SIAM J. MATRIX ANAL. APPL.
Vol. 41, No. 4, pp. 1505–1527




# NEAREST MATRIX POLYNOMIALS WITH A SPECIFIED ELEMENTARY DIVISOR*

BISWAJIT DAS† AND SHREEMAYEE BORA†

**Abstract.** The problem of finding the distance from a given $n \times n$ matrix polynomial of degree $k$ to the set of matrix polynomials having the elementary divisor $(\lambda - \lambda_0)^j$, $j \geqslant r$, for fixed scalar $\lambda_0$ and $2 \leqslant r \leqslant kn$ is considered. For regular matrix polynomials the problem is shown to be equivalent to finding minimal structure-preserving perturbations such that a certain block Toeplitz matrix becomes suitably rank deficient. This is then used to characterize the distance via two different optimizations. The first one shows that if $\lambda_0$ is not already an eigenvalue of the matrix polynomial, then the problem is equivalent to computing a generalized notion of a structured singular value. The distance is computed via algorithms like BFGS and MATLAB's `globalsearch` algorithm from the second optimization. Upper and lower bounds of the distance are also derived, and numerical experiments are performed to compare them with the computed values of the distance.

**Key words.** matrix polynomial, elementary divisor, Jordan chain, Toeplitz matrix

**AMS subject classifications.** 15A18, 65F35, 65F15, 47A56, 15B05, 47J10

**DOI.** 10.1137/19M1286505

## 1. Introduction.

Given a matrix polynomial $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ of degree $k$, where $A_i$, $i = 0, \ldots, k$ are $n \times n$ real or complex matrices, this paper investigates the distance from $P(\lambda)$ to a nearest matrix polynomial with an elementary divisor $(\lambda - \lambda_0)^j, j \geqslant r$, for a given $\lambda_0 \in \mathbb{C}$ and integer $r$, where $2 \leqslant r \leqslant kn$.

Although the problem is considered only for finite values of $\lambda_0$, the analysis also covers the infinite case, which is equivalent to the reversal polynomial defined by $\operatorname{rev} P(\lambda) := \sum_{i=0}^{k} \lambda^i A_{k-i}$ having an elementary divisor $\lambda^j, j \geqslant r$. In particular, in such cases the distance under consideration is important from the point of view of control theory for the following reasons. If $P(\lambda) = \lambda A_1 - A_0$ is a regular matrix pencil, then there exist invertible matrices $E$ and $F$ such that

$$EP(\lambda)F = \lambda \begin{bmatrix} I_{p_1} & 0 \\ 0 & N_{p_2} \end{bmatrix} + \begin{bmatrix} J_{p_1} & 0 \\ 0 & I_{p_2} \end{bmatrix}, p_1 + p_2 = n,$$

where $J_{p_1}$ is a block diagonal matrix containing all the Jordan blocks associated with finite eigenvalues of $P(\lambda)$ and $N_{p_2}$ is a nilpotent matrix of size $p_2$ with nilpotent Jordan blocks on the diagonal. The block $N_{p_2}$ arises in the decomposition only if $P(\lambda)$ has an eigenvalue at $\infty$. The matrix pencil on the right-hand side of the above decomposition is called the Weierstrass canonical form of the pencil $P(\lambda)$. If $\infty$ is an eigenvalue, then the smallest positive integer $\nu$ such that $N_{p_2}^{\nu} = 0$ is called the index of the pencil. If $\nu > 1$, then this is equivalent to the existence of a Jordan chain of length at least 2 at $\infty$ for $P(\lambda)$ or equivalently an elementary divisor $\lambda^j, j \geqslant 2$, for $\operatorname{rev} P(\lambda)$. In such a case the associated differential algebraic equation $A_1\dot{x}(t) = A_0x(t) + Bu(t)$ may not have any solution for certain choices of initial conditions unless the controller $u(t)$ is sufficiently smooth. In fact, the larger the length of a Jordan chain at $\infty$, the greater

*Received by the editors September 10, 2019; accepted for publication (in revised form) by H. Fassbender June 29, 2020; published electronically October 1, 2020.
https://doi.org/10.1137/19M1286505
**Funding:** The work of the first author is supported by the MHRD, Government of India.
†Department of Mathematics, Indian Institute of Technology Guwahati, Guwahati 781039, Assam, India (biswajit.das@iitg.ac.in, shbora@iitg.ac.in).

are the smoothness requirements on $u(t)$. In particular, for dynamical systems arising from matrix pencils as above to be stable or asymptotically stable, it is necessary that the matrix pencil has index at most one. Moreover, for the stability of such systems, it is necessary that the purely imaginary eigenvalues of $P(\lambda)$ are not associated with Jordan chains of length 2 or more. For more details, see [4, 26, 8] and references therein.

It is well known that arbitrarily small perturbations to matrix pencils with $\lambda_0$ as an eigenvalue of algebraic multiplicity $r$ can result in a matrix pencil having an elementary divisor $(\lambda - \lambda_0)^r$. In fact, this result can also be extended to all matrix polynomials, a proof of which is provided in section 3. Due to this fact, the distance problem under consideration is also equivalent to finding the distance to a nearest matrix polynomial with an eigenvalue at $\lambda_0$ with algebraic multiplicity at least $r$. This problem has been considered in the literature in various forms. The distance to a nearest matrix polynomial with a prescribed multiple eigenvalue is considered in [21], and bounds on the distance are obtained under certain conditions. In [22] this work is extended to find the distance from a given matrix polynomial to a nearest matrix polynomial with a specified eigenvalue of algebraic multiplicity at least $r$ and a Jordan chain of length at most $k$ and an upper bound of the distance to a nearest matrix polynomial with a specified eigenvalue of algebraic multiplicity at least $r$. The latter is done by constructing a perturbation to the given matrix polynomial which has the desired feature. However, the construction is possible under certain conditions. The results are extended to matrix polynomials in [17], where a similar construction is made to find an upper bound on the distance to a nearest matrix polynomial with specified eigenvalues of desired multiplicities. The distance from an $n \times m$ matrix pencil $A + \lambda B$ with $n \geqslant m$ to a nearest matrix pencil having specified eigenvalues such that the sum of their multiplicities is at least $r$ is considered in [19]. Under the assumption that $\operatorname{rank} B \geqslant r$ and only $A$ is perturbed, the distance is shown to be given by a certain singular value optimization under certain conditions. These ideas are extended in [16] to find the same distance from a square matrix polynomial that has no infinite eigenvalues. Under certain conditions similar to those in [19], a singular value optimization is shown to be equal to the distance when only the constant coefficient of the matrix polynomial is perturbed. A lower bound is found for the general case when all coefficient matrices are perturbed. The techniques are further extended to find the same distance for more general nonlinear eigenvalue problems in [15].

The analysis of the distance problem in this paper has several key features. First, the stated distance is considered for a square matrix polynomial that is either regular or singular, and perturbations are considered on all the coefficient matrices of the polynomial. Note that with the exception of [19], where a rectangular matrix pencil is considered, in all other works in the literature the matrix pencil or polynomial is assumed to be regular. However, [19] considers perturbations only to the constant coefficient matrix of the pencil. As a singular matrix polynomial $P(\lambda)$ is arbitrarily close to a regular matrix polynomial with an elementary divisor $(\lambda - \lambda_0)^j, j \geqslant r$ (see section 3), it is possible to assume that the matrix polynomial $P(\lambda)$ is regular in the rest of the paper. A necessary and sufficient condition is obtained for $P(\lambda)$ to have $\lambda_0$ as an eigenvalue of algebraic multiplicity at least $r$. Due to this, it is possible to show that finding the stated distance is equivalent to finding a structure-preserving perturbation such that the nullity of a certain block Toeplitz matrix is at least $r$. This leads to a lower bound on the distance and allows for several characterizations of the distance in terms of optimization problems. Under the mild assumption that $\lambda_0$ is not an eigenvalue of $P(\lambda)$, for different choices of norms it is established that computing the distance from $P(\lambda)$ to a nearest matrix polynomial with an elementary divisor $(\lambda - \lambda_0)^j, j \geqslant r$, is equivalent to computing a generalized version of a structured singular value or $\mu$-value. It is well known that the $\mu$-value computation problem can be an NP-hard problem [3]. Due to the form of the generalized $\mu$-value, these results may throw light on the computational complexity of the distance problem. The characterization in terms of generalized $\mu$-values also yields a lower bound on the distance. Alternatively, the distance is characterized by another optimization problem which is computed via BFGS and MATLAB's `globalsearch` algorithm. This also results in an upper bound on the distance. A special case for which the solution of the distance problem has a closed-form expression is also discussed. Finally, computed values of the distance via BFGS and MATLAB's `globalsearch` are compared with upper and lower bounds.

**2. Preliminaries.** Standard notations are followed throughout the paper. The set of $n \times n$ complex matrices is denoted by $\mathbb{C}^{n\times n}$. The $i$th singular value of a matrix $A$ is denoted by $\sigma_i(A)$. Also, the smallest singular value of $A$ is denoted by $\sigma_{\min}(A)$.

Consider the matrix polynomial of degree $k$ of the form $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$, $A_i \in \mathbb{C}^{n\times n}$. There exist two $n \times n$ matrix polynomials $E(\lambda)$ and $F(\lambda)$ with nonzero determinants independent of $\lambda$ such that $P(\lambda) = E(\lambda)D(\lambda)F(\lambda)$, where

$$D(\lambda) = \begin{bmatrix} d_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & d_t(\lambda) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

is a diagonal matrix with monic scalar polynomials $d_i(\lambda)$ such that $d_{i-1}(\lambda)$ divides $d_i(\lambda)$. This is called the Smith form of $P(\lambda)$. Important concepts associated with $P(\lambda)$ may be defined via its Smith form. The nonzero diagonal elements $d_1(\lambda), \ldots, d_t(\lambda)$ are called the invariant polynomials of $P(\lambda)$. The number of such invariant polynomials is the normal rank of $P(\lambda)$. The polynomial $P(\lambda)$ is said to be regular if its normal rank is equal to its size $n$. Otherwise, it is said to be a nonregular or singular matrix polynomial.

Each invariant polynomial may be written as a product of linear factors

$$d_i(\lambda) = (\lambda - \lambda_{i1})^{c_{i1}} \ldots (\lambda - \lambda_i q_i)^{c_{iq_i}},$$

where $\lambda_{i1}, \ldots, \lambda_{iq_i}$ are distinct complex numbers and $c_{i1}, \ldots, c_{iq_i}$ are positive integers. The factors $(\lambda - \lambda_{ij})^{c_{ij}}$ are called elementary divisors of $P(\lambda)$. Any $\lambda_0 \in \mathbb{C}$ is a finite eigenvalue of $P(\lambda)$ if $(\lambda - \lambda_0)^c$ is an elementary divisor of $P(\lambda)$ for some positive integer $c$. The algebraic multiplicity of $\lambda_0$ as an eigenvalue of $P(\lambda)$ is the sum of all the powers of the term $(\lambda - \lambda_0)$ in all the invariant polynomials, and the geometric multiplicity of $\lambda_0$ as an eigenvalue is the number of invariant polynomials which have $(\lambda - \lambda_0)^c$ as a factor. Clearly, if the matrix polynomial $P(\lambda)$ is regular, then the eigenvalues of $P(\lambda)$ are the roots of $\det(P(\lambda))$ with algebraic multiplicity equal to the multiplicity of the root.

Having an elementary divisor $(\lambda-\lambda_0)^r$ is also equivalent to the existence of vectors $x_0,\ldots,x_{r-1}\in\mathbb{C}^n$, $x_0\neq 0$, satisfying the equations

$$\sum_{i=0}^{p}\frac{1}{i!}P^i(\lambda_0)x_{p-i}=0, p=0,\ldots,r-1,$$

where $P^i(\lambda)$ denotes the $i$th derivative of $P(\lambda)$ with respect to $\lambda$. The vectors $x_0,\ldots,x_{r-1}$ are said to form a Jordan chain of length $r$ of $P(\lambda)$ corresponding to $\lambda_0$.

Given any $n\times n$ matrix pencil $L(\lambda)=A-\lambda E$, there exist two $n\times n$ invertible matrices $P$ and $Q$ such that $P(A-\lambda E)Q$ is a block diagonal matrix pencil, with each diagonal block being one of the following forms: $\lambda I_q-J_q(\alpha)\in\mathbb{C}^{q\times q}, \lambda J_q(0)-I_q\in\mathbb{C}^{q\times q}, \lambda G_q-F_q\in\mathbb{C}^{q\times(q+1)}, \lambda G_q^T-F_q^T\in\mathbb{C}^{(q+1)\times q}$, where

$$J_q(\alpha)=\begin{bmatrix}\alpha & 1 & & \\ & \alpha & \ddots & \\ & & \ddots & 1\\ & & & \alpha\end{bmatrix}, F_q=\begin{bmatrix}1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0\end{bmatrix}, G_q=\begin{bmatrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & 0 & 1\end{bmatrix}$$

for some $\alpha\in\mathbb{C}$. The blocks $\lambda I_q-J_q(\alpha)$, $\lambda J_q(0)-I_q$, $\lambda G_q-F_q$, and $\lambda G_q^T-F_q^T$ correspond to a finite eigenvalue $\alpha$, the infinite eigenvalue, right singular blocks, and left singular blocks, respectively. This is called the Kronecker canonical form (KCF) of the pencil. For a matrix pencil, having an elementary divisor $(\lambda-\lambda_0)^r$ is equivalent to having a block $\lambda I_r-J_r(\lambda_0)$ in its KCF. Also, clearly the algebraic multiplicity of $\lambda_0$ as an eigenvalue of $L(\lambda)$ is the sum of all the sizes of the blocks corresponding to $\lambda_0$ and its geometric multiplicity is the number of such blocks.

The normwise distance of $P(\lambda)$ to the set of all matrix polynomials having a elementary divisor $(\lambda-\lambda_0)^j$, where $j\geqslant r$ will be considered with respect to the norms

$$\delta_F(P,\lambda_0,r)=\inf\left\{|||\Delta P|||_F|P+\Delta P \text{ has an elementary divisor } (\lambda-\lambda_0)^j, j\geqslant r\right\},$$

$$\delta_2(P,\lambda_0,r)=\inf\left\{|||\Delta P|||_2|P+\Delta P \text{ has an elementary divisor } (\lambda-\lambda_0)^j, j\geqslant r\right\},$$

where $|||P|||_F:=(\sum_{i=0}^k\|A_i\|_F^2)^{1/2}$ and $|||P|||_2:=\|[A_0\cdots A_k]\|_2$, $\|\cdot\|_F$ and $\|\cdot\|_2$ being the Frobenius and 2-norms on matrices, respectively. Also, the matrix polynomial $\Delta P(\lambda)=\sum_{i=0}^k\lambda^i\Delta A_i$ is such that any of the coefficient matrix $\Delta A_i$ may be zero.

Due to the importance of the case $\lambda_0=0$ in practical applications and also because the results for this case involve expressions that are relatively simpler than the general case, in many instances initially the results for this special case are obtained and then extend for other choices of $\lambda_0$. The following lemma will be useful for making these extensions.

Lemma 2.1. *Given any $n\times n$ matrix polynomial $Q(\lambda)=\sum_{i=0}^k\lambda^iB_i$ and $\lambda_0\in\mathbb{C}$,*

$$\begin{bmatrix}Q(\lambda_0) & Q'(\lambda_0) & \cdots & \frac{1}{p!}Q^p(\lambda_0)\end{bmatrix}=\begin{bmatrix}B_0 & \cdots & B_k\end{bmatrix}M(\lambda_0;r),$$

*where $M(\lambda_0;r):=H(\lambda_0)\otimes I_n$, $H(\lambda_0)$ being a $(k+1)\times(p+1)$ matrix with $p=\min\{r,k\}$, given by*

$$H(\lambda_0) = \begin{bmatrix} 1 & & & & \\ \lambda_0 & 1 & & & \\ \lambda_0^2 & 2\lambda_0 & 1 & & \\ \lambda_0^3 & 3\lambda_0^2 & 3\lambda_0 & \ddots & \\ \vdots & \vdots & \vdots & \dots & 1 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \lambda_0^k & k\lambda_0^{k-1} & \frac{k!\lambda_0^{k-2}}{(k-2)!2!} & \dots & \frac{k!\lambda_0^{k-p}}{(k-p)!p!} \end{bmatrix}.$$

*Proof.* The proof follows from the fact that for each $i = 1, \dots, p+1$, $\frac{Q^{(i-1)}(\lambda_0)}{(i-1)!}$ is given by the product of $\begin{bmatrix} B_0 & \cdots & B_k \end{bmatrix}$ with $H_i(\lambda_0) \otimes I_n$, $H_i(\lambda_0)$ being the $i$th column of $H(\lambda_0)$. ∎

## 3. Polynomials for which the distance is zero.

Given an $n \times n$ matrix polynomial $P(\lambda)$ of degree $k$, it is interesting to identify cases when the distance to a nearest matrix polynomial with an elementary divisor $(\lambda - \lambda_0)^j$, $j \geqslant r$ (for any given $r$ satisfying $2 \leqslant r \leqslant kn$) is zero. One such situation is obviously the case that $k = 1$ and $\lambda_0$ is an eigenvalue of $P(\lambda)$ of algebraic multiplicity at least $r$. The main result of this section is a proof of the fact that the distance of interest is zero in two cases. The first one is that $P(\lambda)$ is a matrix polynomial of *any* degree and $\lambda_0$ is an eigenvalue of $P(\lambda)$ of algebraic multiplicity at least $r$. The second one is that $P(\lambda)$ is a singular matrix polynomial. For the matrix pencils, this follows from the next theorem and for the matrix polynomials from Theorem 3.2.

THEOREM 3.1. *For a given $n \times n$ matrix pencil $L(\lambda) = A - \lambda E$,*

(a) *if $L(\lambda)$ is regular and algebraic multiplicity of $\lambda_0$ as an eigenvalue of $L(\lambda)$ is $j$, $1 \leqslant j \leqslant n$, then it is arbitrarily close to a regular matrix pencil having an elementary divisor $(\lambda - \lambda_0)^j$;*

(b) *if $L(\lambda)$ is singular, then it is arbitrarily close to a regular matrix pencil having an elementary divisor $(\lambda - \lambda_0)^n$.*

*Proof.*

(a) The proof of this part is obvious owing to the structure of the Kronecker canonical form of a matrix pencil having $\lambda_0$ as an eigenvalue of algebraic multiplicity $j$.

(b) Suppose $L(\lambda)$ is an $n \times n$ singular matrix pencil. We have to show that there is a regular pencil with an elementary divisor of the form $(\lambda - \lambda_0)^n$ that is arbitrarily close to $L(\lambda)$. This can be proved by using parts (iii), (iv), and (vi) of [9, Theorem III]. In particular, by repeated application of part (iii) and/or part (iv), it may be shown that there exists a singular pencil with only singular blocks in its Kronecker canonical form that is arbitrarily close to $L(\lambda)$. Part (vi) of the same theorem then implies that this singular pencil is arbitrarily close to a regular pencil with an eigenvalue of algebraic multiplicity $n$ at $\lambda_0$, and the proof follows from part (a). ∎

For a detailed proof of Theorem 3.1 using elementary perturbation theory, see [5].

The above result may be extended to matrix polynomials by considering the first companion linearization

$$C_1(\lambda) = \lambda \begin{bmatrix} A_k & & & \\ & I_n & & \\ & & \ddots & \\ & & & I_n \end{bmatrix} + \begin{bmatrix} A_{k-1} & A_{k-2} & \cdots & A_0 \\ -I_n & & & 0 \\ & \ddots & & \\ & & -I_n & 0 \end{bmatrix}$$

of $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$. It is an example of a block Kronecker linearization as introduced in [7], where it was shown that if $L(\lambda)$ is a block Kronecker linearization of $P(\lambda)$ and $\Delta L(\lambda)$ is a pencil of the same size as $L(\lambda)$ with $|||\Delta L|||_F < \epsilon$, for some sufficiently small $\epsilon > 0$, then $L(\lambda) + \Delta L(\lambda)$ is a strong linearization of $P(\lambda) + \Delta P(\lambda)$ such that $|||\Delta P|||_F < C\epsilon$ for some positive constant $C$. Due to this result, the following theorem is an immediate consequence of Theorem 3.1.

THEOREM 3.2. *For a given $n \times n$ matrix polynomial $P(\lambda)$ of degree $k$,*

(a) *if $P(\lambda)$ is regular and algebraic multiplicity of $\lambda_0$ as an eigenvalue of $P(\lambda)$ is $j$, $1 \leqslant j \leqslant kn$, then it is arbitrarily close to a regular matrix polynomial having an elementary divisor $(\lambda - \lambda_0)^j$;*

(b) *if $P(\lambda)$ is singular, then it is arbitrarily close to a regular matrix polynomial having an elementary divisor $(\lambda - \lambda_0)^{kn}$.*

Note that Theorem 3.2 is also implied by the work done in [6], which uses the above-mentioned result in [7] to construct the closure hierarchy graphs of matrix polynomials from those of their first companion linearizations.

From part (a) of Theorem 3.2, it is now clear that if the algebraic multiplicity of $\lambda_0$ as an eigenvalue of $P(\lambda)$ is at least $r$, then the distance to a nearest matrix polynomial having an elementary divisor $(\lambda - \lambda_0)^j$, $j \geqslant r$, is zero. From part (b) of the theorem it follows that this distance is zero if $P(\lambda)$ is a singular matrix polynomial because $r \leqslant kn$. In view of this, it is now possible to assume without loss of generality that the distances $\delta_s(P, \lambda_0, r)$ for $s = 2$ or $F$ are being computed for a regular matrix polynomial $P(\lambda)$, which does not have $\lambda_0$ as an eigenvalue of algebraic multiplicity $r$ or more. This also has the effect of removing the uncertainty that was earlier associated with the situation that perturbations being made to the matrix polynomial for the desired objectives could result in a singular matrix polynomial.

## 4. A characterization via block Toeplitz matrices.

One of the aims of this work is to show that for appropriate choices of norms, computing the distance to a matrix polynomial with an elementary divisor $(\lambda - \lambda_0)^j, j \geqslant r$, is equivalent to finding a structured singular value or generalized $\mu$-value. The next result is an important step in this direction. Since the expression for the optimization is more aesthetic if $r$ is replaced by $r+1$, in the rest of the paper the distance is considered in the form $\delta_s(P, \lambda_0, r+1)$, where $s = 2$ or $F$. The following set will be frequently used:

$$\Gamma := \{[\gamma_1 \cdots \gamma_r] : \gamma_i > 0, 1 \leqslant i \leqslant r\}. \tag{4.1}$$

For any $\gamma = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_r \end{bmatrix} \in \Gamma$ given by (4.1) and $\alpha \in \mathbb{C}$, let $T_\gamma(Q, \alpha)$ be a function from the set of all $n \times n$ matrix polynomial $Q(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ to the set of $(r+1)n \times (r+1)n$ matrices defined by

$$T_\gamma(Q, \alpha) := \begin{bmatrix} Q(\alpha) & & & & \\ \gamma_1 Q'(\alpha) & Q(\alpha) & & & \\ \gamma_1\gamma_2 \frac{Q''(\alpha)}{2!} & \gamma_2 Q'(\alpha) & Q(\alpha) & & \\ \vdots & \vdots & \ddots & \ddots & \\ \left(\prod_{i=1}^{r}\gamma_i\right)\frac{Q^r(\alpha)}{r!} & \left(\prod_{i=2}^{r}\gamma_i\right)\frac{Q^{r-1}(\alpha)}{(r-1)!} & \cdots & \gamma_r Q'(\alpha) & Q(\alpha) \end{bmatrix}. \tag{4.2}$$

Let

$$T(Q, \alpha) := T_\gamma(Q, \alpha) \text{ with } \gamma = [1 \cdots 1]. \tag{4.3}$$

Block Toeplitz matrices of the form $T(Q,\alpha)$ have been used to analyze the left and right Jordan chain structures of $Q(\lambda)$ at its eigenvalues [1, 2, 10, 11, 13, 24, 25]. In particular, in [25] it was proved that the algebraic multiplicity, say, $m(P,\lambda_0)$, of an eigenvalue $\lambda_0$ of an $n\times n$ regular matrix polynomial $P(\lambda)$ satisfies

$$m(P,\lambda_0) = nl - \operatorname{rank} T^{(l-1)}(P,\lambda_0), \tag{4.4}$$

where

$$T^{(q)}(P,\lambda_0) := \begin{bmatrix} P(\lambda_0) & & & & \\ P'(\lambda_0) & P(\lambda_0), & & & \\ \frac{P''(\lambda_0)}{2!} & P'(\lambda_0) & P(\lambda_0) & & \\ \vdots & \vdots & \ddots & \ddots & \\ \frac{P^{q}(\lambda_0)}{q!} & \frac{P^{q-1}(\lambda_0)}{(q-1)!} & \cdots & P'(\lambda_0) & P(\lambda_0) \end{bmatrix}, q = 0, 1, \ldots$$

and $l \geqslant g$, $g$ being the length of the largest Jordan chain of $P(\lambda)$ associated with $\lambda_0$. As a consequence, we have the following result.

THEOREM 4.1. *Let $P(\lambda)$ be an $n\times n$ matrix polynomial of degree $k$, $\lambda_0\in\mathbb{C}$ and $r$ be any real number satisfying $1\leqslant r\leqslant kn-1$. Then $\lambda_0\in\mathbb{C}$ is an eigenvalue of $P(\lambda)$ of algebraic multiplicity at least $r+1$ if and only if $T_\gamma(P,\lambda_0)$ as defined by* (4.2) *has rank at most $(r+1)(n-1)$.*

*Proof.* Let $\lambda_0$ be an eigenvalue of $P(\lambda)$ of algebraic multiplicity $r+1$. Suppose there are $p$ Jordan chains

$$\{x_{11}, x_{12}, \ldots, x_{1k_1}\}, \{x_{21}, x_{22}, \ldots, x_{2k_2}\}, \ldots, \{x_{p1}, x_{p2}, \ldots, x_{pk_p}\}$$

of $P(\lambda)$ corresponding to $\lambda_0$, where $x_{ij}\in\mathbb{C}^n$ for $i=1,\ldots,p$, $j=1,\ldots,k_i$, satisfying $\sum_{i=1}^{p} k_i \geqslant r+1$. The $i$th Jordan chain satisfies the following equations for $i=1,\ldots,p$:

$$\begin{aligned}
&P(\lambda_0)x_{i1} = 0\\
&P(\lambda_0)x_{i2} + P'(\lambda_0)x_{i1} = 0\\
&P(\lambda_0)x_{i3} + P'(\lambda_0)x_{i2} + \frac{P''(\lambda_0)}{2!}x_{i1} = 0\\
&\vdots\\
&P(\lambda_0)x_{ik_i} + P'(\lambda_0)x_{i(k_i-1)} + \frac{P''(\lambda_0)}{2!}x_{i(k_i-2)} + \cdots + \frac{P^{k_i-1}(\lambda_0)}{(k_i-1)!}x_{i1} = 0.
\end{aligned}$$

It may be assumed that $\{x_{11}, x_{21}, \ldots, x_{p1}\}$ is a linear independent set. Clearly, the $i$th Jordan chain contributes the $k_i$ vectors

$$\begin{bmatrix} 0\\ \vdots\\ 0\\ \frac{x_{i1}}{\gamma_{(r-k_i+2)}\cdots\gamma_r}\\ \vdots\\ \frac{x_{i(k_i-3)}}{\gamma_{r-2}\gamma_{r-1}\gamma_r}\\ \frac{x_{i(k_i-2)}}{\gamma_{r-1}\gamma_r}\\ \frac{x_{i(k_i-1)}}{\gamma_r}\\ x_{ik_i}\end{bmatrix}, \begin{bmatrix} 0\\ 0\\ \vdots\\ 0\\ \frac{x_{i1}}{\gamma_{(r-k_i+3)}\cdots\gamma_r}\\ \vdots\\ \frac{x_{i(k_i-3)}}{\gamma_{r-1}\gamma_r}\\ \frac{x_{i(k_i-2)}}{\gamma_r}\\ x_{i(k_i-1)}\end{bmatrix}, \begin{bmatrix} 0\\ 0\\ 0\\ \vdots\\ 0\\ \frac{x_{i1}}{\gamma_{(r-k_1+4)}\cdots\gamma_r}\\ \vdots\\ \frac{x_{i(k_i-3)}}{\gamma_r}\\ x_{i(k_i-2)}\end{bmatrix}, \ldots, \begin{bmatrix} 0\\ 0\\ 0\\ 0\\ 0\\ 0\\ \vdots\\ 0\\ x_{i1}\end{bmatrix}$$

of length$(r+1)n$ to the null space $N(T_\gamma(P,\lambda_0))$ of $T_\gamma(P,\lambda_0)$ for $i=1,\ldots,p$. All the above vectors are linearly independent, as $\{x_{11},x_{21},\ldots,x_{p1}\}$ are linearly independent. Hence, the nullity of $T_\gamma(P,\lambda_0)$ is at least $r+1$ so that $\operatorname{rank} T_\gamma(P,\lambda_0) \leqslant (n-1)(r+1)$.

To prove the converse, suppose that $\operatorname{rank} T_\gamma(P,\lambda_0) \leqslant (n-1)(r+1)$. Noting that $\operatorname{rank} T_\gamma(P,\lambda_0) = \operatorname{rank} T^{(r)}(P,\lambda_0)$ for all $\gamma \in \Gamma$, this may be rephrased as

$$n(r+1) - \operatorname{rank} T^{(r)}(P,\lambda_0) \geqslant r+1.$$

Also, clearly $\lambda_0$ is an eigenvalue of $P(\lambda)$. We assume without loss of generality that the lengths of the Jordan chains of $P(\lambda)$ associated with $\lambda_0$ are at most $r$. By (4.4), the algebraic multiplicity of $\lambda_0$ is $n(r+1) - \operatorname{rank} T^{(r)}(P,\lambda_0)$. Since this is at least $r+1$, the proof follows. ∎

*Remark* 4.2. Theorem 4.1 is established in [21] for the particular case that $P(\lambda)$ has an eigenvalue of algebraic multiplicity 2. Under the assumption that the leading coefficient matrix is full rank, another characterization of a matrix polynomial $P(\lambda)$ having a specified eigenvalue of algebraic multiplicity $r$ is obtained in [16] via the rank of a different matrix involving $r(r+1)/2$ parameters.

In view of part (a) of Theorem 3.2, the following corollary of Theorem 4.1 is immediate.

COROLLARY 4.3. *Given any $n\times n$ matrix polynomial $P(\lambda)$, consider the collection $\mathcal{S}(P,\lambda_0)$ of all $n\times n$ matrix polynomials $(\Delta P)(\lambda) := \sum_{i=0}^k \lambda^i \Delta A_i$ such that the block Toeplitz matrices $T_\gamma(P+\Delta P,\lambda_0)$ as defined in (4.2) with $\gamma = [1\cdots 1]^T$ have rank at most $(r+1)(n-1)$. For any choice of norm $|||\cdot|||$, the distance to a nearest matrix polynomial with an elementary divisor $(\lambda-\lambda_0)^j, j \geqslant r+1$, is given by* $\inf\{|||\Delta P||| : \Delta P(\lambda) \in \mathcal{S}(P,\lambda_0)\}$.

## 5. The distance as the reciprocal of a generalized $\mu$-value.

Corollary 4.3 implies that for any given choice of norm, finding the distance from $P(\lambda)$ to a nearest matrix polynomial with an elementary divisor $\lambda^{r+1}$ is equivalent to finding the smallest structure-preserving perturbation to the block Toeplitz matrix $T(P,0)$ so that the rank of the perturbed matrix is at most $(r+1)(n-1)$. This fact is used in this section to show that if $\lambda_0 \in \mathbb{C}$ is not already an eigenvalue of $P(\lambda)$, then computing the distance from $P(\lambda)$ to a nearest matrix polynomial with the desired elementary divisor $(\lambda-\lambda_0)^j, j\geqslant r+1$, with respect to the norms $|||P|||_2$ and $|||P|||_F$ is the reciprocal of a generalized notion of a $\mu$-value. Towards this end, we consider a nonempty subset of $\mathbb{C}^{p\times q}$ as a perturbation class and introduce the notion of a $\mu$-value, also referred to as a structured singular value, and its generalized version.

DEFINITION 5.1 ($\mu$-value [20, 14]). *Let $S \subset \mathbb{C}^{p\times q}$ be a perturbation class, and let $\|.\|$ be a norm on $\mathbb{C}^{p\times q}$. The $\mu$-value of $M \in \mathbb{C}^{q\times p}$ with respect to $S$ and $\|.\|$ is*

$$\mu_{S,\|.\|}(M) := (\inf\{\|\Delta\| : \Delta \in S, 1 \in \sigma(\Delta M)\})^{-1}. \tag{5.1}$$

*If there is no such $\Delta \in S$, then $\mu_{S,\|.\|}(M) = 0$.*

The generalized $\mu$-value is now defined as follows.

DEFINITION 5.2 (generalized $\mu$-value). *Let $S \subset \mathbb{C}^{p\times q}$ be a perturbation class, and let $\|.\|$ be a norm on $\mathbb{C}^{p\times q}$. The generalized $\mu$-value of $M \in \mathbb{C}^{q\times p}$ with respect to $S$ and $\|.\|$ is defined as*

$$\mu^r_{S,\|.\|}(M) := (\inf\{\|\Delta\| : \Delta \in S, \operatorname{rank}(I-\Delta M) \leqslant p-r\})^{-1}. \tag{5.2}$$

*If there is no such $\Delta \in S$, then $\mu^r_{S,\|.\|}(M) = 0$.*

The above definition was first introduced in [12]. The following lemma provides a useful factorization of $T(Q,\lambda_0)$ as in (4.3). The proof of the lemma follows from direct multiplication of the stated factors and is therefore skipped.

LEMMA 5.3. *For a given positive integer $r$ and an $n\times n$ matrix polynomial $Q(\lambda)$ of degree $k$,*

$$T(Q,\lambda_0)=\left(I_{r+1}\otimes\begin{bmatrix}Q(\lambda_0) & Q'(\lambda_0) & \cdots & \frac{Q^{\min\{r,k\}}(\lambda_0)}{\min\{r,k\}!}\end{bmatrix}\right)E,$$

*where*

$$E=\begin{cases}\begin{bmatrix}E_1^T & E_2^T & \cdots & E_{r+1}^T\end{bmatrix}^T\otimes I_n & \text{if } r\leqslant k,\\ \begin{bmatrix}\tilde{E}_1^T & \tilde{E}_2^T & \cdots & \tilde{E}_{r+1}^T\end{bmatrix}^T\otimes I_n & \text{if } r>k\end{cases}$$

*such that $E_i$, $i=1,\dots,r,r+1$ are the $(r+1)\times(r+1)$ matrices,*

$$E_1=\begin{bmatrix}1 & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}, E_2=\begin{bmatrix} & 1 & & \\ 1 & & & \\ & & & \\ & & & \end{bmatrix},\dots, E_r=\begin{bmatrix} & & 1 & \\ & \iddots & & \\ 1 & & & \\ & & & \end{bmatrix}, E_{r+1}=\begin{bmatrix} & & & 1\\ & & 1 & \\ & \iddots & & \\ 1 & & & \end{bmatrix},$$

*and $\tilde{E}_i\in\mathbb{C}^{(k+1)\times(r+1)}$ are the first $k+1$ rows of $E_i$.*

The next theorem is the main result of this section.

THEOREM 5.4. *Let $P(\lambda)=\sum_{i=0}^{k}\lambda^i A_i$ be an $n\times n$ matrix polynomial of degree $k$ and $1\leqslant r<kn$. For $\Delta A_i\in\mathbb{C}^{n\times n}$, $i=0,\dots,k$, let $S_1$ be the perturbation class of all perturbations of the type $\left(I_{r+1}\otimes\begin{bmatrix}\Delta A_0 & \Delta A_1 & \cdots & \Delta A_k\end{bmatrix}\right)$ and $S_2$ be the perturbation class of all perturbations of the type $\left(I_{r+1}\otimes\begin{bmatrix}\Delta A_0 & \Delta A_1 & \cdots & \Delta A_{\min\{r,k\}}\end{bmatrix}\right)$. For any $\lambda_0\in\mathbb{C}$ which is not an eigenvalue of $P(\lambda)$, let $T(P,\lambda_0)$ be defined by (4.3) and $E$ and $M(\lambda_0;r)$ be as given in Lemma 5.3 and Lemma 2.1, respectively. Then*

$$\delta_2(P,\lambda_0,r+1)=\begin{cases}\left[\mu^{r+1}_{S_1,\|\cdot\|_2}\left((I_{r+1}\otimes M(\lambda_0;r))E\ (T(P,\lambda_0))^{-1}\right)\right]^{-1} & \text{if } \lambda_0\neq 0,\\ \left[\mu^{r+1}_{S_2,\|\cdot\|_2}\left(E\ (T(P,0))^{-1}\right)\right]^{-1} & \text{otherwise}\end{cases}$$

*and*

$$\delta_F(P,\lambda_0,r+1)=\begin{cases}\dfrac{\left[\mu^{r+1}_{S_1,\|\cdot\|_F}\left((I_{r+1}\otimes M(\lambda_0;r))E\ (T(P,\lambda_0))^{-1}\right)\right]^{-1}}{\sqrt{r+1}} & \text{if } \lambda_0\neq 0,\\ \dfrac{\left[\mu^{r+1}_{S_2,\|\cdot\|_F}\left(E\ (T(P,0))^{-1}\right)\right]^{-1}}{\sqrt{r+1}} & \text{otherwise}\end{cases}$$

*Proof.* Recall that $S(P,\lambda_0)$ is the collection of all $n\times n$ matrix polynomials $\Delta P(\lambda)=\sum_{i=0}^{k}\lambda^i\Delta A_i$ such that $T(P+\Delta P,\lambda_0)$ has rank at most $(r+1)(n-1)$. As $P(\lambda_0)$ is invertible, $-\Delta P(\lambda)\in S(P,\lambda_0)$ if and only if

$$\text{nullity}(I-T(\Delta P,\lambda_0)T(P,\lambda_0)^{-1})\geqslant r+1. \tag{5.3}$$

Due to Lemma 5.3 the above relation may be written as

$$\text{nullity}\left(I-\left(I_{r+1}\otimes\begin{bmatrix}\Delta P(\lambda_0) & \Delta P'(\lambda_0) & \cdots & \frac{\Delta P^p(\lambda_0)}{p!}\end{bmatrix}\right)E\ (T(P,\lambda_0))^{-1}\right)\geqslant r+1, \tag{5.4}$$

where $p = \min\{r, k\}$. By Lemma 2.1 this is equivalent to

$$\text{nullity}\left(I - \left(I_{r+1} \otimes \begin{bmatrix}\Delta A_0 & \Delta A_1 & \cdots & \Delta A_k\end{bmatrix} M(\lambda_0; r)\right) E\ (T(P, \lambda_0))^{-1}\right) \geqslant r + 1,$$

which may also be written as

$$\text{nullity}\left(I - (I_{r+1} \otimes [\Delta A_0 \cdots \Delta A_k])\,(I_{r+1} \otimes M(\lambda_0; r))\, E\ (T(P, \lambda_0))^{-1}\right) \geqslant r+1. \tag{5.5}$$

Then

$$\inf\{|||\Delta P|||_2 | \Delta P(\lambda) \in S(P, \lambda_0)\} = \left[\mu^{r+1}_{S_1, \|\cdot\|_2}\left((I_{r+1} \otimes M(\lambda_0; r)) E\ (T(P, \lambda_0))^{-1}\right)\right]^{-1}.$$

Similarly,

$$\inf\{|||\Delta P|||_F | \Delta P(\lambda) \in S(P, \lambda_0)\} = \frac{\left[\mu^{r+1}_{S_1, \|\cdot\|_F}\left((I_{r+1} \otimes M(\lambda_0; r)) E\ (T(P, \lambda_0))^{-1}\right)\right]^{-1}}{\sqrt{r+1}}.$$

The proof for the case $\lambda_0 \neq 0$ now follows from Corollary 4.3.

If $\lambda_0 = 0$, then (5.4) will be of the form

$$\text{nullity}\left(I - \left(I_{r+1} \otimes \begin{bmatrix}\Delta A_0 & \cdots & \Delta A_p\end{bmatrix}\right) E\ (T(P, 0))^{-1}\right) \geqslant r + 1. \tag{5.6}$$

Then setting $\Delta A_i = 0$ for $i = r + 1, \ldots, k$ if $r < k$,

$$\inf\{|||\Delta P|||_2 | \Delta P(\lambda) \in S(P, 0)\} = \left[\mu^{r+1}_{S_2, \|\cdot\|_2}\left(E\ (T(P, 0))^{-1}\right)\right]^{-1} \text{ and}$$

$$\inf\{|||\Delta P|||_F | \Delta P(\lambda) \in S(P, 0)\} = \frac{\left[\mu^{r+1}_{S_2, \|\cdot\|_F}\left(E\ (T(P, 0))^{-1}\right)\right]^{-1}}{\sqrt{r+1}}.$$

Hence, the proof follows from Corollary 4.3. ∎

**6. An alternative formulation of the distance as an optimization.** An alternative formulation for the distance $\delta_s(P, \lambda_0, r + 1)$ is obtained in this section for $s = 2$ or $F$.

THEOREM 6.1. *Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ be an $n \times n$ matrix polynomial of degree $k$. For a given integer $r$ such that $0 < r < kn$, consider the sets $\Gamma$ as in (4.1) and*

$$\mathbb{C}_0^{(r+1)n} := \{[x_0^T \cdots x_r^T]^T : x_i \in \mathbb{C}^n, i = 0, \ldots, r, x_0 \neq 0\}. \tag{6.1}$$

*Let $\mathbb{C}^{r,n}_{\mathcal{T},\Gamma}$ be the collection of all block Toeplitz-like matrices $X$ given by*

$$X = \begin{cases} \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_r \\ & \gamma_1 x_0 & \gamma_2 x_1 & \cdots & \gamma_r x_{r-1} \\ & & \gamma_1\gamma_2 x_0 & \cdots & \gamma_{r-1}\gamma_r x_{r-2} \\ & & & \ddots & \vdots \\ & & & & \left(\prod_{i=1}^{r} \gamma_i\right) x_0 \end{bmatrix} & \text{if } r \leqslant k, \\ \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_k & \cdots & x_r \\ & \gamma_1 x_0 & \gamma_2 x_1 & \cdots & \gamma_k x_{k-1} & \cdots & \gamma_r x_{r-1} \\ & & \gamma_1\gamma_2 x_0 & \cdots & \gamma_{k-1}\gamma_k x_{k-2} & \cdots & \gamma_{r-1}\gamma_r x_{r-2} \\ & & & \ddots & \vdots & \vdots & \vdots \\ & & & & \left(\prod_{i=1}^{k} \gamma_i\right) x_0 & \cdots & \left(\prod_{i=r-k+1}^{r} \gamma_i\right) x_{r-k} \end{bmatrix} & \text{otherwise,} \end{cases}$$

*where* $[x_0^T \cdots x_r^T]^T \in \mathbb{C}_0^{(r+1)n}$ *and* $[\gamma_1 \cdots \gamma_r] \in \Gamma$. *Then for* $s = 2$ *or* $F$,

$$\delta_s(P, 0, r+1) = \begin{cases} \inf_{X \in \mathbb{C}^{r,n}_{\mathcal{T},\Gamma}} \left\| \begin{bmatrix} A_0 & \cdots & A_r \end{bmatrix} X X^\dagger \right\|_s & \text{if } r \leqslant k, \\ \inf_{X \in \mathbb{C}^{r,n}_{\mathcal{T},\Gamma}} \left\| \begin{bmatrix} A_0 & \cdots & A_k \end{bmatrix} X X^\dagger \right\|_s & \text{otherwise}, \end{cases} \tag{6.2}$$

*and in general,*

$$\delta_s(P, \lambda_0, r+1) = \begin{cases} \inf_{X \in \mathbb{C}^{r,n}_{\mathcal{T},\Gamma}} \left\| \begin{bmatrix} P(\lambda_0) & \cdots & \frac{1}{r!}P^r(\lambda_0) \end{bmatrix} X (M(\lambda_0; r)X)^\dagger \right\|_s & \text{if } r \leqslant k, \\ \inf_{X \in \mathbb{C}^{r,n}_{\mathcal{T},\Gamma}} \left\| \begin{bmatrix} P(\lambda_0) & \cdots & \frac{1}{k!}P^k(\lambda_0) \end{bmatrix} X (M(\lambda_0; r)X)^\dagger \right\|_s & \text{otherwise}, \end{cases} \tag{6.3}$$

*where* $M(\lambda_0; r)$ *is as given in Lemma* 2.1.

*Proof.* Initially consider the case that $r \leqslant k$. Let $\Delta P(\lambda) = \sum_{i=0}^k \lambda^i \Delta A_i$ be any $n \times n$ matrix polynomial of degree $k$ such that $P(\lambda) + \Delta P(\lambda)$ is a regular matrix polynomial. Then $P(\lambda) + \Delta P(\lambda)$ has an elementary divisor $(\lambda - \lambda_0)^j$, where $j \geqslant r+1$ if and only if there exists vectors $x_0, x_1, \ldots, x_r \in \mathbb{C}^n$ with $x_0 \neq 0$ and $r$ positive scalars $\gamma_1, \ldots, \gamma_r$ such that

$$\begin{aligned} &(P + \Delta P)(\lambda_0)x_0 = 0 \\ &(P + \Delta P)(\lambda_0)x_1 + \gamma_1 (P + \Delta P)'(\lambda_0))x_0 = 0 \\ &(P + \Delta P)(\lambda_0)x_2 + \gamma_2 (P + \Delta P)'(\lambda_0)x_1 + \gamma_1\gamma_2 \frac{(P + \Delta P)''(\lambda_0)}{2!} x_0 = 0 \\ &\cdots\cdots\cdots \\ &(P + \Delta P)(\lambda_0)x_r + \gamma_r (P + \Delta P)'(\lambda_0)x_{r-1} + \cdots + \left(\prod_{i=1}^r \gamma_i\right) \frac{(P + \Delta P)^r(\lambda_0)}{r!} x_0 = 0. \end{aligned}$$

This is equivalent to

$$\begin{bmatrix} (P+\Delta P)(\lambda_0) & & & & \\ \gamma_1 (P+\Delta P)'(\lambda_0) & (P+\Delta P)(\lambda_0) & & & \\ \frac{\gamma_1\gamma_2}{2!}(P+\Delta P)''(\lambda_0) & \gamma_2 (P+\Delta P)'(\lambda_0) & (P+\Delta P)(\lambda_0) & & \\ \vdots & \vdots & & \ddots & \\ \frac{\gamma_1\cdots\gamma_r}{r!}(P+\Delta P)^r(\lambda_0) & \frac{\gamma_2\cdots\gamma_r}{(r-1)!}(P+\Delta P)^{r-1}(\lambda_0) & \cdots & \cdots & (P+\Delta P)(\lambda_0) \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{r-1} \\ x_r \end{bmatrix} = 0,$$

which can be written in the form

$$\begin{bmatrix} \Delta P(\lambda_0) & \Delta P'(\lambda_0) & \cdots & \frac{1}{r!}\Delta P^r(\lambda_0) \end{bmatrix} X = - \begin{bmatrix} P(\lambda_0) & P'(\lambda_0) & \cdots & \frac{1}{r!}P^r(\lambda_0) \end{bmatrix} X, \tag{6.4}$$

where

$$X = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_r \\ & \gamma_1 x_0 & \gamma_2 x_1 & \cdots & \gamma_r x_{r-1} \\ & & \gamma_1\gamma_2 x_0 & \cdots & \gamma_{r-1}\gamma_r x_{r-1} \\ & & & \ddots & \vdots \\ & & & & \left(\prod_{i=1}^r \gamma_i\right) x_0 \end{bmatrix} \in \mathbb{C}^{r,n}_{\mathcal{T},\gamma}.$$

When $\lambda_0 = 0$, (6.4) takes the form

$$\begin{bmatrix} \Delta A_0 & \Delta A_1 & \cdots & \Delta A_r \end{bmatrix} X = - \begin{bmatrix} A_0 & A_1 & \cdots & A_r \end{bmatrix} X. \tag{6.5}$$

By [23, Lemma 1.3], the minimum 2 or Frobenius norm solution of this equation is given by

$$\begin{bmatrix}\Delta A_0 & \Delta A_1 & \cdots & \Delta A_r\end{bmatrix} = -\begin{bmatrix}A_0 & A_1 & \cdots & A_r\end{bmatrix} XX^\dagger. \tag{6.6}$$

Setting $\Delta A_i = 0$ for $i = r+1, \ldots, k$ if $r < k$, $(P + \Delta P)(\lambda)$ has an elementary divisor $\lambda^j, j \geq r+1$. Therefore, in this case, $\delta_s(P, 0, r+1) = \inf_{X \in \mathbb{C}^{r,n}_{\mathcal{T},\Gamma}} \left\| \begin{bmatrix}A_0 & \cdots & A_r\end{bmatrix} XX^\dagger \right\|_s$ for $s = 2$ or $F$. When $\lambda_0 \neq 0$, Lemma 2.1 implies that

$$\begin{bmatrix}\Delta P(\lambda_0) & \Delta P'(\lambda_0) & \cdots & \frac{1}{r!}\Delta P^r(\lambda_0)\end{bmatrix} = \begin{bmatrix}\Delta A_0 & \cdots & \Delta A_k\end{bmatrix} M(\lambda_0; r).$$

Using this in (6.4), the minimum 2 or Frobenius norm solution of the resulting equation is given by

$$\begin{bmatrix}\Delta A_0 & \cdots & \Delta A_k\end{bmatrix} = -\begin{bmatrix}P(\lambda_0) & P'(\lambda_0) & \cdots & \frac{1}{r!}P^r(\lambda_0)\end{bmatrix} X(M(\lambda_0; r)X)^\dagger,$$

thus proving that if $r \leqslant k$, then for $s = 2$ or $F$,

$$\delta_s(P, \lambda_0, r+1) = \inf_{X \in \mathbb{C}^{r,n}_{\mathcal{T},\Gamma}} \left\| \begin{bmatrix}P(\lambda_0) & P'(\lambda_0) & \cdots & \frac{1}{r!}P^r(\lambda_0)\end{bmatrix} X(M(\lambda_0; r)X)^\dagger \right\|_s.$$

If $r > k$, then $P(\lambda) + \Delta P(\lambda)$ has an elementary divisor $(\lambda - \lambda_0)^j$, where $j \geqslant r+1$ if and only if there exists vectors $x_0, x_1, \ldots, x_r \in \mathbb{R}^n$ with $x_0 \neq 0$ and $r$ positive scalars $\gamma_1, \ldots, \gamma_r$ such that

$$\begin{aligned}
&(P+\Delta P)(\lambda_0)x_0 = 0\\
&(P+\Delta P)(\lambda_0)x_1 + \gamma_1(P+\Delta P)'(\lambda_0)x_0 = 0\\
&(P+\Delta P)(\lambda_0)x_2 + \gamma_2(P+\Delta P)'(\lambda_0)x_1 + \gamma_1\gamma_2\frac{(P+\Delta P)''(\lambda_0)}{2!}x_0 = 0\\
&\cdots\cdots\cdots\\
&(P+\Delta P)(\lambda_0)x_k + \gamma_k(P+\Delta P)'(\lambda_0)x_{k-1} + \cdots + \left(\prod_{i=1}^{k}\gamma_i\right)\frac{(P+\Delta P)^k(\lambda_0)}{k!}x_0 = 0\\
&\cdots\cdots\cdots\\
&(P+\Delta P)(0)x_r + \gamma_r(P+\Delta P)'(\lambda_0)x_{r-1} + \cdots + \left(\prod_{i=r-k+1}^{r}\gamma_i\right)\frac{(P+\Delta P)^k(\lambda_0)}{k!}x_{r-k} = 0.
\end{aligned}$$

This set of equations can be written in the form

$$\begin{bmatrix}
(P+\Delta P)(\lambda_0) & & & & & \\
\gamma_1(P+\Delta P)'(\lambda_0) & (P+\Delta P)(\lambda_0) & & & & \\
\vdots & \vdots & \ddots & & & \\
\frac{\gamma_1\cdots\gamma_k}{k!}(P+\Delta P)^k(\lambda_0) & & & (P+\Delta P)(\lambda_0) & & \\
 & \ddots & & & \ddots & \\
 & & & \frac{\gamma_{r-k+1}\cdots\gamma_r}{k!}(P+\Delta P)^k(\lambda_0) & \cdots & (P+\Delta P)(\lambda_0)
\end{bmatrix}
\begin{bmatrix}x_0\\ x_1\\ \vdots\\ x_k\\ \vdots\\ x_r\end{bmatrix} = 0.$$

This is equivalent to

$$\begin{bmatrix}\Delta P(\lambda_0) & \Delta P'(\lambda_0) & \cdots & \frac{1}{k!}\Delta P^k(\lambda_0)\end{bmatrix} X = -\begin{bmatrix}P(\lambda_0) & P'(\lambda_0) & \cdots & \frac{1}{k!}P^k(\lambda_0)\end{bmatrix} X,$$

where

$$X = \begin{bmatrix} x_0 & x_1 & x_2 & \cdots & x_k & \cdots & x_r \\ & \gamma_1 x_0 & \gamma_2 x_1 & \cdots & \gamma_k x_{k-1} & \cdots & \gamma_r x_{r-1} \\ & & \gamma_1\gamma_2 x_0 & \cdots & \gamma_{k-1}\gamma_k x_{k-2} & \cdots & \gamma_{r-1}\gamma_r x_{r-2} \\ & & & \ddots & \vdots & \vdots & \vdots \\ & & & & \left(\prod_{i=1}^{k}\gamma_i\right) x_0 & \cdots & \left(\prod_{i=r-k+1}^{r}\gamma_i\right) x_{r-k} \end{bmatrix} \in \mathbb{C}^{r,n}_{\mathcal{T},\Gamma}.$$

Therefore, the proof follows by arguing as in the previous case. ∎

*Remark* 6.2. The parameters $\gamma_i$ can all be taken to be 1 in the optimization that computes $\delta_s(P, \lambda_0, r+1)$ for $s = 2$ or $F$. As shall be seen in section 8, this will also decrease the number of variables in the optimization. But these parameters play an important role when computing the upper bound for $\delta_s(P, \lambda_0, r+1)$ from this characterization. However, there is no particular advantage in choosing them to be nonzero real or complex numbers when deriving the upper bound.

## 7. Lower bounds on the distance.

The first lower bound on the distance $\delta_F(P, \lambda_0, r+1)$ is derived from Theorem 4.1.

THEOREM 7.1. *Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ be an $n \times n$ matrix polynomial of degree $k$ and $r < kn$ be a positive integer. For $\Gamma$ as given in* (4.1)*, let $\gamma := \begin{bmatrix} \gamma_1, \dots, \gamma_r \end{bmatrix} \in \Gamma$ and $T_\gamma(P, \lambda_0)$ be defined by* (4.2)*. For $s = 2$ or $F$, the distance $\delta_s(P, \lambda_0, r+1)$ to a nearest matrix polynomial having $(\lambda - \lambda_0)^j$ as an elementary divisor with $j \geqslant r+1$ satisfies*

$$\delta_s(P, \lambda_0, r+1) \geqslant \begin{cases} \sup_{\gamma\in\Gamma} \frac{f(\gamma)}{\sqrt{F_1(\gamma)}} & \textit{if } r \leqslant k, \\ \sup_{\gamma\in\Gamma} \frac{f(\gamma)}{\sqrt{F_2(\gamma)}} & \textit{otherwise}, \end{cases}$$

*where $f(\gamma) := \sigma_{(r+1)n-r}(T_\gamma(P, \lambda_0))$,*

$$F_1(\gamma) := \|M(\lambda_0; r)\|_2^2 \max_{1\leqslant p\leqslant r} \left(1 + \sum_{t=1}^{p}\prod_{i=t}^{p} \gamma^2_{(r+1-i)}\right), \ \textit{and}$$

$$F_2(\gamma) := \|M(\lambda_0; r)\|_2^2 \max\left\{ \max_{1\leqslant p\leqslant k}\left(1 + \sum_{t=1}^{p}\prod_{i=t}^{p}\gamma^2_{(r+1-i)}\right), \max_{k+1\leqslant p\leqslant r}\left(1 + \sum_{t=p-k+1}^{p}\prod_{i=t}^{p}\gamma^2_{(r+1-i)}\right)\right\},$$

*$M(\lambda_0; r)$ being as defined in Lemma* 2.1.

*Proof.* Consider the case that $r \leqslant k$. Let $\Delta P(\lambda) = \sum_{i=0}^{k} \lambda^i \Delta A_i$ be an $n \times n$ matrix polynomial such that $(P + \Delta P)(\lambda)$ has an elementary divisor $(\lambda - \lambda_0)^j$ with $j \geqslant r+1$. By Theorem 4.1, $\sigma_{(r+1)n-r}(T_\gamma(P + \Delta P, \lambda_0)) = 0$, where $T_\gamma(P + \Delta P, \lambda_0)$ is as defined in (4.2). By the perturbation theory for singular values,

$$\begin{aligned} f(\gamma) &= \left|\sigma_{(r+1)n-r}(T_\gamma(P + \Delta P, \lambda_0)) - \sigma_{(r+1)n-r}(T_\gamma(P, \lambda_0))\right| \\ &\leqslant \|T_\gamma(\Delta P, \lambda_0)\|_2 . \end{aligned}$$

Observe that

$$T_\gamma(\Delta P, \lambda_0) = \left(I_{r+1} \otimes \begin{bmatrix} \Delta P(\lambda_0) & \cdots & \frac{1}{r!}\Delta P^r(\lambda_0) \end{bmatrix}\right) \begin{bmatrix} \hat{E}^T_{\gamma,1} & \hat{E}^T_{\gamma,2} & \cdots & \hat{E}^T_{\gamma,r+1} \end{bmatrix}^T,$$

where $\hat{E}\gamma_{,j} = E_{\gamma,j} \otimes I_n$, $E_{\gamma,j}, j = 1, \ldots, r+1$, being the $(r+1) \times (r+1)$ matrices

$$E_{\gamma,1} = \begin{bmatrix} 1 & & & \\ & & & \\ & & & \\ & & & \end{bmatrix}, E_{\gamma,2} = \begin{bmatrix} & 1 & & \\ \gamma_1 & & & \\ & & & \\ & & & \end{bmatrix}, \ldots, E_{\gamma,r} = \begin{bmatrix} & & & 1 & \\ & & \gamma_{r-1} & & \\ & \iddots & & & \\ \prod_{i=1}^{r-1} \gamma_i & & & & \\ & & & & \end{bmatrix},$$

$$E_{\gamma,r+1} = \begin{bmatrix} & & & & 1 \\ & & & \gamma_r & \\ & & \gamma_r\gamma_{r-1} & & \\ & \iddots & & & \\ \prod_{i=1}^{r} \gamma_i & & & & \end{bmatrix}.$$

Therefore, $f(\gamma) \leqslant \left\| \begin{bmatrix} \Delta P(\lambda_0) & \cdots & \frac{1}{r!}\Delta P^r(\lambda_0) \end{bmatrix} \right\|_2 \left\| \begin{bmatrix} \hat{E}_{\gamma,1}^T & \hat{E}_{\gamma,2}^T & \cdots & \hat{E}_{\gamma,r+1}^T \end{bmatrix}^T \right\|_2$. Now,

$$\left\| \begin{bmatrix} \Delta P(\lambda_0) & \cdots & \frac{1}{r!}\Delta P^r(\lambda_0) \end{bmatrix} \right\|_2 \leqslant |||\Delta P|||_2 \|M(\lambda_0; r)\|_2,$$

where the last inequality holds due to Lemma 2.1. This shows for $s = 2$ or $F$,

$$\frac{f(\gamma)}{\sqrt{F_1(\gamma)}} \leqslant \delta_s(P, \lambda_0, r+1),$$

and the proof follows by taking the supremum of the left-hand side as $\gamma$ varies over $\Gamma$. The lower bound for the case $r > k$ can be proved by similar arguments. ∎

A second lower bound is obtained from the formulation of $\delta_F(P, \lambda_0, r+1)$ as the reciprocal of a generalized $\mu$-value as derived in Theorem 5.4. The following lemma, the proof of which is evident from elementary properties of singular values, will be used to derive this bound.

LEMMA 7.2. *Let $M \in \mathbb{C}^{p\times q}$ and $N \in \mathbb{C}^{q\times p}$. Then for $i = 1, \ldots, \min\{p, q\}$,*

$$\inf\{\|M\|_2 : \operatorname{rank}(I_p - MN) \leqslant p - i\} = (\sigma_i(N))^{-1}.$$

THEOREM 7.3. *Let $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ be an $n \times n$ matrix polynomial of degree $k$ and $1 \leqslant r < kn$. For any $\lambda_0 \in \mathbb{C}$ which is not an eigenvalue of $P(\lambda)$, let $T(P, \lambda_0)$ be as defined by* (4.3) *and $E$ and $M(\lambda_0; r)$ be as defined in Lemma* 5.3 *and Lemma* 2.1*, respectively. For a given positive integer $t$, let $\mathbb{S}^t$ be the collection of column vectors of length $t$ with positive entries and*

$$W_{t,n}(a) := \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_t \end{bmatrix} \otimes I_n,$$

*where $a = \begin{bmatrix} a_1, \cdots, a_t \end{bmatrix}^T \in \mathbb{S}^t$. Then for $s = 2$ or $F$,*

$$\delta_s(P, \lambda_0, r+1) \geqslant \sup_{a \in \mathbb{S}^{r+1}} (\sigma_{r+1}(B(\lambda_0, a, P)))^{-1},$$

*where $B(\lambda_0, a, P) := W_{r+1,(k+1)n}(a)(I_{r+1} \otimes M(\lambda_0; r))E\ (T(P, \lambda_0))^{-1}W_{r+1,n}^{-1}(a)$.*

*If* $\lambda_0 = 0$, *then for* $s = 2$ *or* $F$,

$$\delta_s(P, 0, r+1) \geqslant \sup_{a \in \mathbb{S}^{r+1}} \sigma_{r+1}((\widehat{B}(0, a, P)))^{-1}$$

*for* $\widehat{B}(0, a, P) := W_{r+1,(\min\{r,k\}+1)n}(a) E\ (T(P, 0))^{-1} W_{r+1,n}^{-1}(a)$.

*Proof.* If $-\Delta P(\lambda) \in S(P, \lambda_0)$, then from inequality (5.5), the nullity of

$$I - W_{r+1,n}^{-1}(a)\,(I_{r+1} \otimes [\Delta A_0 \cdots \Delta A_k])\,W_{r+1,(k+1)n}(a)(I_{r+1} \otimes M(\lambda_0; r)) E\ (T(P, \lambda_0))^{-1}$$

is at least $r+1$. Therefore, nullity $(I - (I_{r+1} \otimes [\Delta A_0 \cdots \Delta A_k])\, B(\lambda_0, a, P)) \geqslant r+1$. By Lemma 7.2, $\left\| \begin{bmatrix} \Delta A_0 & \cdots & \Delta A_k \end{bmatrix} \right\|_2 \geqslant (\sigma_{r+1}(B(\lambda_0, a, P)))^{-1}$ so that for $s = 2$ or $F$, $\delta_s(P, \lambda_0, r+1) \geqslant \sup_{a \in \mathbb{S}^{r+1}} (\sigma_{r+1}(B(\lambda_0, a, P)))^{-1}$.

Similarly, if $-\Delta P(\lambda) \in S(P, 0)$, then from inequality (5.6), the nullity of

$$I - W_{r+1,n}^{-1}(a) \left(I_{r+1} \otimes \begin{bmatrix} \Delta A_0 & \cdots & \Delta A_p \end{bmatrix}\right) W_{r+1,(p+1)n}(a) E\ (T(P, 0))^{-1}$$

is at least $r+1$, where $p = \min\{r, k\}$. This is equivalent to

$$\text{nullity} \left( I - \left(I_{r+1} \otimes \begin{bmatrix} \Delta A_0 & \cdots & \Delta A_p \end{bmatrix}\right) \widehat{B}(0, a, P) \right) \geqslant r+1.$$

By Lemma 7.2, $\left\| \begin{bmatrix} \Delta A_0 & \cdots & \Delta A_p \end{bmatrix} \right\|_2 \geqslant (\sigma_{r+1}(\widehat{B}(0, a, P)))^{-1}$. Therefore,

$$\delta_s(P, 0, r+1) \geqslant \sup_{a \in \mathbb{S}^{r+1}} (\sigma_{r+1}(\widehat{B}(0, a, P)))^{-1} \text{ for } s = 2 \text{ or } F. \qquad \square$$

**8. Upper bound on the distance.** In this section an upper bound on the distance $\delta_F(P, \lambda_0, r+1)$ that can be used in conjunction with the lower bound obtained in Theorem 7.1 is derived.

THEOREM 8.1. *Let* $P(\lambda) = \sum_{i=0}^{k} \lambda^i A_i$ *be an* $n \times n$ *matrix polynomial of degree* $k$ *and* $r < kn$ *be a positive integer. For* $\Gamma$ *as given in* (4.1), *let* $\gamma := \begin{bmatrix} \gamma_1, \ldots, \gamma_r \end{bmatrix} \in \Gamma$, *and let* $f(\gamma) := \sigma_{(r+1)n-r}(T_\gamma(P, \lambda_0))$, *where* $T_\gamma(P, \lambda_0)$ *is as defined in* (4.2). *Suppose that* $v(\gamma) = \begin{bmatrix} v_0^T & v_1^T & \cdots & v_r^T \end{bmatrix}^T$ *and* $u(\gamma) = \begin{bmatrix} u_0^T & u_1^T & \cdots & u_r^T \end{bmatrix}^T$ *are the corresponding right and left singular vectors with* $v_i, u_i \in \mathbb{C}^n$ *for* $i = 0, 1, \ldots, r$ *dependent on* $\gamma$. *Also, let* $\Gamma_0 \subset \Gamma$ *be the collection of all* $\gamma \in \Gamma$ *with the property that the vector* $v_0$ *formed by the first* $n$ *entries of a right singular vector* $v(\gamma)$ *associated with the singular value* $f(\gamma)$ *of* $T_\gamma(P, \lambda_0)$ *is nonzero. Then for* $s = 2$ *or* $F$,

$$\delta_s(P, \lambda_0, r+1) \leqslant \inf_{\gamma \in \Gamma_0} f(\gamma) \left\| U(\gamma)(M(\lambda_0; r) V(\gamma))^\dagger \right\|_s, \tag{8.1}$$

*where* $M(\lambda_0; r)$ *is as defined in Lemma* 2.1, $U(\gamma) = \begin{bmatrix} u_0 & \cdots & u_r \end{bmatrix}$,

$$V(\gamma) := \begin{cases} \begin{bmatrix} v_0 & v_1 & v_2 & \cdots & v_r \\ & \gamma_1 v_0 & \gamma_2 v_1 & \cdots & \gamma_r v_{r-1} \\ & & \gamma_1\gamma_2 v_0 & \cdots & \gamma_{r-1}\gamma_r v_{r-2} \\ & & & \ddots & \vdots \\ & & & & \left(\prod_{i=1}^{r} \gamma_i\right) v_0 \end{bmatrix}^\dagger & \textit{if } r \leqslant k, \\ \begin{bmatrix} v_0 & v_1 & v_2 & \cdots & v_k & \cdots & v_r \\ & \gamma_1 v_0 & \gamma_2 v_1 & \cdots & \gamma_k v_{k-1} & \cdots & \gamma_r v_{r-1} \\ & & \gamma_1\gamma_2 v_0 & \cdots & \gamma_{k-1}\gamma_k v_{k-2} & \cdots & \gamma_{r-1}\gamma_r v_{r-2} \\ & & & \ddots & \vdots & \vdots & \vdots \\ & & & & \left(\prod_{i=1}^{k} \gamma_i\right) v_0 & \cdots & \left(\prod_{i=r-k+1}^{r} \gamma_i\right) v_{r-k} \end{bmatrix}^\dagger & \textit{otherwise}, \end{cases}$$

*and the infimum is taken to be* $\infty$ *if* $\Gamma_0 = \emptyset$.

*Proof.* From (6.2) and (6.3) it is clear that if $\gamma \in \Gamma_0$, then $V(\gamma)$ satisfies

$$\delta_s(P, \lambda_0, r+1) \leqslant \begin{cases} \left\| \begin{bmatrix} P(\lambda_0) & \cdots & \frac{1}{r!}P^r(\lambda_0) \end{bmatrix} V(\gamma)(M(\lambda_0; r)V(\gamma))^\dagger \right\|_s & \text{if } r \leqslant k, \\ \left\| \begin{bmatrix} P(\lambda_0) & \cdots & \frac{1}{k!}P^k(\lambda_0) \end{bmatrix} V(\gamma)(M(\lambda_0; r)V(\gamma))^\dagger \right\|_s & \text{otherwise} \end{cases}$$

for $s = 2$ or $F$. As $v(\gamma)$ is a right singular vector of $T_\gamma(P, \lambda_0)$ corresponding to $f(\gamma)$, it is clear that $\begin{bmatrix} P(\lambda_0) & \cdots & \frac{1}{p!}P^p(\lambda_0) \end{bmatrix} V(\gamma) = U(\gamma)$, where $p = \min\{r, k\}$. In either case, $\delta_s(P, \lambda_0, r+1) \leqslant f(\gamma) \left\| U(\gamma)(M(\lambda_0; r)V(\gamma))^\dagger \right\|_s$, and the proof follows by taking the infimum of the right-hand side of the above inequality as $\gamma$ varies over $\Gamma_0$. ∎

*Remark* 8.2. A matrix polynomial for which $\Gamma_0 = \emptyset$ has never been encountered in practice. Therefore, it is conjectured that the upper bound in Theorem 8.1 is never $\infty$. In fact numerical experiments show that in many cases this upper bound is very close to the computed value of the distance.

**9. Some special cases.** The quantities $\delta_s(P, 0, 2), s = 2, F$, are a measure of the distance to a matrix polynomial nearest to $P(\lambda) = \sum_{i=0}^k \lambda^i A_i$, having a defective eigenvalue at 0. In this case the problem is equivalent to finding a nearest matrix pencil to $\lambda A_1 + A_0$ in the chosen norm that has 0 as a defective eigenvalue. Note that this distance is of significant practical interest, as when $P(\lambda)$ is replaced by rev $P(\lambda)$, then it is the distance to a nearest matrix polynomial with a defective eigenvalue at $\infty$. This problem was considered in [18] for the matrix pencils, where several results that apply only to this special case were obtained.

First, the upper bound for the distances in Theorem 8.1 is given by

$$\inf_{\gamma \in \Gamma_0} f(\gamma) \left\| \begin{bmatrix} u_0 & u_1 \end{bmatrix} \begin{bmatrix} v_0 & v_1 \\ 0 & \gamma v_0 \end{bmatrix}^\dagger \right\|_s$$

for $s = 2$ or $F$, where $\left[\begin{smallmatrix} v_0 \\ v_1 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} u_0 \\ u_1 \end{smallmatrix}\right]$ are the right and left singular vectors of $\left[\begin{smallmatrix} P(0) & 0 \\ \gamma P'(0) & P(0) \end{smallmatrix}\right]$ corresponding to its $(2n-1)$th singular value $f(\gamma)$. In this case, $\gamma$ can be allowed to vary over all positive real numbers, as the restriction $v_0 \neq 0$ can be removed. To see this, assume that $\gamma > 0$ is such that the corresponding vector $v_0 = 0$. Then clearly $u_0 = 0$ and

$$f(\gamma) \left\| \begin{bmatrix} 0 & u_1 \end{bmatrix} \begin{bmatrix} 0 & v_1 \\ 0 & 0 \end{bmatrix}^\dagger \right\|_s = f(\gamma).$$

Let $\Delta P(\lambda) = \sum_{i=0}^k \lambda^i \Delta A_i$, where $\Delta A_0 = -f(\gamma) u_1 v_1^*$ and $\Delta A_i = 0$ for all $i = 2, \ldots, k$. Then

$$[\Delta A_0 \quad \Delta A_1] = -f(\gamma) \begin{bmatrix} 0 & u_1 \end{bmatrix} \begin{bmatrix} 0 & v_1 \\ 0 & 0 \end{bmatrix}^\dagger \text{ and } |||\Delta P|||_s = f(\gamma) \left\| \begin{bmatrix} 0 & u_1 \end{bmatrix} \begin{bmatrix} 0 & v_1 \\ 0 & 0 \end{bmatrix}^\dagger \right\|_s = f(\gamma)$$

for $s = 2$ or $F$, and the relations

$$\begin{bmatrix} P(0) & \\ \gamma P'(0) & P(0) \end{bmatrix} \begin{bmatrix} 0 \\ v_1 \end{bmatrix} = f(\gamma) \begin{bmatrix} 0 \\ u_1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & u_1^* \end{bmatrix} \begin{bmatrix} P(0) & \\ \gamma P'(0) & P(0) \end{bmatrix} = f(\gamma) \begin{bmatrix} 0 & v_1^* \end{bmatrix}$$

imply that $A_0 v_1 = f(\gamma) u_1, u_1^* A_0 = f(\gamma) v_1^*$, and $u_1^* A_1 = 0$. Therefore,

$$u_1^*(P + \Delta P)(0) = u_1^* A_0 - f(\gamma) v_1^* = 0, \ (P + \Delta P)(0) v_1 = A_0 v_1 - f(\gamma) u_1 = 0$$

and $u_1^*(P+\Delta P)'(0)v_1 = u_1^* A_1 v_1 = 0$. So unless $(P+\Delta P)(\lambda)$ is singular, 0 is a multiple eigenvalue of $(P+\Delta P)(\lambda)$. In either case the objective is achieved, as the polynomial $(P+\Delta P)(\lambda)$ is arbitrarily close to having an elementary divisor $\lambda^j$, $j \geq 2$.

Second, a formula for the Frobenius norm distance to a nearest matrix polynomial with a defective eigenvalue at 0 may be found for the special case that 0 is already an eigenvalue of $P(\lambda)$ (so that $\operatorname{rank} A_0 = n-1$) and the allowable perturbations to $P(\lambda)$ have the property that their coefficient matrices have rank at most 1. The formula is given by the following theorem. The proof is identical to that of [18, Theorem 5.4].

THEOREM 9.1. *Let $P(\lambda) = \sum_{i=0}^k \lambda^i A_i$ be an $n \times n$ matrix polynomial of degree $k$, where* $\operatorname{rank} A_0 = n-1$. *Suppose $A_0 = U\Sigma V^*$ is a singular value decomposition of $A_0$ and $a_{i,j}$ is the entry of $U^* A_1 V$ in the $i$th row and $j$th column. Define $X$ and $Y$ as*

$$X{:=}\begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_{n-1} & \\ a_{n,1} & \cdots & \cdots & a_{n,n-1} & a_{n,n} \end{bmatrix}, Y{:=}\begin{bmatrix} \sigma_1 & & & & a_{1,n} \\ & \sigma_2 & & & a_{2,n} \\ & & \ddots & & \vdots \\ & & & \sigma_{n-1} & a_{n-1,n} \\ & & & & a_{n,n} \end{bmatrix},$$

*where $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_{n-1} > 0$ are the singular values of $A_0$. Then the distance with respect to the norms $|||\cdot|||_F$ to the nearest matrix polynomial with a defective eigenvalue at 0 under the restriction that the coefficient matrices of the perturbating matrix polynomial have rank at most 1 is given by* $\min\{\sigma_{\min}(X), \sigma_{\min}(Y)\}$.

## 10. Numerical experiments.

This section presents numerical experiments conducted to illustrate the upper and lower bounds on the distances and their values computed via BFGS and MATLAB's `globalsearch` algorithm from the formulation in Theorem 6.1.

Computing $\delta_F(P, \lambda_0, r+1)$ from the optimization in Theorem 6.1 via BFGS requires the gradient of the objective function $f(X) := \left\| HX(M(\lambda_0; r)X)^\dagger \right\|_F$, where $X$ varies depending on whether $r \leqslant k$ or $r > k$ and $H = \begin{bmatrix} P(\lambda_0) & \cdots & \frac{1}{p!}P^p(\lambda_0) \end{bmatrix}$, $p = \min\{r, k\}$. By Lemma 2.1, $H = \begin{bmatrix} A_0 & \cdots & A_k \end{bmatrix} M(\lambda_0; r)$. Therefore,

$$f(X) = \left\| G(M(\lambda_0; r)X)(M(\lambda_0; r)X)^\dagger \right\|_F,$$

where $G = \begin{bmatrix} A_0 & \cdots & A_k \end{bmatrix}$. Only real matrix polynomials are considered in the experiments. Since $M(\lambda_0; r)$ has full column rank, for any $X = X_0$ if there exists a neighborhood $S$ of $X_0$ such that $\operatorname{rank} X_0 = \operatorname{rank} X$ for all $X \in S$, then $f(X)$ is differentiable at $X_0$. If we use any numerical scheme to find the infimum of $f(X)$, then generically at every step there exists a neighborhood $S$ of $X$ where every element of $S$ is of full rank, and consequently we can find gradient of $f(X)$ at those points. Additionally, the matrix $X$ involved in the objective function $f(X)$ has a block Toeplitz structure which needs to be incorporated when finding the gradient of $f(X)$. For simplicity, the gradient is initially computed for the function $(f(X))^2$ without taking the structure of $X$ into consideration with the changes due to the structure being incorporated later. Therefore, the function under consideration is

$$g(X) := (f(X))^2 = \left\| G(M(\lambda_0; r)X)(M(\lambda_0; r)X)^\dagger \right\|_F^2.$$

Considering $g(X)$ as a real valued function of the entries of $X$,

$$\nabla g(X)\Big|_{X=X_0} = \mathrm{vec}\left(\frac{dg}{dX}\Big|_{X=X_0}\right).$$

Now, setting $Y = M(\lambda_0; r)X$,

$$dg = 2\langle GYY^\dagger, Gd(YY^\dagger)\rangle, \text{ where } \langle A, B\rangle = \mathrm{trace} A^T B.$$

Expanding the right-hand side gives

$$dg = 2\langle G^T GYY^\dagger (Y^\dagger)^T, dY\rangle + 2\langle Y^T G^T GYY^\dagger, dY^\dagger\rangle, \tag{10.1}$$

where

$$dY^\dagger = (I - Y^\dagger Y)dY^T (Y^\dagger)^T (Y^\dagger) + (Y^\dagger)(Y^\dagger)^T dY^T (I - YY^\dagger) - (Y^\dagger)dY(Y^\dagger),$$

Therefore,

$$\begin{aligned} dg &= 2\langle G^T GYY^\dagger (Y^\dagger)^T, dY\rangle + 2\langle Y^T G^T GYY^\dagger, (I - Y^\dagger Y)dY^T (Y^\dagger)^T (Y^\dagger)\rangle \\ &\quad + 2\langle Y^T G^T GYY^\dagger, (Y^\dagger)(Y^\dagger)^T dY^T (I - YY^\dagger)\rangle - 2\langle Y^T G^T GYY^\dagger, (Y^\dagger)dY(Y^\dagger)\rangle \\ &= 2\langle G^T GYY^\dagger (Y^\dagger)^T, dY\rangle + 2\langle (I - Y^\dagger Y)^T Y^T G^T GYY^\dagger (Y^\dagger)^T (Y^\dagger), dY^T\rangle \\ &\quad + 2\langle (Y^\dagger)(Y^\dagger)^T Y^T G^T GYY^\dagger (I - YY^\dagger)^T, dY^T\rangle \\ &\quad - 2\langle (Y^\dagger)^T Y^T G^T GYY^\dagger (Y^\dagger)^T, dY\rangle \\ &= 2\langle \mathcal{F}(G, Y), dY\rangle, \end{aligned}$$

where

$$\begin{aligned} \mathcal{F}(G, Y) := {} & G^T GYY^\dagger (Y^\dagger)^T + (Y^\dagger)^T Y^\dagger (Y^\dagger)^T Y^T G^T GY(I - Y^\dagger Y) \\ & + (I - YY^\dagger)(Y^\dagger)^T Y^T G^T GYY^\dagger (Y^\dagger)^T - (Y^\dagger)^T Y^T G^T GYY^\dagger (Y^\dagger)^T. \end{aligned}$$

As $dY = M(\lambda_0; r)X$, $\frac{dg}{dX} = 2M(\lambda_0; r)^T \mathcal{F}(G, Y) =: \psi(X)$. Now at a fixed $X_0$, $\frac{dg}{dX}\Big|_{X=X_0} = \psi(X_0)$. Due to the structure of $X$, $\nabla g(X)\Big|_{X=X_0}$ is given by

$$\nabla g(X)\Big|_{X=X_0} = \begin{bmatrix} \sum_{i=1}^{r+1}\left(\begin{bmatrix} \psi(X_0)_{(i-1)n+1,i} \\ \vdots \\ \psi(X_0)_{in,i} \end{bmatrix}\right) \\ \sum_{i=1}^{r}\left(\begin{bmatrix} \psi(X_0)_{(i-1)n+1,i+1} \\ \vdots \\ \psi(X_0)_{in,i+1} \end{bmatrix}\right) \\ \vdots \\ \begin{bmatrix} \psi(X_0)_{1,r+1} \\ \vdots \\ \psi(X_0)_{n,r+1} \end{bmatrix} \end{bmatrix} \text{ if } r \leqslant k$$

and by

$$\nabla g(X)\Big|_{X=X_0} = \begin{bmatrix} \sum_{i=1}^{k+1}\left(\begin{bmatrix}\psi(X_0)_{(i-1)n+1,i}\\ \vdots \\ \psi(X_0)_{in,i}\end{bmatrix}\right) \\ \vdots \\ \sum_{i=1}^{k+1}\left(\begin{bmatrix}\psi(X_0)_{(i-1)n+1,i+r-k}\\ \vdots \\ \psi(X_0)_{in,i+r-k}\end{bmatrix}\right) \\ \sum_{i=1}^{k}\left(\begin{bmatrix}\psi(X_0)_{(i-1)n+1,i+r-k+1}\\ \vdots \\ \psi(X_0)_{in,i+r-k+1}\end{bmatrix}\right) \\ \vdots \\ \begin{bmatrix}\psi(X_0)_{1,r+1}\\ \vdots \\ \psi(X_0)_{n,r+1}\end{bmatrix} \end{bmatrix} \quad \text{if } r > k.$$

Due to the difficulties in computing the gradient of the objective function, the optimization for $\delta_2(P, \lambda_0, r+1)$ in Theorem 6.1 is performed only via the MATLAB's `globalsearch.m`. Also, in each case, the optimizations involved in the lower and upper bounds are computed via the `globalsearch.m` algorithm.

*Example* 10.1. Consider the matrix polynomial

$$P(\lambda) = \begin{bmatrix}-0.1414 & -0.1490\\ 1.1928 & 0.9702\end{bmatrix} + \lambda\begin{bmatrix}0.8837 & 0.9969\\ 0.2190 & 0.0259\end{bmatrix} + \lambda^2\begin{bmatrix}0.6346 & 0.9689\\ 0.6252 & -0.0649\end{bmatrix} + \lambda^3\begin{bmatrix}-1.9867 & 1.2800\\ 0.6097 & -0.1477\end{bmatrix}.$$

Table 1 records the values of the distance $\delta_F(P, 0, r)$ computed via BFGS and `globalsearch.m` algorithms using the formulation in Theorem 6.1 for various values of $r$ together with lower bounds from Theorem 7.1 and Theorem 7.3 and the upper bound from Theorem 8.1. Table 2 records the same for $\delta_F(P, 1, r)$ as $r$ varies from 2 to 6. Likewise, Table 3 and Table 4 record the corresponding quantities for the distances $\delta_2(P, 0, r)$ and $\delta_2(P, 1, r)$, respectively, except that in these cases the distance is computed only via the `globalsearch.m` algorithm.

*Example* 10.2. Consider the matrix polynomial

$$P(\lambda) = \begin{bmatrix}2.7694 & 0.7254 & -0.2050\\ -1.3499 & -0.0631 & -0.1241\\ 3.0349 & .7147 & 1.4897\end{bmatrix} + \lambda\begin{bmatrix}1.4090 & -1.2075 & 0.4889\\ 1.4172 & 0.7172 & 1.0347\\ 0.6715 & 1.6302 & 0.7269\end{bmatrix} + \lambda^2\begin{bmatrix}-0.3034 & 0.8884 & -0.8095\\ 0.2939 & -1.1471 & -2.9443\\ -0.7873 & -1.0689 & 1.4384\end{bmatrix}.$$

TABLE 1
*Comparison of upper and lower bounds with the distance $\delta_F(P, 0, r)$ calculated by BFGS and* `globalsearch.m` *for Example* 10.1.

| Distance measured | Lower bound (Theorem 7.3) | Lower bound (Theorem 7.1) | BFGS | `globalsearch` | Upper bound (Theorem 8.1) |
|---|---|---|---|---|---|
| $\delta_F(P, 0, 2)$ | 0.10797922 | 0.10683102 | 0.14992951 | 0.14992951 | 0.1504944 |
| $\delta_F(P, 0, 3)$ | 0.17943541 | 0.17354340 | 0.27433442 | 0.27433442 | 0.27996519 |
| $\delta_F(P, 0, 4)$ | 0.83444419 | 0.65889251 | 1.41424988 | 1.41424988 | 1.4189444 |
| $\delta_F(P, 0, 5)$ | 0.90827444 | 0.75348431 | 1.46326471 | 1.46326471 | 1.47185479 |
| $\delta_F(P, 0, 6)$ | 0.99263034 | 0.85789363 | 1.66359899 | 1.66359899 | 1.72452708 |

Table 2
*Comparison of upper and lower bounds with the distance $\delta_F(P,1,r)$ calculated by BFGS and* `globalsearch.m` *for Example* 10.1.

| Distance measured | Lower bound (Theorem 7.3) | Lower bound (Theorem 7.1) | BFGS | `globalsearch` | Upper bound (Theorem 8.1) |
|---|---|---|---|---|---|
| $\delta_F(P,1,2)$ | 1.35798224 | 0.70551994 | 1.35814780 | 1.35814780 | 1.39370758 |
| $\delta_F(P,1,3)$ | 1.35690676 | 0.57675049 | 1.42078740 | 1.42078740 | 1.57015806 |
| $\delta_F(P,1,4)$ | 1.35798160 | 0.56881053 | 1.42220397 | 1.42220397 | 1.76028594 |
| $\delta_F(P,1,5)$ | 1.35689708 | 0.56908887 | 1.45865399 | 1.45865399 | 1.82967789 |
| $\delta_F(P,1,6)$ | 1.35690633 | 0.56789237 | 1.46349849 | 1.46349849 | 1.57008146 |

Table 3
*Comparison of upper and lower bounds with the distance $\delta_2(P,0,r)$ calculated by* `globalsearch.m` *for Example* 10.1.

| Distance measured | Lower bound (Theorem 7.3) | Lower bound (Theorem 7.1) | `globalsearch` | Upper bound (Theorem 8.1) |
|---|---|---|---|---|
| $\delta_2(P,0,2)$ | 0.10797922 | 0.10683102 | .10797922 | 0.11368413 |
| $\delta_2(P,0,3)$ | 0.17943541 | 0.17354340 | .19516063 | 0.21687613 |
| $\delta_2(P,0,4)$ | 0.83444419 | 0.65889251 | 1.04436762 | 1.05968598 |
| $\delta_2(P,0,5)$ | 0.90827444 | 0.75348431 | 1.13265970 | 1.20943709 |
| $\delta_2(P,0,6)$ | 0.99263034 | 0.85789363 | 1.55726928 | 1.70199290 |

Table 4
*Comparison of upper and lower bounds with the distance $\delta_2(P,1,r)$ calculated by* `globalsearch.m` *for Example* 10.1.

| Distance measured | Lower bound (Theorem 7.3) | Lower bound (Theorem 7.1) | `globalsearch` | Upper bound (Theorem 8.1) |
|---|---|---|---|---|
| $\delta_2(P,1,2)$ | 1.35798224 | 0.70551994 | 1.35798224 | 1.35827634 |
| $\delta_2(P,1,3)$ | 1.35690676 | 0.57675049 | 1.35805109 | 1.35813196 |
| $\delta_2(P,1,4)$ | 1.35798160 | 0.56881053 | 1.35805159 | 1.56108421 |
| $\delta_2(P,1,5)$ | 1.35689708 | 0.56908887 | 1.35805160 | 1.52575381 |
| $\delta_2(P,1,6)$ | 1.35690633 | 0.56789237 | 1.416503376 | 1.43921050 |

Table 5 and Table 6 record the computed values of the distances $\delta_F(P,0,r)$ and $\delta_F(P,-1,r)$, respectively, obtained via BFGS and `globalsearch.m` algorithms for all possible values of $r$ together with the upper and lower bounds. The corresponding quantities for the distances $\delta_2(P,0,r)$ and $\delta_2(P,-1,r)$ are recorded in Table 7 and Table 8, respectively, except that in these cases the computed value of the distance is obtained only via the `globalsearch.m` algorithm.

In almost every case the lower bound from Theorem 7.3 is better than the lower bound from Theorem 7.1. The perturbations $\Delta P(\lambda)$ constructed to find the upper bound in Theorem 8.1 may also be obtained by using nonzero singular values of $T_\gamma(P,\lambda_0)$ other than $f(\gamma)$ and a corresponding pair of left and right singular vectors.

TABLE 5

*Comparison of upper and lower bounds with the distance $\delta_F(P,0,r)$ calculated by BFGS and* `globalsearch.m` *for Example* 10.2.

| Distance measured | Lower bound (Theorem 7.3) | Lower bound (Theorem 7.1) | BFGS | `globalsearch` | Upper bound (Theorem 8.1) |
|---|---|---|---|---|---|
| $\delta_F(P,0,2)$ | 0.25800277 | 0.25750097 | 0.25904415 | 0.25904415 | 0.268796 |
| $\delta_F(P,0,3)$ | 0.43621850 | 0.38556596 | 0.69617957 | 0.69617957 | 0.82200773 |
| $\delta_F(P,0,4)$ | 0.88752500 | 0.83727454 | 1.84231345 | 1.84231345 | 2.04437686 |
| $\delta_F(P,0,5)$ | 1.19949290 | 1.13421484 | 1.84468801 | 1.84468801 | 2.43953618 |
| $\delta_F(P,0,6)$ | 1.28885600 | 1.07999296 | 2.60665217 | 2.60665222 | 2.76918876 |

TABLE 6

*Comparison of upper and lower bounds with the distance $\delta_F(P,-1,r)$ calculated by BFGS and* `globalsearch.m` *for Example* 10.2.

| Distance measured | Lower bound (Theorem 7.3) | Lower bound (Theorem 7.1) | BFGS | `globalsearch` | Upper bound (Theorem 8.1) |
|---|---|---|---|---|---|
| $\delta_F(P,-1,2)$ | 0.99413714 | 0.49049043 | 1.14436402 | 1.14436402 | 1.14869786 |
| $\delta_F(P,-1,3)$ | 1.23816383 | 0.57712979 | 2.22703947 | 2.22703947 | 2.37565159 |
| $\delta_F(P,-1,4)$ | 1.33820455 | 0.56416354 | 2.33112163 | 2.33112163 | 2.51177974 |
| $\delta_F(P,-1,5)$ | 1.36050277 | 0.59682624 | 2.44152499 | 2.44152500 | 2.89719526 |
| $\delta_F(P,-1,6)$ | 1.46702487 | 0.61024547 | 2.62503371 | 2.64810973 | 2.93776340 |

TABLE 7

*Comparison of upper and lower bounds with the distance $\delta_2(P,0,r)$ calculated by* `globalsearch.m` *for Example* 10.2.

| Distance measured | Lower bound (Theorem 7.3) | Lower bound (Theorem 7.1) | `globalsearch` | Upper bound (Theorem 8.1) |
|---|---|---|---|---|
| $\delta_2(P,0,2)$ | 0.25800277 | 0.25750097 | 0.25802766 | 0.2581792 |
| $\delta_2(P,0,3)$ | 0.43621850 | 0.38556596 | 0.47215137 | 0.58937606 |
| $\delta_2(P,0,4)$ | 0.88752500 | 0.83727454 | 1.11581440 | 1.57310992 |
| $\delta_2(P,0,5)$ | 1.19949290 | 1.13421484 | 1.49604879 | 1.83989133 |
| $\delta_2(P,0,6)$ | 1.28885600 | 1.07999296 | 1.90820166 | 2.39309442 |

However, the resulting upper bound obtained by taking the infimum of $|||\Delta P|||_s$, $s = 2$ or $F$ over all permissible $\gamma$ does not seem to be an improvement over the one already obtained. For instance, in Example 10.1, the matrix $T_\gamma(P,0)$ corresponding to the distance $\delta_2(P,0,4)$ is of size 8, and the upper bound from Theorem 8.1 reported in Table 3 is constructed by using $\sigma_5(T_\gamma(P,0))$ and its corresponding left and right singular vectors. If the same bound is constructed by considering the three smallest singular values $\sigma_6(T_\gamma(P,0))$, $\sigma_7(T_\gamma(P,0))$, and $\sigma_8(T_\gamma(P,0))$ and corresponding left and right singular vectors, then the values are 1.55784600, 1.65319413, and 2.42096365, respectively. Similar observations have been made by considering the other singular value of $T_\gamma(P,0)$.

Table 8
*Comparison of upper and lower bounds with the distance* $\delta_2(P, -1, r)$ *calculated by* `globalsearch.m` *for Example* 10.2.

| Distance measured | Lower bound (Theorem 7.3) | Lower bound (Theorem 7.1) | `globalsearch` | Upper bound (Theorem 8.1) |
|---|---|---|---|---|
| $\delta_2(P, -1, 2)$ | 0.99413714 | 0.49049043 | 0.99413892 | 1.08915666 |
| $\delta_2(P, -1, 3)$ | 1.23816383 | 0.57712979 | 1.44794214 | 1.95311420 |
| $\delta_2(P, -1, 4)$ | 1.33820455 | 0.56416354 | 1.49553573 | 1.92278887 |
| $\delta_2(P, -1, 5)$ | 1.36050277 | 0.59682624 | 1.70157792 | 2.04844570 |
| $\delta_2(P, -1, 6)$ | 1.46702487 | 0.61024547 | 2.19715515 | 2.64000204 |

The functions involved in the optimizations for computing the bounds and distances have several local optimums. For larger values of $k$ and $n$, the number of parameters increases. Consequently, the computational complexity and possibly the number of locally optimal values increase. Therefore, finding good initial parameter values leading to a globally optimal value also becomes more challenging. This affects the computations, and in most cases the bounds become relatively less tight for larger values of $k$ and $n$. This highlights the need for more research on techniques for performing the optimizations and identifying good initialization strategies for these problems.

## 11. Conclusion.

Given a square matrix polynomial $P(\lambda)$, the problem of finding the distance to a nearest matrix polynomial with an elementary divisor of the form $(\lambda - \lambda_0)^j, j \geqslant r$, for a given $\lambda_0 \in \mathbb{C}$ and $r \geqslant 2$ has been considered. The problem has been characterized in terms of different optimization problems. One of them shows that the solution is the reciprocal of a generalized notion of a $\mu$-value. The other optimization is used to compute the distance via numerical software like BFGS and MATLAB's `globalsearch`. Upper and lower bounds have been derived from the characterizations, and numerical experiments performed to compare them with the computed values of the distance show that they are quite tight in many cases. The optimizations involved in the calculations are computationally quite expensive. But this is also the case with other optimizations proposed in the literature for computing similar distances. Also, due to the nature of the optimizations, it cannot be claimed that globally optimal values have been computed. In particular, this applies to the computed values of the bounds from Theorem 7.3 and Theorem 8.1. However, in many cases they are very close to the computed values of the distance. This leaves the question whether they may actually give the exact solution of the distance problem open for future research.

**Acknowledgment.** We thank the referees for their comments that have significantly improved the quality of the paper.

## REFERENCES

[1] J. Z. Anaya and D. Henrion, *An improved Toeplitz algorithm for polynomial matrix null-space computation*, Appl. Math. Comput., 207 (2009), pp. 256–272.

[2] D. L. Boley, *The algebraic structure of pencils and block Toeplitz matrices*, Linear Algebra Appl., 279 (1998), pp. 255–279.

[3] R. P. Braatz, P. M. Young, J. C. Doyle, and M. Morari, *Computational complexity of $\mu$ calculation*, IEEE Trans. Automat. Control, 39 (1994), pp. 1000–1002.

[4] R. Byers and N. K. Nichols, *On the stability radius of a generalized state-space system*, Linear Algebra Appl., 188 (1993), pp. 113–134.

[5] B. Das and S. Bora, *Nearest Matrix Polynomials with a Specified Elementary Divisor*, preprint, arXiv:1911.01299, 2019.

[6] A. Dmytryshyn, S. Johansson, B. Kågström, and P. Van Dooren, *Geometry of matrix polynomial spaces*, Found. Comput. Math., 19 (2019), pp. 1–28.

[7] F. M. Dopico, P. W. Lawrence, J. Pérez, and P. Van Dooren, *Block Kronecker linearizations of matrix polynomials and their backward errors*, Numer. Math., 140 (2018), pp. 373–426.

[8] N. H. Du, V. H. Linh, and V. Mehrmann, *Robust stability of differential-algebraic equations*, in Surveys in Differential-Algebraic Equations I, Springer, New York, 2013, pp. 63–95.

[9] V. Futorny, T. Klymchuk, V. V. Sergeichuk, and N. Shvai, *A Constructive Proof of Pokrzywa's Theorem about Perturbations of Matrix Pencils*, preprint, arXiv:1907.03213, 2019.

[10] F. R. Gantmacher, *The Theory of Matrices*, Vol. 2, Chelsea, New York, 1959.

[11] J. M. Gracia, I. de Hoyos, and I. Zaballa, *Perturbation of linear control systems*, Linear Algebra Appl., 121 (1989), pp. 353–383.

[12] G. Halikias, G. Galanis, N. Karcanias, and E. Milonidis, *Nearest common root of polynomials, approximate greatest common divisor and the structured singular value*, IMA J. Math. Control Inform., 30 (2013), pp. 423–442.

[13] N. Karcanias and G. Kalogeropoulos, *On the Segré, Weyr characteristics of right (left) regular matrix pencils*, Internat. J. Control, 44 (1986), pp. 991–1015.

[14] M. Karow, *Geometry of Spectral Value Sets*, Ph.D. thesis, University of Bremen, Bremen, Germany, 2003.

[15] M. Karow, D. Kressner, and E. Mengi, *Nonlinear eigenvalue problems with specified eigenvalues*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 819–834.

[16] M. Karow and E. Mengi, *Matrix polynomials with specified eigenvalues*, Linear Algebra Appl., 466 (2015), pp. 457–482.

[17] E. Kokabifar, P. Psarrakos, and G. Loghmani, *On the distance from a matrix polynomial to matrix polynomials with some prescribed eigenvalues*, Linear Algebra Appl., 544 (2018), pp. 158–185.

[18] A. Kothyari, B. Das, S. Bora, and M. N. Belur, *On the distance to singular descriptor dynamical systems with impulsive initial conditions*, IEEE Trans. Automat. Control, 64 (2019), pp. 1137–1149.

[19] D. Kressner, E. Mengi, I. Nakić, and N. Truhar, *Generalized eigenvalue problems with specified eigenvalues*, IMA J. Numer. Anal., 34 (2014), pp. 480–501.

[20] A. Packard and J. Doyle, *The complex structured singular value*, Automatica J. IFAC, 29 (1993), pp. 71–109.

[21] N. Papathanasiou and P. Psarrakos, *The distance from a matrix polynomial to matrix polynomials with a prescribed multiple eigenvalue*, Linear Algebra Appl., 429 (2008), pp. 1453–1477.

[22] P. J. Psarrakos, *Distance bounds for prescribed multiple eigenvalues of matrix polynomials*, Linear Algebra Appl., 436 (2012), pp. 4107–4119.

[23] J.-G. Sun, *Backward perturbation analysis of certain characteristic subspaces*, Numer. Math., 65 (1993), pp. 357–382.

[24] P. Van Dooren, P. Dewilde, and J. Vandewalle, *On the determination of the Smith-Macmillan form of a rational matrix from its Laurent expansion*, IEEE Trans. Circuits Syst., 26 (1979), pp. 180–189.

[25] J. Vandewalle and P. Dewilde, *On the determination of the order and the degree of a zero of a rational matrix*, IEEE Trans. Automat. Control, 19 (1974), pp. 608–609.

[26] A. Varga, *On stabilization methods of descriptor systems*, Syst. Control Lett., 24 (1995), pp. 133–138.