $\tau(P,B,\alpha)$ depends on α_0 continuously when $\alpha_0 > 0$ and is continuous from the right when $\alpha_0 = 0$. (Consider the distance of $(K_k, K_{k-1}, \ldots, B)$ to any fixed uncontrollable tuple as a function of α_0 with all other α_j fixed. If such a distance function is bounded around a given α_0, then it is continuous from the right and the minimum of these continuous distance functions is $\tau(P,B,\alpha)$ as a function of α_0.) Therefore if $\alpha_0 = 0$, which is particularly the case when $s_\alpha(|\lambda|) = 0$, then the limiting value of $\xi(P,B,\alpha)$ from the right must approach $\tau(P,B,\alpha)$.

THEOREM 2.5. *With the conditions stated in Lemma 2.2 except that α_0 is allowed to be any nonnegative real number (possibly zero), the equality*

$$\tau(P,B,[\alpha_k, \alpha_{k-1}, \ldots, \alpha_0]) = \lim_{\alpha_0' \to \alpha_0^+} \xi(P,B,[\alpha_k, \alpha_{k-1}, \ldots, \alpha_0'])$$

holds where τ is defined in either the spectral norm or the Frobenius norm.

Specifically when $\tau(P,B,\alpha) = \|[0 \ 0 \cdots \Delta B]\| = \|\Delta B\|$, that is a closest uncontrollable system can be obtained just by perturbing B (this has to be the case when $\alpha = 0$), the result above amounts to a minimization problem over the vectors that are constrained to lie in the left eigenspace of P, $\mathcal{S}_P$, which we can see as follows. If we restrict the perturbations only to B and without loss of generality assume $\alpha = 0$, then the definition of the higher order distance to uncontrollability simplifies as

$$\begin{aligned} \tau(P,B) &= \inf\{\|\Delta B\| : v^*[P(\lambda) \ \ B + \Delta B] = 0, \ \exists v \in \mathbb{C}^n, \ \lambda \in \mathbb{C}\} \\ &= \inf\{\|\Delta B\| : v^* B = -v^* \Delta B, \ v \in \mathcal{S}_P\}. \end{aligned}$$

The last minimization problem must be attained at a ΔB such that $\|\Delta B\| = \|v^* \Delta B\|$, where $v \in \mathcal{S}_P$, because otherwise we can obtain a matrix ΔB smaller in norm by replacing all of the singular values larger than $\|v^* \Delta B\|$ with 0 that still satisfies the constraint $v^* B = -v^* \Delta B$. Therefore the last minimization problem is equivalent to

$$\tau(P,B) = \inf\{\|v^* \Delta B\| : v \in \mathcal{S}_p, v^* B = -v^* \Delta B\} = \inf_{v \in \mathcal{S}_p} \|v^* B\|.$$

Now we can verify Theorem 2.5 for this special case, as indeed

$$\lim_{\alpha_0\to 0^+} \xi(P,B,[0\ 0\ \cdots\ \alpha_0]) = \lim_{\alpha_0\to 0^+} \inf_{\lambda\in\mathbb{C}} \sigma_{\min}\left(\begin{bmatrix} P(\lambda) \\ \alpha_0 \end{bmatrix} B\right)$$
$$= \lim_{\alpha_0\to 0^+} \inf_{\lambda\in\mathbb{C}, v\in\mathbb{C}^n} \left\| v^* \begin{bmatrix} P(\lambda) \\ \alpha_0 \end{bmatrix} B \right\|.$$

Furthermore as $\alpha_0 \to 0^+$, any solution pair λ, v of the minimization problem must correspond to an eigenvalue of P and the associated left eigenvector, respectively. Therefore the minimization problem reduces to

$$\lim_{\alpha_0\to 0^+} \xi(P,B,[0\ 0\ \cdots\ \alpha_0]) = \inf_{v\in\mathcal{S}_P} \|v^*B\| = \tau(P,B).$$

3. A practical algorithm exploiting the singular value characterization. In Theorem 2.3 we established the equality

$$\tau(P,B,\alpha) = \xi(P,B,\alpha) = \inf_{r\geq 0,\theta\in[0,2\pi)} f(r,\theta)$$

when $\alpha_0 \neq 0$, where

$$f(r,\theta) = \sigma_{\min}\left(\begin{bmatrix} \dfrac{P(re^{i\theta})}{\sqrt{s_\alpha(r)}} \\ \ \end{bmatrix} B\right).$$

When $\alpha_0 = 0$, the limit of $\xi(P,B,\alpha)$ as $\alpha_0 \to 0^+$ approaches the distance to uncontrollability. Therefore, in essence the computation of the distance to uncontrollability can be achieved by minimizing $f(r,\theta)$. In this section we present a trisection algorithm to minimize the function $f(r,\theta)$ in polar coordinates. Let δ_1 and δ_2 trisect the interval $[L,U]$ containing the distance to uncontrollability (see Figure 3.1). At each iteration the algorithm updates either the upper bound to δ_1 or the lower bound to δ_2 depending on whether the δ-level set of $f(r,\theta)$

$$\{re^{i\theta} : f(r,\theta) = \delta\}$$

is intersected by any line in the set of lines passing through the origin with slopes multiples of η, where δ and η are determined by δ_1 and δ_2 as

$$\delta = \delta_1, \quad \eta = \frac{2}{k}\arccos\left(1 - \frac{1}{2}\left(\frac{\delta_1-\delta_2}{ckK_{\max}}\right)^2\right).$$

Above c is a positive real constant depending on the modulus of a point in the complex plane where $\xi(P,B,\alpha)$ is attained and $K_{\max}$ is a positive real constant depending on the norms of the coefficient matrices. (The constants c and $K_{\max}$ are defined precisely in the paragraph preceding Theorem 3.2.) We say the angle η subtends all of the components of the δ-level set of f, when no component has a pair of points whose angles differ by more than η. At each iteration we verify only one of the following (even though both of them may sometimes be true);

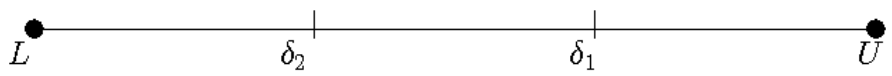

FIG. 3.1. *The trisection algorithm keeps track of an interval* $[L,U]$ *containing* $\xi(P,B,\alpha)$. *At each iteration either* L *is updated to* δ_2 *or* U *is updated to* δ_1.

- the δ-level set of f is not empty,
- the angle η subtends all of the components of the δ-level set of f.

By the definition of $\xi(P, B, \alpha)$ when the δ-level set is not empty

$$\delta = \delta_1 \geq \xi(P, B, \alpha) \tag{3.1}$$

and when η subtends all of the components of the δ-level set we will see below that

$$\xi(P, B, \alpha) > \delta_2 \tag{3.2}$$

because of the choice of η and δ. The algorithm we present is inspired by the trisection algorithm of [3] for the first order distance to uncontrollability. However, the technique we use to verify which one of (3.1) and (3.2) holds is new and has no similarity with the verification technique used in [3] to trisect an interval known to contain the first order distance to uncontrollability. A straightforward modification of the technique for the first order distance to uncontrollability would require the solution of polynomial eigenvalue problems quadratic in size and double in degree as compared to the original polynomial eigenvalue problem, which is too expensive even for systems of small size.

The trisection algorithm starts with the trivial upper bound $U = \sigma_{\min}([K_k/\alpha_k \;\; B])$ (or when $\alpha_k = 0$, $U = \sigma_{\min}(B)$) and the lower bound $L = 0$. At each iteration we either update the upper bound to δ_1 if the inequality (3.1) is verified or the lower bound to δ_2 if the inequality (3.2) is verified. First we need to be equipped with a technique that checks for a given δ and θ whether there exists an r satisfying

$$f(r, \theta) = \delta, \tag{3.3}$$

that is whether the line with slope θ passing through the origin, say $\mathcal{L}(\theta)$, intersects the δ-level set of f. Our first result in this section shows how this can be achieved by solving a polynomial eigenvalue problem of double size and of double degree. Similar results relating the δ-level set of $g(x, y) = \sigma_{\min}(A - (x + yi)I)$, where $A \in \mathbb{C}^{n\times n}$, $x, y \in \mathbb{R}$ and the imaginary eigenvalues of a matrix $G(x, \delta)$ of double size can be found in [4] and [2]. More precisely these results suggest how to find the intersection points of the δ-level set of $g(x, y)$ and a vertical line; that is the results deduce that if $\delta = g(x, y)$, then yi is an eigenvalue of $G(x, \delta)$.

THEOREM 3.1. *Given $\theta \in [0, 2\pi)$ and a positive real number δ, the matrix $[\frac{P(re^{i\theta})}{\sqrt{s_\alpha(r)}} \;\; B]$ has δ as a singular value if and only if the matrix polynomial of double size $Q(\lambda, \theta, \delta) = \sum_{j=0}^{2k} \lambda^j Q_j(\theta, \delta)$ has the imaginary eigenvalue ri where*

$$Q_0(\theta, \delta) = \begin{bmatrix} -\delta\alpha_0^2 I & K_0^* \\ K_0 & BB^*/\delta - \delta I \end{bmatrix},$$

and, when l is odd,

$$Q_l(\theta, \delta) = \begin{bmatrix} 0 & (-1)^{(l+1)/2} i K_l^* e^{-il\theta} \\ (-1)^{(l+1)/2} i K_l e^{il\theta} & 0 \end{bmatrix} \quad 1 \leq l \leq k,$$
$$Q_l(\theta, \delta) = 0 \quad k+1 \leq l < 2k,$$

and, when l is even,

$$Q_l(\theta, \delta) = \begin{bmatrix} (-1)^{l/2+1}\delta\alpha_{l/2}^2 I & (-1)^{l/2} K_l^* e^{-il\theta} \\ (-1)^{l/2} K_l e^{il\theta} & 0 \end{bmatrix} \quad 1 \leq l \leq k,$$
$$Q_l(\theta, \delta) = \begin{bmatrix} (-1)^{l/2+1}\delta\alpha_{l/2}^2 I & 0 \\ 0 & 0 \end{bmatrix} \quad k+1 \leq l \leq 2k.$$

Proof. The matrix $[\frac{P(re^{i\theta})}{\sqrt{s_\alpha(r)}} \;\; B]$ has δ as a singular value if and only if both of the equations

$$\begin{bmatrix} \frac{P(re^{i\theta})}{\sqrt{s_\alpha(r)}} & B \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \delta u,$$
$$\begin{bmatrix} \left(\frac{P(re^{i\theta})}{\sqrt{s_\alpha(r)}}\right)^* \\ B^* \end{bmatrix} u = \delta \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

are satisfied. From the bottom block of the second equation we have $v_2 = B^*u/\delta$. By eliminating v_2 from the other equation, we obtain

$$\begin{aligned} &\begin{bmatrix} -\delta I & \left(\frac{P(re^{i\theta})}{\sqrt{s_\alpha(r)}}\right)^* \\ \frac{P(re^{i\theta})}{\sqrt{s_\alpha(r)}} & BB^*/\delta - \delta I \end{bmatrix} \begin{bmatrix} v_1 \\ u \end{bmatrix} \\ &= \begin{bmatrix} -\delta s_\alpha(r) I & \left(P(re^{i\theta})\right)^* \\ P(re^{i\theta}) & BB^*/\delta - \delta I \end{bmatrix} \begin{bmatrix} v_1/\sqrt{s_\alpha(r)} \\ u \end{bmatrix} \\ &= \sum_{j=0}^{2k} (ri)^j Q_j(\theta, \delta) \begin{bmatrix} v_1/\sqrt{s_\alpha(r)} \\ u \end{bmatrix} = 0. \end{aligned}$$

Therefore ri is an eigenvalue of $Q(\lambda, \theta, \delta)$. ∎

Suppose $\delta \le \lim_{\lambda\to\infty} \sigma_{\min}([\frac{P(\lambda)}{\sqrt{s_\alpha(|\lambda|)}} \;\; B])$. To establish the existence of an r satisfying (3.3), it is sufficient that the polynomial $Q(\lambda, \theta, \delta)$ has an imaginary eigenvalue. When $Q(\lambda, \theta, \delta)$ has an imaginary eigenvalue $r'i$, $f(r', \theta) \le \delta$. Since $\delta \le f(r, \theta)$ in the limit as $r \to \infty$, by the continuity of f with respect to r we deduce $f(\hat{r}, \theta) = \delta$ for some $\hat{r} \ge r'$.

For our trisection algorithm it suffices to check whether any of the lines $\mathcal{L}(0), \mathcal{L}(\eta), \mathcal{L}(2\eta), \ldots, \mathcal{L}(\lfloor \frac{\pi}{\eta} \rfloor \eta)$ intersect the δ-level set of f as illustrated in Figure 3.2. When there is an intersection point the δ-level set is not empty; otherwise the angle η subtends all of the components. The only part of the algorithm that is not clarified so far is how we conclude a lower bound on $\xi(P, B, \alpha)$ when η subtends all of the components, in particular the relation between δ_2 in (3.2) and the pair δ and η. For the next theorem addressing these issues let (r_*, θ_*) be a point where $\xi(P, B, \alpha)$ is attained. We assume the existence of a constant c known *a priori* satisfying

$$c \ge \max_{0\le j\le k} \frac{r_*^j}{\sqrt{s_\alpha(r_*)}} = \max\left(\frac{1}{\sqrt{s_\alpha(r_*)}}, \frac{r_*^k}{\sqrt{s_\alpha(r_*)}}\right). \tag{3.4}$$

Finding a constant c may be tedious in some special cases. However, when both α_k and α_0 are nonzero we can set $c = \frac{1}{\min(\alpha_0, \alpha_k)}$. We furthermore use the notation $K_{\max} = \max_{1\le j\le k} \|K_j\|$. The algorithms in [14, 15, 3] for the first order distance to uncontrollability benefit from an analogous result in [14] which can be stated as, given a $\delta \ge \tau(A, B)$ for all $\eta \in [0, 2(\delta - \tau(A, B))]$ there exists a pair of real numbers x, y satisfying $\sigma_{\min}([A - (x + yi)I \;\; B]) = \sigma_{\min}([A - (x + \eta + yi)I \;\; B]) = \delta$. Throughout the rest of this section we omit the parameters of $\xi(P, B, \alpha)$ assuming P, B, and α are fixed.

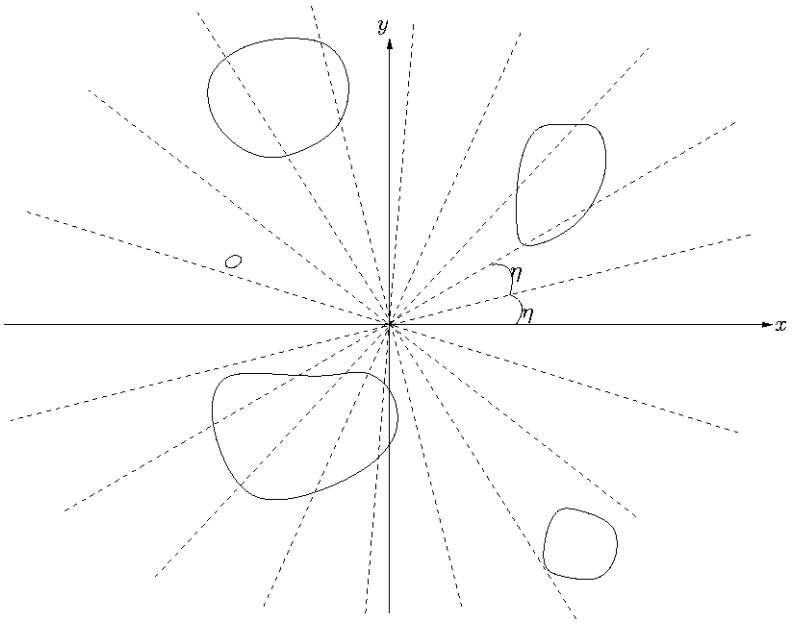

FIG. 3.2. *To verify which one of* (3.1) *and* (3.2) *hold we check the intersection points of the* δ*-level set of* f *and the set of lines with slopes multiples of* η *ranging from 0 to* π*. The closed curves are the* δ*-level sets.*

THEOREM 3.2. *Let*

$$\lim_{\lambda\to\infty} \sigma_{\min}\left(\left[\begin{array}{cc} \dfrac{P(\lambda)}{\sqrt{s_\alpha(|\lambda|)}} & B \end{array}\right]\right) \geq \delta > \xi.$$

Given any $\eta \in [0, \frac{1}{k}\arccos(1-\frac{1}{2}(\frac{\delta-\xi}{ckK_{\max}})^2)]$*, there exist* r_1 *and* r_2 *(depending on* η*) such that*

$$\sigma_{\min}\left(\left[\begin{array}{cc} \dfrac{P(r_1 e^{i(\theta_*+\eta)})}{\sqrt{s_\alpha(r_1)}} & B \end{array}\right]\right) = \delta \quad \text{and} \quad \sigma_{\min}\left(\left[\begin{array}{cc} \dfrac{P(r_2 e^{i(\theta_*-\eta)})}{\sqrt{s_\alpha(r_2)}} & B \end{array}\right]\right) = \delta.$$

Proof. We prove the first equality. The proof of the second equality is similar. Assume

$$\sigma_{\min}\left(\left[\begin{array}{cc} \dfrac{P(re^{i(\theta_*+\eta)})}{\sqrt{s_\alpha(r)}} & B \end{array}\right]\right) > \delta \tag{3.5}$$

holds for all r for an η in the interval specified. Since the singular values of a matrix X are the eigenvalues of the symmetric matrix

$$\left[\begin{array}{cc} 0 & X \\ X^* & 0 \end{array}\right],$$

they are globally Lipschitz with constant 1 (see Weyl's Theorem [19, Theorem (4.3.1)]) meaning

$$\delta - \xi < \sigma_{\min}\left(\begin{bmatrix} \frac{P(r_* e^{i(\theta_*+\eta)})}{\sqrt{s_\alpha(r_*)}} & B \end{bmatrix}\right) - \sigma_{\min}\left(\begin{bmatrix} \frac{P(r_* e^{i\theta_*})}{\sqrt{s_\alpha(r_*)}} & B \end{bmatrix}\right)$$
$$\leq \left\| \begin{bmatrix} \frac{P(r_* e^{i(\theta_*+\eta)})}{\sqrt{s_\alpha(r_*)}} & B \end{bmatrix} - \begin{bmatrix} \frac{P(r_* e^{i\theta_*})}{\sqrt{s_\alpha(r_*)}} & B \end{bmatrix} \right\| = \left\| \frac{\sum_{j=1}^{k} r_*^j e^{ij\theta_*} K_j (e^{ij\eta} - 1)}{\sqrt{s_\alpha(r_*)}} \right\|.$$

Notice that $\eta \leq \pi/k$ implying $\cos k\eta \leq \cos j\eta$ for $j = 0, \ldots, k$. Therefore

$$kcK_{\max}\sqrt{2 - 2\cos k\eta} \geq \sum_{j=1}^{k} c\|K_j\sqrt{2 - 2\cos j\eta}\| \geq \left\| \frac{\sum_{j=1}^{k} r_*^j e^{ij\theta_*} K_j (e^{ij\eta} - 1)}{\sqrt{s_\alpha(r_*)}} \right\| > \delta - \xi$$

or

$$1 - \frac{1}{2}\left(\frac{\delta - \xi}{kcK_{\max}}\right)^2 > \cos k\eta.$$

Since the *cosine* function is strictly decreasing in the interval $[0, \pi]$, we obtain the contradiction that

$$\eta > \frac{1}{k}\arccos\left(1 - \frac{1}{2}\left(\frac{\delta - \xi}{kcK_{\max}}\right)^2\right).$$

Thus, (3.5) cannot hold, so there exists r_1' satisfying

$$\sigma_{\min}\left(\begin{bmatrix} \frac{P(r_1' e^{i(\theta_*+\eta)})}{\sqrt{s_\alpha(r_1')}} & B \end{bmatrix}\right) \leq \delta.$$

The first equality must therefore hold for some $r_1 \geq r_1'$ because of the continuity of $f(r, \theta_* + \eta)$ with respect to r and the fact that $\lim_{r\to\infty} f(r, \theta_* + \eta) \geq \delta$. ∎

As we have already indicated in (3.1), we first set $\delta = \delta_1$. The assignment

$$\eta = \frac{2}{k}\arccos\left(1 - \frac{1}{2}\left(\frac{\delta_1 - \delta_2}{ckK_{\max}}\right)^2\right) \tag{3.6}$$

leads us to the lower bound (3.2) in the case that none of the lines $\mathcal{L}(0), \mathcal{L}(\eta), \mathcal{L}(2\eta), \ldots, \mathcal{L}(\lfloor\frac{\pi}{\eta}\rfloor\eta)$ intersect the δ-level set of f, which we can see as follows. According to Theorem 3.2 for all θ in the interval

$$\left[\theta_* - \frac{1}{k}\arccos\left(1 - \frac{1}{2}\left(\frac{\delta - \xi}{ckK_{\max}}\right)^2\right), \theta_* + \frac{1}{k}\arccos\left(1 - \frac{1}{2}\left(\frac{\delta - \xi}{ckK_{\max}}\right)^2\right)\right], \tag{3.7}$$

the line $\mathcal{L}(\theta)$ intersects the δ-level set of f. When none of the lines $\mathcal{L}(0), \mathcal{L}(\eta), \mathcal{L}(2\eta), \ldots, \mathcal{L}(\lfloor\frac{\pi}{\eta}\rfloor\eta)$ intersects the δ-level set of f, it follows that η must be greater than the length of the interval in (3.7), that is

$$\eta = \frac{2}{k}\arccos\left(1 - \frac{1}{2}\left(\frac{\delta_1 - \delta_2}{ckK_{\max}}\right)^2\right) > \frac{2}{k}\arccos\left(1 - \frac{1}{2}\left(\frac{\delta - \xi}{ckK_{\max}}\right)^2\right).$$

From this inequality it is straightforward to deduce the lower bound (3.2). Algorithm 1 summarizes the approach described.

As the accuracy and efficiency of the algorithm depend on the extraction of the imaginary eigenvalues of the matrix polynomial $Q(\lambda, \theta, \delta)$, it is worth pointing out how these eigenvalues can be computed numerically in a reliable fashion. The matrix polynomial $Q(\lambda, \theta, \delta)$ has a special structure; its even coefficients are Hermitian, while its odd coefficients are skew-Hermitian. The eigenvalues of polynomials with this structure are either imaginary or in pairs $(\lambda, -\bar{\lambda})$ [20]. The standard way to solve a polynomial eigenvalue problem of size $2n$ and degree $2k$ is to reduce it to an equivalent generalized eigenvalue problem $\mathcal{H} - \lambda\mathcal{N}$ of size $4nk$ by a transformation called linearization. The most widely used linearization is the companion form [21]. In [21] vector spaces of linearizations that are generalizations of the companion form are introduced. There are two issues one needs to consider when selecting a linearization. First the structure must be preserved, that is the matrices $\mathcal{H}, \mathcal{N}$ in the transformation above must be Hermitian and skew-Hermitian, respectively. Sec-

Algorithm 1 Trisection algorithm for the higher order distance to uncontrollability

Call:	$[L, U] \leftarrow$ `HODU`(P, B, α, tol, c).
Input:	$P \in \mathbb{C}^{k\times n\times n}$ (the matrix polynomial), $B \in \mathbb{C}^{n\times m}$, $\alpha \in \mathbb{R}^k$ (nonnegative scaling factors, not all zero), tol (desired tolerance), c (a positive real number satisfying (3.4)).
Output:	L, U with $L < U$, $U - L \leq tol$. The interval $[L, U]$ contains the higher order distance to uncontrollability.

Initially set

$$U \leftarrow \sigma_{\min}\left(\begin{bmatrix} \frac{K_k}{\alpha_k} & B \end{bmatrix}\right) \quad \text{if } \alpha_k > 0,$$
$$U \leftarrow \sigma_{\min}(B) \quad \text{if } \alpha_k = 0,$$

and $L \leftarrow 0$.
while $U - L > tol$ **do**
  % Trisection step
  Set $\delta_1 \leftarrow L + 2(U - L)/3$ and $\delta_2 \leftarrow L + (U - L)/3$.
  Set $\delta \leftarrow \delta_1$ and η as defined in (3.6)
  Set $Intersection \leftarrow FALSE$.
  for $\theta = 0$ to π in increments of η **do**
    Compute the eigenvalues of $Q(\lambda, \theta, \delta)$.
    if $Q(\lambda, \theta, \delta)$ has an imaginary eigenvalue **then**
      % An intersection point is detected
      Update the upper bound, $U \leftarrow \delta_1$.
      $Intersection \leftarrow TRUE$.
      Break. (Leave the for loop.)
    end if
  end for
  if $\neg Intersection$ **then**
    % No intersection point is detected
    Update the lower bound, $L \leftarrow \delta_2$.
  end if
end while
Return $[L, U]$.

ond the eigenvalues of the pencil $\mathcal{H} - \lambda\mathcal{N}$ have different condition numbers than the eigenvalues of the matrix polynomial $Q(\lambda, \theta, \delta)$. Ideally we must use a linearization preserving the structure that does not degrade the conditioning of the eigenvalues of the original problem. The linearizations in the vector spaces specified in [21] that preserve the even-odd structure of $Q(\lambda, \theta, \delta)$ are identified in [20]. Furthermore in [17] it was shown that in these vector spaces there are linearizations preserving the conditioning of the eigenvalues of $Q(\lambda, \theta, \delta)$. How best to find such a linearization preserving the structure and the conditioning combined with an even-odd generalized eigenvalue solver is still under investigation. When such an implementation is used, simple imaginary eigenvalues remain on the imaginary axis even in the presence of rounding errors. Therefore tolerances are not needed.

At each iteration the algorithm requires the solution of the eigenvalue problems $Q(\lambda, 0, \delta), Q(\lambda, \eta, \delta), \ldots, Q(\lambda, \lfloor \frac{\pi}{\eta} \rfloor \eta, \delta)$, each typically at a cost of $O(n^3k^3)$. The overall complexity of an iteration is

$$O\left(\frac{n^3k^4}{\arccos\left(1 - \frac{1}{2}\left(\frac{\delta_1 - \delta_2}{ckK_{\max}}\right)^2\right)}\right). \tag{3.8}$$

It is apparent that the initial iterations for which $\delta_1 - \delta_2$ is relatively large are cheaper, while the last iteration for which $\delta_1 - \delta_2 \approx tol/2$ is the most expensive.

4. Numerical results. All of the numerical experiments in this section are performed with MATLAB 6.5 running on a PC with 1000 MHz Intel processor and 256MB RAM.

4.1. Computing the distance to uncontrollability for first order systems. Even though it is much slower than the methods in [14, 3, 15], the trisection algorithm suggested can be applied to estimate the first order distance to uncontrollability with $k = 1$, $K_1 = I$, and $\alpha = [0\ 1]$ so that perturbations to $K_1 = I$ are not allowed. It is well known that in this case the distance to uncontrollability is attained at a point λ_* with $|\lambda_*| = c \le 2(\|K_0\| + \|B\|)$. We choose K_0 as the Toeplitz matrix

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ -2 & 1 & 3 & 0 \\ 0 & -2 & 1 & 3 \\ 0 & 0 & -2 & 1 \end{bmatrix}$$

and $B = [2\ 2\ 2\ 2]^T$. When we require an interval of length 10^{-2} or less, Algorithm 1 returns $[0.473, 0.481]$ in 12 iterations which contains the distance to uncontrollability 0.477. Table 4.1 lists the cumulative running time after each iteration in seconds. Overall we observe that reaching one digit accuracy is considerably cheaper than two digit accuracy. When we allow the perturbations to the leading coefficient by setting $\alpha = [1\ 1]$, there is a closer uncontrollable system at a distance of $\tau(P, B, \alpha) \le 0.145$ which is the upper bound returned by Algorithm 1.

4.2. A quadratic brake model. In [12] the vibrations of a drum brake system are modeled by the quadratic equation

$$Mx^{(2)}(t) + K(\mu)x(t) = f(t) \tag{4.1}$$

TABLE 4.1
Total running time of the trisection algorithm after each iteration on a Toeplitz matrix and a vector pair.

Iteration	**Total running time**	**Interval** $[L, U]$
1	0.400	[0.000,0.667]
2	1.680	[0.222,0.667]
3	2.510	[0.222,0.519]
4	5.369	[0.321,0.519]
5	9.670	[0.387,0.519]
6	16.110	[0.431,0.519]
7	20.140	[0.431,0.489]
8	34.580	[0.450,0.489]
9	56.770	[0.463,0.489]
10	70.470	[0.463,0.481]
11	118.40	[0.469,0.481]
12	190.93	[0.473,0.481]

TABLE 4.2
The intervals computed by the trisection algorithm for the brake system for various μ values in an absolute sense in the second column and in a relative sense in the third column.

μ	**Interval** $[L, U]$ **(Absolute)**	**Interval** $[L, U]$ **(Relative)**
0.05	[0.051,0.059]	[0.038,0.046]
0.10	[0.097,0.105]	[0.071,0.079]
0.15	[0.140,0.148]	[0.104,0.112]
0.20	[0.184,0.191]	[0.137,0.145]
0.50	[0.418,0.426]	[0.325,0.333]
1	[0.676,0.684]	[0.574,0.581]
10	[0.990,0.997]	[0.984,0.991]
100	[0.993,1.000]	[0.987,0.994]
1000	[0.993,1.000]	[0.987,0.994]

with the mass and stiffness matrices

$$M = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}, \quad K(\mu) = g \begin{bmatrix} (\sin\gamma + \mu\cos\gamma)\sin\gamma & -\mu - (\sin\gamma + \mu\cos\gamma)\cos\gamma \\ (\mu\sin\gamma - \cos\gamma)\sin\gamma & 1 + (-\mu\sin\gamma + \cos\gamma)\cos\gamma \end{bmatrix}.$$

Suppose the force on the brake system has just the vertical component determined by the input

$$f(t) = \begin{bmatrix} f_x(t) \\ f_y(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t).$$

For the parameters $m = 5$, $g = 1$ and $\gamma = \frac{\pi}{100}$, we consider two cases. First by setting $\alpha = [1\ 0\ 1]$, we impose equal importance on the perturbations to the mass and stiffness matrices. Notice that for small μ and γ, the system is close to being uncontrollable. In the second column in Table 4.2 the intervals of length 10^{-2} or less containing the distance to uncontrollability returned by Algorithm 1 are provided for various values of μ. The algorithm iterates 16 times to reach two digit accuracy. Second we assign scaling to the perturbations proportional to the norms of the mass and stiffness matrices, that is $\alpha = [\|M\|\ 0\ \|K\|]$. The intervals returned by Algorithm 1 for this second case are given in the rightmost column in Table 4.2. As expected the distance to uncontrollability again increases with respect to μ. The system (4.1) is closer to being uncontrollable in a relative sense than in an absolute sense.

If we allow perturbations to all coefficients with equal scaling (e.g., $\alpha = [1\ 1\ 1]$), then usually the first order distance uncontrollability of the embedded system (1.5) is

TABLE 4.3
Running time of the trisection algorithm in seconds with respect to the size and order of the systems with normally distributed coefficient matrices.

Size / order	First order	Quadratic	Cubic
5	10 (10)	192 (12)	1237 (13)
10	83 (12)	1392 (11)	12485 (12)
15	271 (13)	6390 (14)	37324 (12)

considerably smaller than the actual value $\tau(P, B, \alpha)$, since the perturbations are not constrained so that the structure of the embedding can be preserved. For instance, for the drum brake system with $\alpha = [1\ 1\ 1]$ and $\mu = 0.1$, $\tau(P, B, \alpha) \in [0.097, 0.105]$ (up to two digit accuracy it does not make any difference whether we allow perturbations to the zero coefficient K_1 or not) whereas the standard unstructured distance to uncontrollability of the embedding lies in the interval $[0.012, 0.020]$.

4.3. Running time with respect to the size and order of the system. We run the trisection algorithm on systems with random coefficients of various size and order. To be precise the entries of all of the coefficient matrices are chosen from a normal distribution with zero mean and variance one independently. Table 4.3 illustrates how the running time in seconds varies with respect to the size and order of the system. In all of the examples intervals of length at most 10^{-2} containing the absolute distance to uncontrollability (α is the vector of ones) are returned. The numbers in parentheses correspond to the number of trisection iterations needed. The variation in the running time with respect to the size and order is consistent with the complexity suggested by (3.8).

Acknowledgments. A MATLAB implemention of the trisection algorithm is available on the author's web page.[2] Most of this work was completed during the author's Ph.D. study at New York University and some part was completed during the author's visit to the numerical analysis and modeling group at the Technical University of Berlin. The author is grateful to Michael Overton and Daniel Kressner for reading a preliminary version of this paper, Volker Mehrmann for pointing out the importance of the even-odd matrix polynomials and insightful discussions regarding preserving the even-odd structure, and two anonymous referees.

REFERENCES

[1] D. BOLEY, *Estimating the sensitivity of the algebraic structure of pencils with simple eigenvalue estimates*, SIAM J. Matrix Anal. Appl., 11 (1990), pp. 632–643.
[2] J. V. BURKE, A. S. LEWIS, AND M. L. OVERTON, *Optimization and pseudospectra, with applications to robust stability*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 80–104.
[3] J. V. BURKE, A. S. LEWIS, AND M. L. OVERTON, *Pseudospectral components and the distance to uncontrollability*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 350–361.
[4] R. BYERS, *A bisection method for measuring the distance of a stable matrix to the unstable matrices*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 875–881.
[5] R. BYERS, *Detecting nearly uncontrollable pairs*, in Proceedings of the International Symposium MTNS-89, vol. III, Amsterdam, 1989, Progr. Systems Control Theory 5, M. A. Kaashoek, J. H. van Schuppen, and A. C. M. Ran, eds., Birkhäuser Boston, Boston, 1990, pp. 447–457.
[6] R. BYERS, *The descriptor controllability radius*, in Proceedings of the International Symposium MTNS-93, Regensburg, Germany, vol. II, U. Helmke, R. Mennicken, and H. Saurer, eds., Akademie Verlag, Berlin, 1994, pp. 85–88.

[2] http://www.cs.nyu.edu/mengi/robust_stability/dist_uncont_higher.html

[7] K-W. Chu, *Controllability of descriptor systems*, Internat. J. Control, 46 (1987), pp. 1761–1770.
[8] K-W. Chu, *A controllability condensed form and a state feedback pole assignment algorithm or descriptor systems*, IEEE Trans. Automat. Control, 33 (1988), pp. 366–370.
[9] G. E. Dullerud and F. Paganini, *A Course in Robust Control Theory: A Convex Approach*, Springer-Verlag, New York, 2000.
[10] R. Eising, *Between controllable and uncontrollable*, Systems Control Lett., 4 (1984), pp. 263–264.
[11] M. Gao and M. Neumann, *A global minimum search algorithm for estimating the distance to uncontrollability*, Linear Algebra Appl., 188/189 (1993), pp. 305–350.
[12] L. Gaul and N. Wagner, *Eigenpath Dynamics of Nonconservative Mechanical Systems Such as Disc Brakes*, in IMAC XXII, Dearborn, MI, 2004.
[13] Y. Genin, R. Stefan, and P. Van Dooren, *Real and complex stability radii of polynomial matrices*, Linear Algebra Appl., 351/352 (2002), pp. 381–410.
[14] M. Gu, *New methods for estimating the distance to uncontrollability*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 989–1003.
[15] M. Gu, E. Mengi, M. L. Overton, J. Xia, and J. Zhu, *Fast methods for estimating the distance to uncontrollability*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 477–502.
[16] C. He, *Estimating the distance to uncontrollability: A fast method and a slow one*, Systems Control Lett., 26 (1995), pp. 275–281.
[17] N. J. Higham, D. S. Mackey, and F. Tisseur, *The conditioning of linearizations of matrix polynomials*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1005–1028.
[18] N. J. Higham and F. Tisseur, *More on pseudospectra for polynomial eigenvalue problems and applications in control theory*, Linear Algebra Appl., 351/352 (2002), pp. 435–453.
[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
[20] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann, *Structured polynomial eigenvalue problems: Good vibrations from good linearizations*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1029–1051.
[21] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann, *Vector spaces of linearizations for matrix polynomials*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 971–1004.
[22] C. C. Paige, *Properties of numerical algorithms relating to computing controllability*, IEEE Trans. Automat. Control, 26 (1981), pp. 130–138.
[23] F. Tisseur and N. J. Higham, *Structured pseudospectra for polynomial eigenvalue problems with applications*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 187–208.
[24] M. Wicks and R. A. DeCarlo, *Computing the distance to an uncontrollable system*, IEEE Trans. Automat. Control, 36 (1991), pp. 39–49.