Linear Algebra and its Applications 466 (2015) 457–482

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Matrix polynomials with specified eigenvalues

Michael Karow [a], Emre Mengi [b,*]

[a] *Department of Mathematics, TU-Berlin, Sekretariat MA 4-5, Strasse des 17. Juni 136, D-10623, Berlin, Germany*
[b] *Department of Mathematics, Koç University, Rumelifeneri Yolu, 34450 Sarıyer-İstanbul, Turkey*

ARTICLE INFO

Article history:
Received 25 January 2014
Accepted 9 October 2014
Available online 5 November 2014
Submitted by F. Dopico

MSC:
65F15
65F18
47A56

Keywords:
Matrix polynomial
Linearization
Singular values
Sylvester equation
Eigenvalue perturbation theory

ABSTRACT

This work concerns the distance in the 2-norm from a given matrix polynomial to a nearest polynomial with a specified number of its eigenvalues at specified locations in the complex plane. Initially, we consider perturbations of the constant coefficient matrix only. A singular value optimization characterization is derived for the associated distance. We also consider the distance in the general setting, when all of the coefficient matrices are perturbed. In this general setting, we obtain a lower bound in terms of another singular value optimization problem. The singular value optimization problems derived facilitate the numerical computation of the distances.

* Corresponding author.
E-mail addresses: karow@math.tu-berlin.de (M. Karow), emengi@ku.edu.tr (E. Mengi).

http://dx.doi.org/10.1016/j.laa.2014.10.010
0024-3795/

1. Introduction

We study the distance from a given matrix polynomial to a nearest matrix polynomial with a specified number of its eigenvalues at specified locations in the complex plane. Formally, suppose $P : \mathbb{C} \to \mathbb{C}^{n\times n}$ denotes a given polynomial defined by

$$P(\lambda) := \sum_{j=0}^{m} \lambda^j A_j. \tag{1}$$

Here $A_j \in \mathbb{C}^{n\times n}$, $j = 1, \ldots, m$ are fixed, and we assume A_m is full rank. Furthermore, suppose that a set $\mathbb{S} := \{\lambda_1, \ldots, \lambda_s\}$ consisting of complex scalars, and a positive integer r, are given. We provide a singular value optimization characterization for the distance

$$\tau_r(\mathbb{S}) := \inf\left\{ \|\Delta\|_2 \;\middle|\; \Delta \in \mathbb{C}^{n\times n} \text{ s.t. } \sum_{j=1}^{s} m_j(P+\Delta) \geq r \right\}. \tag{2}$$

Above $m_j(P+\Delta)$ denotes the algebraic multiplicity of λ_j as an eigenvalue of $\widetilde{P}(\lambda) := P(\lambda)+\Delta$, i.e., the multiplicity of λ_j as a root of the characteristic polynomial $\det(\widetilde{P}(\lambda))$. We also consider this distance in the general setting, when perturbations of every coefficient A_j, $j = 0, \ldots, m$, of $P(\lambda)$ are admissible. Another singular value optimization characterization yielding a lower bound for the distance in this general setting is derived.

The characterizations derived here are generalizations of the singular value optimization characterization for a linear matrix pencil of the form $L(\lambda) = A_0 + \lambda A_1$ in [16], which was inspired by [19]. Unlike that in [19], the derivation here fully depends on a Sylvester equation characterization for the matrix polynomial P to have sufficiently many eigenvalues belonging to $\mathbb{S}$. This yields a neater derivation. The machinery here and in [16] have similarities, but considerably more work is required here. Some of the additional machinery that we depend on in this matrix polynomial setting are as follows:

(1) The Sylvester equation characterizations are derived starting from the characterizations in [16], but the derivation employs the companion form linearizations of matrix polynomials in Section 2;

(2) The singular value optimization characterizations originate from a subtle relation between a pair of optimal left and right singular vectors. The details are worked out in Section 3.2. Analogous relations are observed in [19,16] for linear pencils, but the procedure for the same observation for an arbitrary matrix polynomial is more involved;

(3) The divided difference formulas in Section 3.5 provide means to express the derived singular value characterization in a comprehensible fashion. They are not needed in the linear pencil setting.

1.1. Literature

A simple special case of the problem considered here is the *backward error* for a polynomial eigenvalue problem. This concerns the distance from a matrix polynomial to a nearest one with a prescribed eigenvalue, and is studied rigorously in [29]. The ϵ-*pseudospectrum* for a matrix polynomial P consists of each complex scalar whose backward error is no more than ϵ. There is abundant work in this direction, including [30,13,17,7].

The extensions to more than one prescribed eigenvalue are not addressed much in the literature except in a few special cases. Some studies [7,1,22] focus on *the distance to the set of matrix polynomials with a multiple eigenvalue*: geometric characterizations in terms of the ϵ-pseudospectrum are given in [7,1], while in [22], lower and upper bounds in terms of singular value optimization problems are derived. In the latter work, perturbations of all of the coefficient matrices are taken into account; however, it is not clear how tight the derived bounds are. When only perturbations of the constant coefficient are allowed, these bounds coalesce, resulting in an exact singular value formula for the distance. *The distance to the set of matrix polynomials with a prescribed multiple eigenvalue with a prescribed algebraic multiplicity* is considered in [24]. Once again, lower and upper bounds are obtained for the distance in terms of singular value optimization problems.

The singular value optimization characterizations derived here facilitate the numerical computation of the distances. Especially, when $|\mathbb{S}|$ is small, the resulting singular value optimization problems can be solved numerically, by means of algorithms exploiting the Lipschitzness [23,28] and smoothness [21] properties of singular values.

1.2. Outline

We start with the derivation of the Sylvester equation characterization for the condition $\sum_{j=1}^{s} m_j(P) \geq r$ in the next section. This Sylvester equation characterization can equivalently be expressed as a rank characterization. In Section 3, we observe that this rank characterization leads to a singular value optimization formula bounding $\tau_r(\mathbb{S})$ from below due to the Eckart–Young theorem. Eventually, we establish the exact equality of this singular value optimization formula with $\tau_r(\mathbb{S})$, under a multiplicity and a linear independence assumption. We achieve the equality by constructing an optimal perturbation Δ_* with 2-norm as small as possible and such that $\sum_{j=1}^{s} m_j(P+\Delta_*) \geq r$. In Section 3.5, the singular value optimization formula (see Theorem 3.1) is expressed in terms of divided difference formulas (see Theorem 3.3). In Section 4, we obtain a lower bound for the distance when perturbations of all coefficients are admissible. Section 5 is devoted to arguments indicating mildness of the multiplicity and linear independence assumptions when two eigenvalues are prescribed. Section 6 confirms the validity of the results by means of several numerical examples and by making connections with the ϵ-pseudospectrum for a matrix polynomial.

2. Rank characterization for polynomials with specified eigenvalues

In this section we seek a rank characterization for the condition $\sum_{j=1}^{s} m_j(P) \geq r$. Our methodology exploits the companion form linearization $\mathcal{L}(\lambda) := \mathcal{A} + \lambda\mathcal{B}$ for $P(\lambda)$, where

$$\mathcal{A} := \begin{bmatrix} 0 & I & & 0 \\ & & \ddots & \\ 0 & 0 & & I \\ A_0 & A_1 & & A_{m-1} \end{bmatrix} \quad \text{and} \quad \mathcal{B} := \begin{bmatrix} -I & & 0 & 0 \\ & \ddots & & \\ 0 & & -I & 0 \\ 0 & & 0 & A_m \end{bmatrix}. \tag{3}$$

In particular, we benefit from the fact that the eigenvalues of $\mathcal{L}(\lambda)$ and $P(\lambda)$ are the same with the same algebraic multiplicities. The main results in this section and in the succeeding sections rely on the following notation. The set $\mathbb{S}^r$ represents the r-tuples with elements from $\mathbb{S}$. Suppose

$$\mu = \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_r \end{bmatrix}^T \in \mathbb{S}^r \quad \text{and} \quad \Gamma = \begin{bmatrix} \gamma_{21} & \gamma_{31} & \dots & \gamma_{r,r-1} \end{bmatrix}^T \in \mathbb{C}^{r(r-1)/2}.$$

We make use of the following upper triangular matrix-valued function

$$C(\mu, \Gamma) := \begin{bmatrix} \mu_1 & \gamma_{21} & \dots & \gamma_{r1} \\ 0 & \mu_2 & \ddots & \vdots \\ & & \ddots & \gamma_{r(r-1)} \\ 0 & & & \mu_r \end{bmatrix}.$$

The set $\mathcal{G}(\mu)$ consists of Γ values such that $C(\mu, \Gamma)$ has all eigenvalues with geometric multiplicity one. This is a dense subset of $\mathbb{C}^{r(r-1)/2}$ [9].

Our starting point is the following result concerning the multiplicities of the eigenvalues of a matrix pencil [16, Theorem 3.3].

Theorem 2.1. *Let $L(\lambda) := A + \lambda B$ be a matrix pencil with $A, B \in \mathbb{C}^{n\times n}$ such that $\operatorname{rank}(B) = n$, let the set $\mathbb{S} := \{\lambda_1, \dots, \lambda_s\}$ consist of complex scalars, and let $r \in \mathbb{Z}^+$. The following two conditions are equivalent:*

(1) *$\sum_{j=1}^{s} m_j(A, B) \geq r$ where $m_j(A, B)$ is the algebraic multiplicity of λ_j as an eigenvalue of $L(\lambda) = A + \lambda B$;*

(2) *There exists a $\mu \in \mathbb{S}^r$ such that for all $\Gamma \in \mathcal{G}(\mu)$*

$$\dim\{X \in \mathbb{C}^{n\times r} \mid AX + BXC(\mu, \Gamma) = 0\} \geq r.$$

Due to the assumption that $\operatorname{rank}(A_m) = n$, the leading coefficient $\mathcal{B}$ in (3) of the linearization $\mathcal{L}(\lambda) = \mathcal{A} + \lambda\mathcal{B}$ is full rank. An application of Theorem 2.1 to the linearization $\mathcal{L}$ leads to the following characterization of the multiple eigenvalues of the matrix

polynomial P. This characterization is expressed in terms of $C^j(\mu,\Gamma) \equiv [C(\mu,\Gamma)]^j$, the jth power of $C(\mu,\Gamma)$.

Theorem 2.2. *Let $P(\lambda) := \sum_{j=0}^m \lambda^j A_j$ with $A_j \in \mathbb{C}^{n\times n}$ be a matrix polynomial such that $\operatorname{rank}(A_m) = n$, let the set $\mathbb{S} := \{\lambda_1,\ldots,\lambda_s\}$ consist of complex scalars, and let $r \in \mathbb{Z}^+$. The following two conditions are equivalent:*

(1) *$\sum_{j=1}^s m_j(P) \geq r$ where $m_j(P)$ is the algebraic multiplicity of λ_j as an eigenvalue of $P(\lambda)$;*
(2) *There exists a $\mu \in \mathbb{S}^r$ such that for all $\Gamma \in \mathcal{G}(\mu)$*

$$\dim\left\{ X \in \mathbb{C}^{n\times r} \;\middle|\; \sum_{j=0}^m A_j X C^j(\mu,\Gamma) = 0 \right\} \geq r.$$

Proof. It follows from Theorem 2.1 that the condition $\sum_{j=1}^s m_j(P) \geq r$, equivalently $\sum_{j=1}^s m_j(\mathcal{A},\mathcal{B}) \geq r$, is met if and only if

$$\dim\left\{ \mathcal{X} \in \mathbb{C}^{mn\times r} \;\middle|\; \mathcal{A}\mathcal{X} + \mathcal{B}\mathcal{X}C(\mu,\Gamma) = 0 \right\} \geq r. \tag{4}$$

We partition $\mathcal{X} = [\, X_0^T \;\; X_1^T \;\; \ldots \;\; X_{m-1}^T \,]^T$ where $X_j \in \mathbb{C}^{n\times r}$. Now the condition

$$0 = \mathcal{A}\mathcal{X} + \mathcal{B}\mathcal{X}C(\mu,\Gamma) = \begin{bmatrix} X_1 \\ \vdots \\ X_{m-1} \\ \sum_{j=0}^{m-1} A_j X_j \end{bmatrix} + \begin{bmatrix} -X_0 C(\mu,\Gamma) \\ \vdots \\ -X_{m-2} C(\mu,\Gamma) \\ A_m X_{m-1} C(\mu,\Gamma) \end{bmatrix}$$

can be expressed as $X_j = X_{j-1}C(\mu,\Gamma)$ for $j = 1,\ldots,m-1$ and

$$\sum_{j=0}^{m-1} A_j X_j + A_m X_{m-1} C(\mu,\Gamma) = 0. \tag{5}$$

By eliminating X_j for $j = 1,\ldots,m-1$ in (5) employing $X_j = X_0 C^j(\mu,\Gamma)$, we obtain

$$\sum_{j=0}^{m} A_j X_0 C^j(\mu,\Gamma) = 0.$$

Thus, X_0 is a solution of $\sum_{j=0}^m A_j X C^j(\mu,\Gamma) = 0$ if and only if

$$\mathcal{X}_0 = \begin{bmatrix} X_0^T & (X_0 C(\mu,\Gamma))^T & \ldots & (X_0 C^{m-1}(\mu,\Gamma))^T \end{bmatrix}^T \tag{6}$$

is a solution of $\mathcal{A}\mathcal{X}+\mathcal{B}\mathcal{X}C(\mu,\Gamma) = 0$. Furthermore, all solutions of $\mathcal{A}\mathcal{X}+\mathcal{B}\mathcal{X}C(\mu,\Gamma) = 0$ must be of the form (6). Now, the result follows from (4). □

Intuitively, Theorem 2.2 also follows from the following observations. Consider a pair $(X, C(\mu, \Gamma))$ satisfying the Sylvester equation

$$\sum_{j=0}^{m} A_j X C^j(\mu, \Gamma) = 0. \tag{7}$$

Such a pair is called an *invariant pair* of the matrix polynomial $P(\lambda)$ in [5]. When $C(\mu, \Gamma)$ is diagonal, it can trivially be verified that $P(\mu_j)x_j = 0$ for $j = 1, \ldots, r$, where x_j denotes the jth column of X. Thus, in this case the columns of X are eigenvectors of P. When $C(\mu, \Gamma)$ is in Jordan form, the pair $(X, C(\mu, \Gamma))$ satisfying (7) is called a *Jordan pair*. In this case, it can be shown that the columns of X are Jordan chains of P [11, Chapter 2]. Thus, a matrix X satisfying the Sylvester equation (7) is inherently related to the generalized eigenspaces of P. The dimension of all such X is related to the dimensions of the generalized eigenspaces. Next, we express the Sylvester characterization in Theorem 2.2 as a rank condition in terms of the Kronecker product $\otimes$.

Corollary 2.3. *Let $P(\lambda) := \sum_{j=0}^{m} \lambda^j A_j$ with $A_j \in \mathbb{C}^{n\times n}$ be a matrix polynomial such that $\operatorname{rank}(A_m) = n$, let the set $\mathbb{S} := \{\lambda_1, \ldots, \lambda_s\}$ consist of complex scalars, and let $r \in \mathbb{Z}^+$. The following two conditions are equivalent:*

(1) *$\sum_{j=1}^{s} m_j(P) \geq r$ where $m_j(P)$ is the algebraic multiplicity of λ_j as an eigenvalue $P(\lambda)$;*

(2) *There exists a $\mu \in \mathbb{S}^r$ such that for all $\Gamma \in \mathcal{G}(\mu)$*

$$\operatorname{rank}\left(\sum_{j=0}^{m} \left(C^j(\mu, \Gamma)\right)^T \otimes A_j\right) \leq n \cdot r - r.$$

Proof. We reserve the notation $\operatorname{vec}(\cdot)$ for the linear operator that stacks up the columns of its matrix argument into a vector. The result follows from Theorem 2.2 and the identity

$$\operatorname{vec}(AXB) = \left(B^T \otimes A\right) \operatorname{vec}(X).$$

Specifically, we apply this identity to $\sum_{j=0}^{m} A_j X C^j(\mu, \Gamma) = 0$. Then, we make use of Theorem 2.2 to conclude that the dimension of the null space of $\sum_{j=0}^{m} (C^j(\mu, \Gamma))^T \otimes A_j$ is at least r for some $\mu \in \mathbb{S}^r$ and for all $\Gamma \in \mathcal{G}(\mu)$ if and only if $\sum_{j=1}^{s} m_j(P) \geq r$ holds. □

As an example, we deduce the following when $\mathbb{S} = \{\mu\}$ and $r = 2$ from the corollary above: the matrix polynomial P has μ as a multiple eigenvalue if and only if

$$\operatorname{rank}\left(\sum_{j=0}^{m} \begin{bmatrix} \mu & 0 \\ \gamma & \mu \end{bmatrix}^j \otimes A_j\right) = \operatorname{rank}\left(\begin{bmatrix} P(\mu) & 0 \\ \gamma P'(\mu) & P(\mu) \end{bmatrix}\right) \leq 2n - 2$$

for all $\gamma \neq 0$.

3. Derivation of the singular value formula

For each $\mu \in \mathbb{S}^r$, let us define

$$\begin{aligned} &\mathcal{P}_r(\mu) := \inf\big\{\|\Delta\|_2 \;\big|\; \Delta \in \mathbb{C}^{n\times n} \text{ s.t. } \operatorname{rank}\big(\mathcal{Q}(\mu, \Gamma, P+\Delta)\big) \le n\cdot r - r \;\forall \Gamma \in \mathcal{G}(\mu)\big\}, \\ &\text{where } \mathcal{Q}(\mu,\Gamma,P) := \sum_{j=0}^{m} \big(C^j(\mu,\Gamma)\big)^T \otimes A_j, \end{aligned} \tag{8}$$

and $P + \Delta =: \widetilde{P}$ denotes the polynomial $\widetilde{P}(\lambda) := \sum_{j=1}^m \lambda^j A_j + (A_0 + \Delta)$. From Corollary 2.3, the distance in (2) is alternatively given by

$$\tau_r(\mathbb{S}) = \inf_{\mu\in\mathbb{S}^r} \mathcal{P}_r(\mu).$$

In the remaining, we derive a singular value formula for $\mathcal{P}_r(\mu)$.

The lower bound

$$\mathcal{P}_r(\mu) \ge \sup_{\Gamma\in\mathbb{C}^{r(r-1)/2}} \sigma_{-r}\big(\mathcal{Q}(\mu,\Gamma,P)\big) =: \kappa_r(\mu), \tag{9}$$

is an immediate consequence of the Eckart–Young theorem. Here and elsewhere, $\sigma_{-r}(\cdot)$ denotes the rth smallest singular value of its matrix argument. When deducing the lower bound in (9), we depend on **(1)** the continuity of $\sigma_{-r}(\mathcal{Q}(\mu,\Gamma,P))$ with respect to Γ and **(2)** the denseness of the set $\mathcal{G}(\mu)$. These two properties ensure that the supremum in (9) can be taken over $\mathbb{C}^{r(r-1)/2}$ rather than over its dense subset $\mathcal{G}(\mu)$. We also establish the validity of the reverse inequality $\mathcal{P}_r(\mu) \le \kappa_r(\mu)$, by constructing $\Delta_* \in \mathbb{C}^{n\times n}$ such that

(P1) $\|\Delta_*\|_2 = \kappa_r(\mu)$ and **(P2)** $\operatorname{rank}\big(\mathcal{Q}(\mu,\Gamma,P+\Delta_*)\big) \le n\cdot r - r$ for some $\Gamma \in \mathcal{G}(\mu)$.

3.1. Statement of Δ_*

As shown in Appendix A, the supremum on the right-hand side of (9) is attained if $r \le n$. Let Γ_* be a point where this supremum is attained. In other words,

$$\kappa_r(\mu) = \sigma_{-r}\big(\mathcal{Q}(\mu,\Gamma_*,P)\big). \tag{10}$$

Let $U, V \in \mathbb{C}^{nr}$ be a consistent pair of unit left and right singular vectors associated with this singular value. In particular, U and V satisfy

$$\mathcal{Q}(\mu,\Gamma_*,P)V = \kappa_r(\mu)U \quad \text{and} \quad U^*\mathcal{Q}(\mu,\Gamma_*,P) = \kappa_r(\mu)V^*. \tag{11}$$

Furthermore, let $\mathcal{U}, \mathcal{V} \in \mathbb{C}^{n\times r}$ represent the matrices such that $\operatorname{vec}(\mathcal{U}) = U$ and $\operatorname{vec}(\mathcal{V}) = V$. In the subsequent two subsections, we prove that

$$\Delta_* := -\kappa_r(\mu)\mathcal{U}\mathcal{V}^+ \tag{12}$$

satisfies both of the properties **(P1)** and **(P2)**, assuming $\sigma_{-r}(\mathcal{Q}(\mu,\Gamma_*,P))$ is simple and $\operatorname{rank}(\mathcal{V}) = r$. We refer to these two assumptions as *multiplicity* and *linear independence* assumptions, respectively.

3.2. Norm of Δ_*

In this subsection, we show that $\|\Delta_*\|_2 = \kappa_r(\mu)$. For this purpose, it is sufficient to establish the validity of $\mathcal{U}^*\mathcal{U} = \mathcal{V}^*\mathcal{V}$, implying

$$\begin{aligned}\left\|\mathcal{U}\mathcal{V}^+\right\|_2 &= \max_{w\in\mathbb{C}^n,\ \|w\|_2=1} \sqrt{w^*\left(\mathcal{V}^+\right)^*\mathcal{U}^*\mathcal{U}\mathcal{V}^+w}\\ &= \max_{w\in\mathbb{C}^n,\ \|w\|_2=1} \sqrt{w^*\left(\mathcal{V}^+\right)^*\mathcal{V}^*\mathcal{V}\mathcal{V}^+w} = \left\|\mathcal{V}\mathcal{V}^+\right\|_2 = 1.\end{aligned}$$

The last equality above is due to the fact that $\mathcal{V}\mathcal{V}^+$ is an orthogonal projector.

Throughout the rest of this subsection we prove $\mathcal{U}^*\mathcal{U} = \mathcal{V}^*\mathcal{V}$ under the multiplicity assumption. As we shall see, this property follows from the fact that Γ_* is a local extremum of $\sigma(\Gamma) := \sigma_{-r}(\mathcal{Q}(\mu,\Gamma,P))$. The partial derivatives of $\mathcal{Q}(\mu,\Gamma,P)$ with respect to the real and the imaginary parts of the components γ_{ik} of Γ are given by

$$\begin{aligned}\frac{\partial\mathcal{Q}}{\partial\Re\gamma_{ik}}(\mu,\Gamma,P) &= \sum_{j=1}^{m}\left(\sum_{\ell=0}^{j-1} C^{\ell}(\mu,\Gamma)\frac{\partial C(\mu,\Gamma)}{\partial\Re\gamma_{ik}}C^{j-1-\ell}(\mu,\Gamma)\right)^T\otimes A_j\\ &= \sum_{j=1}^{m}\sum_{\ell=0}^{j-1}\left(C^{\ell}(\mu,\Gamma)e_ie_k^{\top}C^{j-1-\ell}(\mu,\Gamma)\right)^T\otimes A_j \quad \text{for } 1\le i<k\le r,\\ \frac{\partial\mathcal{Q}}{\partial\Im\gamma_{ik}}(\mu,\Gamma,P) &= \sum_{j=1}^{m}\left(\sum_{\ell=0}^{j-1} C^{\ell}(\mu,\Gamma)\frac{\partial C(\mu,\Gamma)}{\partial\Im\gamma_{ik}}C^{j-1-\ell}(\mu,\Gamma)\right)^T\otimes A_j\\ &= \imath\sum_{j=1}^{m}\sum_{\ell=0}^{j-1}\left(C^{\ell}(\mu,\Gamma)e_ie_k^{\top}C^{j-1-\ell}(\mu,\Gamma)\right)^T\otimes A_j \quad \text{for } 1\le i<k\le r.\end{aligned}$$

Here e_i (e_k) denotes the ith (kth) column of the $r\times r$ identity matrix. Let

$$G := \sum_{j=1}^{m}\sum_{\ell=0}^{j-1} C^{j-1-\ell}(\mu,\Gamma_*)\mathcal{U}^*A_j\mathcal{V}C^{\ell}(\mu,\Gamma_*). \tag{13}$$

From the multiplicity assumption, the singular value $\sigma(\Gamma_*)$ is simple, so it follows that the function $\Gamma\mapsto\sigma(\Gamma)$ is analytic at Γ_* with respect to $\Re\gamma_{ij}$ and $\Im\gamma_{ij}$ for each i and j. Furthermore,

$$
\begin{aligned}
0 &= \frac{\partial \sigma}{\partial \Re \gamma_{ik}}(\Gamma_*) \\
&= \Re\left(U^* \frac{\partial \mathcal{Q}}{\partial \Re \gamma_{ik}}(\mu, \Gamma_*, P) V \right) \\
&= \Re\left(\operatorname{vec}(\mathcal{U})^* \frac{\partial \mathcal{Q}}{\partial \Re \gamma_{ik}}(\mu, \Gamma_*, P) \operatorname{vec}(\mathcal{V}) \right) \\
&= \Re\left(\operatorname{vec}(\mathcal{U})^* \operatorname{vec}\left(\sum_{j=1}^{m} \sum_{\ell=0}^{j-1} A_j \mathcal{V} C^\ell(\mu, \Gamma_*) e_i e_k^\top C^{j-1-\ell}(\mu, \Gamma_*) \right) \right) \\
&= \Re\left(\operatorname{tr}\left(\mathcal{U}^* \sum_{j=1}^{m} \sum_{\ell=0}^{j-1} A_j \mathcal{V} C^\ell(\mu, \Gamma_*) e_i e_k^\top C^{j-1-\ell}(\mu, \Gamma_*) \right) \right) \\
&= \Re\big(e_k^\top G e_i \big) \quad \text{for } 1 \le i < k \le r.
\end{aligned}
$$

The last equation follows from the trace identity $\operatorname{tr}(XY) = \operatorname{tr}(YX)$. Analogously, we have

$$
0 = \frac{\partial \sigma}{\partial \Im \gamma_{ik}}(\Gamma_*) = \Re\big(\imath e_k^T G e_i \big) = -\Im\big(e_k^T G e_i \big) \quad \text{for } 1 \le k < i \le r.
$$

Thus, G is upper triangular. Let

$$
M := -\mathcal{U}^* A_0 \mathcal{V} + \sum_{j=1}^{m} \sum_{\ell=1}^{j-1} C^{j-\ell}(\mu, \Gamma_*) \mathcal{U}^* A_j \mathcal{V} C^\ell(\mu, \Gamma_*). \tag{14}
$$

Then, it is easily verified that

$$
\begin{aligned}
G C(\mu, \Gamma_*) &= M + \mathcal{U}^* \sum_{j=0}^{m} A_j \mathcal{V} C^j(\mu, \Gamma_*) \\
&= M + \sigma(\Gamma_*) \mathcal{U}^* \mathcal{U},
\end{aligned}
$$

where the latter equality follows by writing the left-hand equation in (11) in matrix form. Also,

$$
\begin{aligned}
C(\mu, \Gamma_*) G &= M + \left(\sum_{j=0}^{m} C^j(\mu, \Gamma_*) \mathcal{U}^* A_j \right) \mathcal{V} \\
&= M + \sigma(\Gamma_*) \mathcal{V}^* \mathcal{V},
\end{aligned}
$$

where the second equality follows from the right-hand equation in (11). Thus,

$$
\sigma(\Gamma_*)\big(\mathcal{U}^* \mathcal{U} - \mathcal{V}^* \mathcal{V} \big) = G C(\mu, \Gamma_*) - C(\mu, \Gamma_*) G. \tag{15}
$$

Since G and $C(\mu, \Gamma_*)$ are both upper triangular, the right-hand side of this equation is strictly upper triangular. The left hand-side is Hermitian. Hence, both sides vanish. Thus, $\mathcal{U}^*\mathcal{U} = \mathcal{V}^*\mathcal{V}$.

3.3. Rank of $\mathcal{Q}(\mu, \Gamma_*, P + \Delta_*)$

In this subsection, we establish that $\operatorname{rank}(\mathcal{Q}(\mu, \Gamma_*, P + \Delta_*)) \leq n \cdot r - r$ by proving

$$\dim\left\{ X \in \mathbb{C}^{n\times r} \;\middle|\; \sum_{j=0}^{m} A_j X C^j(\mu, \Gamma_*) + \Delta_* X = 0 \right\} \geq r \tag{16}$$

under the assumption that $\mathcal{V}$ is full rank. Our starting point is the left-hand singular value equation in (11). Writing this equation as a matrix equation, and employing the full rank assumption on $\mathcal{V}$ implying $\mathcal{V}^+\mathcal{V} = I$, we deduce

$$\sum_{j=0}^{m} A_j \mathcal{V} C^j(\mu, \Gamma_*) = \kappa_r(\mu)\mathcal{U}\mathcal{V}^+\mathcal{V} \quad \Longrightarrow \quad \sum_{j=0}^{m} A_j \mathcal{V} C^j(\mu, \Gamma_*) + \Delta_* \mathcal{V} = 0.$$

Moreover, the vector space $\mathcal{D} := \{D \in \mathbb{C}^{r\times r} \mid C(\mu, \Gamma_*)D - DC(\mu, \Gamma_*) = 0\}$ consisting of matrices commuting with $C(\mu, \Gamma_*)$ is of dimension at least r [10, Theorem 1, p. 219]. For all $D \in \mathcal{D}$, we have

$$0 = \sum_{j=0}^{m} A_j \mathcal{V} C^j(\mu, \Gamma_*)D + \Delta_* \mathcal{V} D = \sum_{j=0}^{m} A_j (\mathcal{V}D) C^j(\mu, \Gamma_*) + \Delta_* (\mathcal{V}D).$$

Thus, each matrix in the vector space $\mathcal{V}\mathcal{D} := \{\mathcal{V}D \mid D \in \mathcal{D}\}$ is a solution of the Sylvester equation $\sum_{j=0}^{m} A_j X C^j(\mu, \Gamma_*) + \Delta_* X = 0$. Since the dimension of $\mathcal{V}\mathcal{D}$ is at least r, we conclude with (16).

3.4. Main result

Let us first suppose that μ consists of distinct scalars. Then all eigenvalues of $C(\mu, \Gamma)$ have algebraic and geometric multiplicities equal to one for all Γ. This means $\mathcal{G}(\mu) = \mathbb{C}^{r(r-1)/2}$. In particular $\Gamma_* \in \mathcal{G}(\mu)$. It follows from Sections 3.2 and 3.3 that $\mathcal{P}_r(\mu) = \kappa_r(\mu)$ under multiplicity and linear independence assumptions at Γ_*.

Now suppose that there are repeated scalars in μ. In this case, consider $\tilde{\mu}$, arbitrarily close to μ, that is comprised of distinct scalars. Sections 3.2 and 3.3 guarantee that $\mathcal{P}_r(\tilde{\mu}) = \kappa_r(\tilde{\mu})$ under multiplicity and linear independence assumptions. The equality $\mathcal{P}_r(\mu) = \kappa_r(\mu)$ follows from the continuity of $\mathcal{P}_r(\cdot)$ and $\kappa_r(\cdot)$. Thus, we arrive at the following main result of this paper.

Theorem 3.1 *(Distance to polynomials with specified eigenvalues). Let* $P(\lambda) := \sum_{j=0}^{m} \lambda^j A_j$ *with* $A_j \in \mathbb{C}^{n\times n}$ *be a matrix polynomial such that* $\operatorname{rank}(A_m) = n$*, let the set* $\mathbb{S} :=$

$\{\lambda_1, \ldots, \lambda_s\}$ consist of complex scalars, and let $r \in \mathbb{Z}^+$. Consider the singular value optimization problem

$$\kappa_r(\mathbb{S}) := \inf_{\mu \in \mathbb{S}^r} \sup_{\Gamma \in \mathbb{C}^{r(r-1)/2}} \sigma_{-r}\big(\mathcal{Q}(\mu, \Gamma, P)\big). \tag{17}$$

If $r > n$, suppose that the inner supremum is attained at an optimal μ of (17). Additionally, suppose that the multiplicity and linear independence assumptions hold at optimal $\mu \in \mathbb{S}^r$ and $\Gamma \in \mathbb{C}^{r(r-1)/2}$ of (17). Then

(i) *$\tau_r(\mathbb{S}) = \kappa_r(\mathbb{S})$;*
(ii) *A minimal Δ_* in 2-norm such that $\sum_{j=1}^{s} m_j(P + \Delta_*) \geq r$ is given by (12) for an optimal μ of (17).*

3.5. Simplified formula in terms of divided differences

The singular value optimization characterization (17) seems cumbersome at first. In this section, we show that it can be expressed in a comprehensible way using *divided differences*. Let $x_0, \ldots, x_k \in \mathbb{C}$ be given contiguous nodes, i.e., $x_j = x_\ell$ for $\ell > j$ implies $x_j = x_i$ for all $i \in [j, \ell]$. The divided difference of $f : \mathbb{C} \to \mathbb{C}$ at these nodes is defined recursively by the formula [6], [26, Section 8.2.1], [12, Section B.16]

$$f[x_0, x_1, \ldots, x_k] = \begin{cases} \frac{f[x_1, \ldots, x_k] - f[x_0, \ldots, x_{k-1}]}{x_k - x_0} & x_0 \neq x_k \\ \frac{f^{(k)}(x_0)}{k!} & x_0 = x_k \end{cases} \tag{18}$$

and $f[x_j] = f(x_j)$ for $j = 0, \ldots, k$.

The starting point is the theorem below regarding functions of triangular matrices [8, Corollary of Theorem 2], [18, Theorem 3], [12, Theorem 4.11].

Theorem 3.2 *(Functions of triangular matrices). Let T be an $n \times n$ lower triangular matrix with eigenvalues $\mu_i = t_{ii}$, and $f : \mathbb{C} \to \mathbb{C}$ be a function defined on the spectrum of T. The matrix $\mathcal{T} := f(T)$ is lower triangular with $\mathcal{T}_{ii} = f(\mu_i)$ and*

$$\mathcal{T}_{i\ell} = \sum_{(s_0, s_1, \ldots, s_k)} t_{s_1 s_0} t_{s_2 s_1} \ldots t_{s_k s_{k-1}} f[\mu_{s_0}, \ldots, \mu_{s_k}] \quad \text{for } 1 \leq \ell < i \leq n$$

where the summation is over all increasing sequences of positive integers starting with ℓ and ending with i.

Letting $p_j(x) = x^j$, the formula in (17) concerns the optimization of the rth smallest singular value of

$$\mathcal{Q}(\mu, \Gamma, P) = \sum_{j=0}^{m} p_j\big(C(\mu, \Gamma)^T\big) \otimes A_j.$$

Partition $\mathcal{Q}(\mu,\Gamma,P)$ into $n\times n$ blocks. By an application of Theorem 3.2, its $n\times n$ submatrix at the ith block row and ℓth block column for $1\le \ell < i \le n$ is given by

$$\begin{aligned}\sum_{j=0}^{m}\left(p_j\left(C(\mu,\Gamma)^T\right)\right)_{i\ell}A_j &= \sum_{j=0}^{m}\sum_{(s_0,s_1,\ldots,s_k)} \gamma_{s_1s_0}\gamma_{s_2s_1}\cdots\gamma_{s_ks_{k-1}}p_j[\mu_{s_0},\ldots,\mu_{s_k}]A_j\\ &= \sum_{(s_0,s_1,\ldots,s_k)} \gamma_{s_1s_0}\gamma_{s_2s_1}\cdots\gamma_{s_ks_{k-1}}\left(\sum_{j=0}^{m}p_j[\mu_{s_0},\ldots,\mu_{s_k}]A_j\right)\\ &= \sum_{(s_0,s_1,\ldots,s_k)} \gamma_{s_1s_0}\gamma_{s_2s_1}\cdots\gamma_{s_ks_{k-1}}P[\mu_{s_0},\ldots,\mu_{s_k}].\end{aligned}$$

Above we define $P[\mu_{s_0},\ldots,\mu_{s_k}]$ by the divided difference formula (18) by replacing f with the matrix polynomial P. On the other hand, the $n\times n$ submatrix of $\mathcal{Q}(\mu,\Gamma,P)$ at the ith block row and column is given by

$$\sum_{j=0}^{m}\left(p_j\left(C(\mu,\Gamma)^T\right)\right)_{ii}A_j = \sum_{j=0}^{m}\mu_i^jA_j = P(\mu_i).$$

Theorem 3.3 *(Divided difference characterization). Let $P(\lambda):=\sum_{j=0}^{m}\lambda^jA_j$ with $A_j\in\mathbb{C}^{n\times n}$ be a matrix polynomial such that $\operatorname{rank}(A_m)=n$, let the set $\mathbb{S}:=\{\lambda_1,\ldots,\lambda_s\}$ consist of complex scalars, and let $r\in\mathbb{Z}^+$. The matrix $\mathcal{Q}(\mu,\Gamma,P)\in\mathbb{C}^{nr\times nr}$ in the singular value optimization characterization (17) is block lower triangular. Furthermore, the $n\times n$ submatrix of $\mathcal{Q}(\mu,\Gamma,P)$ at rows $1+(i-1)n:in$ and at columns $1+(\ell-1)n:\ell n$ is given by*

$$\begin{cases}\sum_{(s_0,s_1,\ldots,s_k)}\gamma_{s_1s_0}\gamma_{s_2s_1}\cdots\gamma_{s_ks_{k-1}}P[\mu_{s_0},\ldots,\mu_{s_k}] & i>\ell\\ P(\mu_i) & i=\ell\\ 0 & i<\ell\end{cases},$$

where the summation is over all increasing sequences of positive integers starting with ℓ and ending with i.

Let us focus on the singular value optimization characterization (17) for particular values of r. For $r=2$, i.e., two eigenvalues are prescribed, this characterization takes the form

$$\inf_{\mu_1,\mu_2\in\mathbb{S}}\sup_{\gamma\in\mathbb{C}}\sigma_{-2}\left(\begin{bmatrix}P(\mu_1) & 0\\ \gamma P[\mu_1,\mu_2] & P(\mu_2)\end{bmatrix}\right). \tag{19}$$

For $r=3$, i.e., three eigenvalues are prescribed, the characterization becomes

$$\inf_{\mu_1,\mu_2,\mu_3\in\mathbb{S}}\sup_{\gamma_{21},\gamma_{31},\gamma_{32}\in\mathbb{C}}\sigma_{-3}\left(\begin{bmatrix}P(\mu_1) & 0 & 0\\ \gamma_{21}P[\mu_1,\mu_2] & P(\mu_2) & 0\\ \gamma_{21}\gamma_{32}P[\mu_1,\mu_2,\mu_3]+\gamma_{31}P[\mu_1,\mu_3] & \gamma_{32}P[\mu_2,\mu_3] & P(\mu_3)\end{bmatrix}\right).$$

When $r = 2$, the inner supremum can be performed over $\mathbb{R}$ rather than $\mathbb{C}$. This is due to the observation that the singular values of the matrix in (19) remain the same if γ is replaced by $|\gamma|$. Singular value optimization formulas for $r > 3$ can be obtained in a similar fashion.

3.6. Distances to multiple eigenvalues

A particular case of interest is the distance to a nearest polynomial with an eigenvalue of algebraic multiplicity $\geq r$. This distance was initially considered by Wilkinson [31,32] and Ruhe [27] for matrices due to its connection with the sensitivity of eigenvalues. For matrices the case $r = 2$ is studied in [4,19], the case $r = 3$ is the theme in [14,15], while an arbitrary r is worked out in [20]. For matrix polynomials, a singular value characterization is derived in [22] for $r = 2$. Let us consider the distance

$$\mathcal{M}_r(\mu) := \inf\big\{\|\Delta\|_2 \;\big|\; P(\lambda) + \Delta \text{ has } \mu \text{ as an eigenvalue of algebraic multiplicity} \geq r\big\}. \tag{20}$$

An application of Theorem 3.3 with an arbitrary r and $\mathbb{S} = \{\mu\}$ results in the following characterization for this distance.

Corollary 3.4 *(Distance to polynomials with multiple eigenvalues). Let $P(\lambda) := \sum_{j=0}^{m} \lambda^j A_j$ with $A_j \in \mathbb{C}^{n\times n}$ be a matrix polynomial such that $\operatorname{rank}(A_m) = n$. Furthermore, let $\mu \in \mathbb{C}$ and $r \in \mathbb{Z}^+$. Consider the singular value optimization problem*

$$\xi_r(\mu) = \sup_{\Gamma\in\mathbb{C}^{r(r-1)/2}} \sigma_{-r}\big(\mathcal{Q}(\mu, \Gamma, P)\big), \tag{21}$$

where the $n\times n$ submatrix of $\mathcal{Q}(\mu,\Gamma,P) \in \mathbb{C}^{nr\times nr}$ at rows $1+(i-1)n : in$ and at columns $1+(\ell-1)n : \ell n$ is given by

$$\begin{cases} \sum_{(s_0,s_1,\ldots,s_k)} \gamma_{s_1 s_0}\gamma_{s_2 s_1}\cdots\gamma_{s_k s_{k-1}} \frac{P^{(k)}(\mu)}{k!} & i > \ell \\ P(\mu) & i = \ell \\ 0 & i < \ell \end{cases}$$

and the summation is over all increasing sequences of positive integers starting with ℓ and ending with i. If $r > n$, suppose that the supremum in (21) is attained. Suppose also that the multiplicity and linear independence assumptions hold at an optimal Γ in (21). Then $\mathcal{M}_r(\mu) = \xi_r(\mu)$.

Some particular instances of the singular value optimization formula (21) are as follows. When $r = 2$, that is the distance to polynomials with μ as a multiple eigenvalue is under consideration, the formula (21) simplifies to

$$\sup_{\gamma\in\mathbb{C}} \sigma_{-2}\left(\begin{bmatrix} P(\mu) & 0 \\ \gamma P'(\mu) & P(\mu) \end{bmatrix}\right).$$

This formula is already derived in [22]. When $r = 3$, that is the distance to polynomials with μ as a triple eigenvalue is under consideration, we obtain

$$\sup_{\gamma_{21},\gamma_{31},\gamma_{32}\in\mathbb{C}} \sigma_{-3}\left(\begin{bmatrix} P(\mu) & 0 & 0 \\ \gamma_{21}P'(\mu) & P(\mu) & 0 \\ \gamma_{21}\gamma_{32}P''(\mu)/2 + \gamma_{31}P'(\mu) & \gamma_{32}P'(\mu) & P(\mu) \end{bmatrix}\right).$$

4. Perturbations of all coefficients

In this section we consider the more general distance

$$\tau_r^G(\mathbb{S}) := \inf\left\{ |||\delta P||| \;\middle|\; \sum_{j=1}^{s} m_j(P+\delta P) \geq r,\ \delta P(\lambda) = \sum_{j=0}^{m} \lambda^j \delta A_j \ \exists \delta A_0,\dots,\delta A_m \in \mathbb{C}^{n\times n} \right\}.$$

Above the norm $|||\cdot|||$ for a matrix polynomial $\delta P(\lambda) = \sum_{j=0}^m \lambda^j \delta A_j$ is defined by

$$|||\delta P||| = \sqrt{\sum_{j=0}^{m} c_j \cdot \|\delta A_j\|_2^2}, \quad \text{where } c_j = \begin{cases} 1/\omega_j^2 & \omega_j \neq 0 \\ \infty & \omega_j = 0 \end{cases}$$

for fixed non-negative real scalars $\omega_0,\dots,\omega_m$. The purpose of introducing the scalars ω_j is to take into account the possibility that perturbations of all coefficients may not be equally significant. Specifically, if $\omega_j = 0$, then perturbations of the coefficient A_j are not allowed. Furthermore, if $\omega_0 = 1$ and $\omega_1 = \cdots = \omega_m = 0$, then $\tau_r^G(\mathbb{S}) = \tau_r(\mathbb{S})$.

It follows from Corollary 2.3 that

$$\tau_r^G(\mathbb{S}) = \inf_{\mu\in\mathbb{S}^r} \mathcal{P}_r^G(\mu)$$
$$\text{where } \mathcal{P}_r^G(\mu) := \inf\big\{ |||\delta P||| \;\big|\; \operatorname{rank}\big(\mathcal{Q}(\mu,\Gamma,P+\delta P)\big) \leq n\cdot r - r \ \forall \Gamma\in\mathcal{G}(\mu)\big\}.$$

The Eckart–Young theorem implies that for each δP satisfying $\operatorname{rank}(\mathcal{Q}(\mu,\Gamma,P+\delta P)) \leq n\cdot r - r$, we have

$$\begin{aligned} \sigma_{-r}\big(\mathcal{Q}(\mu,\Gamma,P)\big) &\leq \left\| \sum_{j=0}^{m} \big(C^j(\mu,\Gamma)\big)^T \otimes \delta A_j \right\|_2 \\ &\leq \sum_{j=0}^{m} \big\| \big(C^j(\mu,\Gamma)\big)^T \otimes \delta A_j \big\|_2 \\ &= \sum_{j=0}^{m} \big\| C^j(\mu,\Gamma) \big\|_2 \|\delta A_j\|_2 \end{aligned}$$

$$= \begin{bmatrix} \omega_0\|I\|_2 & \omega_1\|C(\mu,\Gamma)\|_2 & \dots & \omega_m\|C^m(\mu,\Gamma)\|_2 \end{bmatrix} \cdot \begin{bmatrix} \frac{\|\delta A_0\|_2}{\omega_0} \\ \frac{\|\delta A_1\|_2}{\omega_1} \\ \vdots \\ \frac{\|\delta A_n\|_2}{\omega_n} \end{bmatrix}$$

$$\leq \sqrt{\sum_{j=0}^{m} \omega_j^2 \left\|C^j(\mu,\Gamma)\right\|_2^2} \cdot |||\delta P|||.$$

Above, we use the triangle inequality in the second inequality and the Cauchy–Schwarz inequality in the last inequality. Since the inequality holds for each Γ, we deduce

$$\sup_{\Gamma\in\mathbb{C}^{r(r-1)/2}} \frac{\sigma_{-r}(\mathcal{Q}(\mu,\Gamma,P))}{\sqrt{\sum_{j=0}^{m}\omega_j^2\|C^j(\mu,\Gamma)\|_2^2}} \leq \mathcal{P}_r^G(\mu).$$

As for the distance, this leads to the lower bound

$$\inf_{\mu\in\mathbb{S}^r} \sup_{\Gamma\in\mathbb{C}^{r(r-1)/2}} \frac{\sigma_{-r}(\mathcal{Q}(\mu,\Gamma,P))}{\sqrt{\sum_{j=0}^{m}\omega_j^2\|C^j(\mu,\Gamma)\|_2^2}} \leq \tau_r^G(\mathbb{S}),$$

which holds regardless of the multiplicity and linear independence assumptions.

5. Prescribing two eigenvalues

We now turn to the multiplicity and linear independence assumptions when two eigenvalues are prescribed. We intend to relax these assumptions in full generality in future work. The singular value function involved is

$$\sigma_{-2}(\mu_1,\mu_2,\gamma) := \sigma_{-2}\big(\mathcal{Q}(\mu_1,\mu_2,\gamma,P)\big) \quad \text{where } \mathcal{Q}(\mu_1,\mu_2,\gamma,P) = \begin{bmatrix} P(\mu_1) & 0 \\ \gamma P[\mu_1,\mu_2] & P(\mu_2) \end{bmatrix}$$

and $\mu_1,\mu_2\in\mathbb{C}$ are the eigenvalues prescribed. Without loss of generality, we can assume γ is real, as the singular values of $\mathcal{Q}(\mu_1,\mu_2,\gamma,P)$ and $\mathcal{Q}(\mu_1,\mu_2,|\gamma|,P)$ are the same.

The result below shows that the multiplicity assumption is not necessary for the validity of the singular value optimization characterization. To prove it, we utilize the ideas explored in [25, Lemma 4].

Theorem 5.1. *There exists a consistent pair consisting of a left singular vector $U\in\mathbb{C}^{2n}$ and a right singular vector $V\in\mathbb{C}^{2n}$ associated with the singular value $\sigma_{-2}(\mu_1,\mu_2,\gamma_*) := \sup_\gamma \sigma_{-2}(\mu_1,\mu_2,\gamma)$ satisfying*

$$\mathcal{V}^*\mathcal{V} = \mathcal{U}^*\mathcal{U}.$$

Above $\mathcal{V},\mathcal{U}\in\mathbb{C}^{n\times 2}$ are such that $\mathrm{vec}(\mathcal{V}) = V$, $\mathrm{vec}(\mathcal{U}) = U$.

Proof. If the singular value $\sigma_{-2}(\mu_1, \mu_2, \gamma_*)$ is simple, this is already shown in Section 3.2. Otherwise, there exist two analytic singular value functions $\tilde{\sigma}_{-j}(\gamma)$, $j = 1, 2$ of the analytic matrix-valued function $\mathcal{Q}(\mu_1, \mu_2, \gamma, P)$ satisfying $\tilde{\sigma}_{-j}(\gamma_*) = \sigma_{-2}(\mu_1, \mu_2, \gamma_*)$, $j = 1, 2$. Their derivatives at γ_* are given by

$$\frac{\partial \tilde{\sigma}_{-j}(\gamma_*)}{\partial \gamma} = \Re\left[\widetilde{U}_j^* \frac{\partial \mathcal{Q}(\mu_1, \mu_2, \gamma, P)}{\partial \gamma} \widetilde{V}_j\right], \quad j = 1, 2$$

where $\widetilde{U}_j, \widetilde{V}_j \in \mathbb{C}^{2n}$, $j = 1, 2$ form a pair of consistent unit left and right singular vectors associated with $\sigma_{-2}(\mu_1, \mu_2, \gamma_*)$. Furthermore, the sets $\{\widetilde{U}_1, \widetilde{U}_2\}$ and $\{\widetilde{V}_1, \widetilde{V}_2\}$ are orthonormal. Suppose one of the two derivatives above, $\partial \tilde{\sigma}_{-j}(\gamma_*)/\partial \gamma$, is zero. Consider $\mathcal{U}_j, \mathcal{V}_j \in \mathbb{C}^{n \times 2}$ such that $\mathrm{vec}(\mathcal{U}_j) = \widetilde{U}_j$, $\mathrm{vec}(\mathcal{V}_j) = \widetilde{V}_j$. The argument in Section 3.2 applies to deduce that G, defined as in (13) by substituting $\mathcal{U}_j$, $\mathcal{V}_j$ for $\mathcal{U}$, $\mathcal{V}$, is upper triangular. Thus $\mathcal{U}_j^* \mathcal{U}_j = \mathcal{V}_j^* \mathcal{V}_j$.

If, on the other hand, neither of the derivatives $\partial \tilde{\sigma}_{-j}(\gamma_*)/\partial \gamma$, $j = 1, 2$ is zero, then they must have opposite sign. Define $U(\alpha) := \alpha \widetilde{U}_1 + (1 - \alpha)\widetilde{U}_2$ and $V(\alpha) := \alpha \widetilde{V}_1 + (1 - \alpha)\widetilde{V}_2$. Observe that $U(\alpha)$, $V(\alpha)$ for each $\alpha \in [0, 1]$ form a consistent pair of a unit left, unit right singular vectors associated with $\sigma_{-2}(\mu_1, \mu_2, \gamma_*)$. Since

$$\frac{\partial \tilde{\sigma}_{-1}(\gamma_*)}{\partial \gamma} = \Re\left[U(1)^* \frac{\partial \mathcal{Q}(\mu_1, \mu_2, \gamma, P)}{\partial \gamma} V(1)\right] \quad \text{and}$$
$$\frac{\partial \tilde{\sigma}_{-2}(\gamma_*)}{\partial \gamma} = \Re\left[U(0)^* \frac{\partial \mathcal{Q}(\mu_1, \mu_2, \gamma, P)}{\partial \gamma} V(0)\right]$$

have opposite signs, we must have

$$\Re\left[U(\alpha_*)^* \frac{\partial \mathcal{Q}(\mu_1, \mu_2, \gamma, P)}{\partial \gamma} V(\alpha_*)\right] = 0$$

for some $\alpha_* \in (0, 1)$. Again, the argument in Section 3.2 applies to conclude $\mathcal{U}(\alpha_*)^* \mathcal{U}(\alpha_*) = \mathcal{V}(\alpha_*)^* \mathcal{V}(\alpha_*)$ where $\mathcal{U}(\alpha_*), \mathcal{V}(\alpha_*) \in \mathbb{C}^{n \times 2}$ satisfy $\mathrm{vec}(\mathcal{U}(\alpha_*)) = U(\alpha_*)$, $\mathrm{vec}(\mathcal{V}(\alpha_*)) = V(\alpha_*)$. □

The main route to the violation of the linear independence assumption is the attainment of the supremum at $\gamma_* = 0$, as established by Theorem 5.2 below. The argument here is similar to Malyshev's argument [19].

Theorem 5.2. *Suppose that $P[\mu_1, \mu_2]$ is full rank. Let $\mathcal{U}, \mathcal{V} \in \mathbb{C}^{n \times 2}$ be such that* $\mathrm{vec}(\mathcal{U})$, $\mathrm{vec}(\mathcal{V})$ *form a consistent pair of unit left and right singular vectors associated with $\sigma_{-2}(\mu_1, \mu_2, \gamma_*)$ satisfying $\mathcal{U}^*\mathcal{U} = \mathcal{V}^*\mathcal{V}$ as in Theorem 5.1. If* $\mathrm{rank}(\mathcal{V}) = 1$, *then $\gamma_* = 0$.*

Proof. Let us partition $\mathcal{U} = [\, u_1 \;\; u_2 \,]$ and $\mathcal{V} = [\, v_1 \;\; v_2 \,]$ where $u_1, u_2, v_1, v_2 \in \mathbb{C}^n$. Since $\mathrm{vec}(\mathcal{U})$, $\mathrm{vec}(\mathcal{V})$ are a consistent pair of right and left singular vectors, we have

$$P(\mu_1)v_1 = \sigma_* u_1 \quad \text{and} \quad \gamma_* P[\mu_1, \mu_2]v_1 + P(\mu_2)v_2 = \sigma_* u_2 \tag{22}$$

$$u_1^* P(\mu_1) + \gamma_* u_2^* P[\mu_1, \mu_2] = \sigma_* v_1^* \quad \text{and} \quad u_2^* P(\mu_2) = \sigma_* v_2^*, \tag{23}$$

where $\sigma_* := \sigma_{-2}(\mu_1, \mu_2, \gamma_*)$. Now suppose $\operatorname{rank}(\mathcal{V}) = 1$. Assuming $v_1 = 0$ yields $u_1 = 0$ and $\gamma_* u_2^* P[\mu_1, \mu_2] = 0$, thus $\gamma_* = 0$. Similarly, assuming $v_2 = 0$ yields $u_2 = 0$ and $\gamma_* P[\mu_1, \mu_2]v_1 = 0$, thus $\gamma_* = 0$.

Finally, assume $v_1 \neq 0$, $v_2 \neq 0$, but $\operatorname{rank}(\mathcal{V}) = 1$. This yields $v_2 = cv_1$ and $u_2 = cu_1$ for some scalar $c \neq 0$. It follows from (22), by first plugging in cv_1 for v_2 and cu_1 for u_2, then eliminating u_1 using the left-hand equation, that

$$\big[\gamma_* - c(\mu_1 - \mu_2)\big] P[\mu_1, \mu_2]v_1 = 0. \tag{24}$$

In a similar fashion, from (23), by first eliminating v_2 and u_2 using $v_2 = cv_1$ and $u_2 = cu_1$, then eliminating v_1 using the right-hand equation, we have

$$\big[(\mu_1 - \mu_2) + \gamma_* \bar{c}\big] u_1^* P[\mu_1, \mu_2] = 0. \tag{25}$$

Employing the full rank assumption on $P[\mu_1, \mu_2]$ in (24) and (25), we deduce $\gamma_*(1+|c|^2) = (\mu_1 - \mu_2)(1 + |c|^2) = 0$, equivalently $\gamma_* = 0$ and $\mu_1 = \mu_2$. □

To summarize, when two eigenvalues are prescribed, **(i)** the multiplicity assumption can be dropped and **(ii)** the linear independence assumption is satisfied if $P[\mu_1, \mu_2]$ is full-rank and the supremum is attained at $\gamma_* \neq 0$.

6. Numerical examples

We illustrate our main results, Theorem 3.1, Theorem 3.3 and Corollary 3.4, on examples that can be visualized by means of the ϵ-pseudospectrum of P. The ϵ-pseudospectrum of P that is relevant to these results is given by

$$\begin{aligned} \Lambda_\epsilon(P) &:= \bigcup_{\|\Delta\|_2 \leq \epsilon} \Lambda(P + \Delta) \\ &= \big\{ z \in \mathbb{C} \;\big|\; \sigma_{-1}\big(P(z)\big) \leq \epsilon \big\}. \end{aligned}$$

Above, $\Lambda(P)$ denotes the spectrum of the polynomial P.

The numerical experiments in the forthcoming two subsections are performed on a 5×5 quadratic matrix polynomial of the form $P(\lambda) = A_0 + \lambda A + \lambda^2 A_2$. To be precise, the entries of the coefficient matrices rounded to four decimal digits are as follows:

$$A_0 = \begin{bmatrix} 0.5377 & -1.3077 & -1.3499 & -0.2050 & 0.6715 \\ 1.8339 & -0.4336 & 3.0349 & -0.1241 & -1.2075 \\ -2.2588 & 0.3426 & 0.7254 & 1.4897 & 0.7172 \\ 0.8622 & 3.5784 & -0.0631 & 1.4090 & 1.6302 \\ 0.3188 & 2.7694 & 0.7147 & 1.4172 & 0.4889 \end{bmatrix},$$

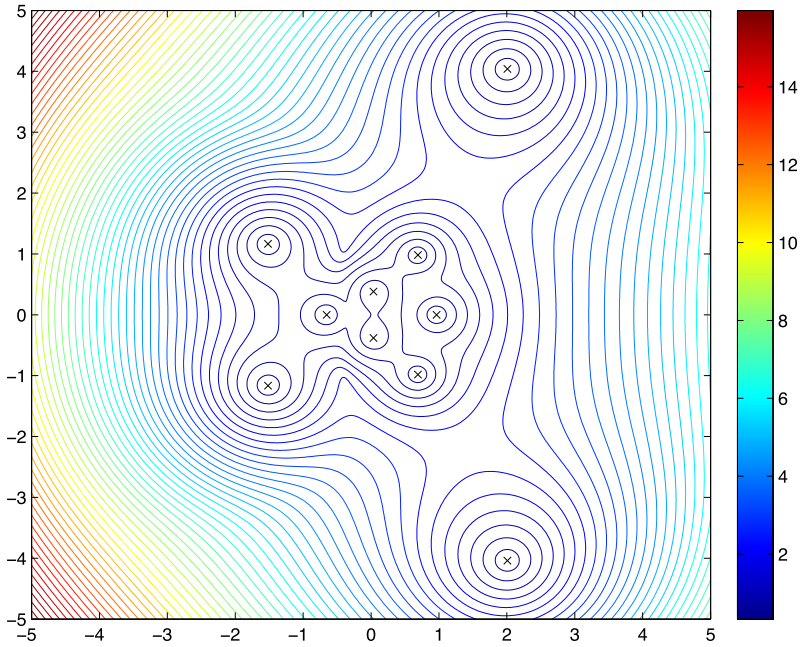

Fig. 1. The ϵ-pseudospectrum of the quadratic random matrix polynomial (26) is displayed for various ϵ. The ten eigenvalues of the polynomial are marked with crosses, and each curve represents the boundary of the ϵ-pseudospectrum for a particular ϵ. The specific value of ϵ can be determined from the color-bar on the right. (For interpretation of the references to color in this figure legend and in other figures, the reader is referred to the web version of this article.)

$$A_1 = \begin{bmatrix} 1.0347 & 0.8884 & 1.4384 & -0.1022 & -0.0301 \\ 0.7269 & -1.1471 & 0.3252 & -0.2414 & -0.1649 \\ -0.3034 & -1.0689 & -0.7549 & 0.3192 & 0.6277 \\ 0.2939 & -0.8095 & 1.3703 & 0.3129 & 1.0933 \\ -0.7873 & -2.9443 & -1.7115 & -0.8649 & 1.1093 \end{bmatrix}, \quad \text{and} \tag{26}$$

$$A_2 = \begin{bmatrix} -0.8637 & 1.5326 & -1.0891 & 0.0859 & -0.6156 \\ 0.0774 & -0.7697 & 0.0326 & -1.4916 & 0.7481 \\ -1.2141 & 0.3714 & 0.5525 & -0.7423 & -0.1924 \\ -1.1135 & -0.2256 & 1.1006 & -1.0616 & 0.8886 \\ -0.0068 & 1.1174 & 1.5442 & 2.3505 & -0.7648 \end{bmatrix}.$$

The entries of each one of $A_0, A_1, A_2 \in \mathbb{R}^{5\times 5}$ above are selected from the normal distribution with zero mean and unit variance, and independently. The ϵ-pseudospectrum of this quadratic matrix polynomial is illustrated in Fig. 1 for various ϵ. Since the coefficient matrices are real, the ϵ-pseudospectrum exhibits symmetry with respect to the real axis. This is easily confirmed by observing that $\overline{P(\bar{z})} = P(z)$. Thus $\sigma_{-1}(P(\bar{z})) = \sigma_{-1}(\overline{P(\bar{z})}) = \sigma_{-1}(P(z))$.

These numerical experiments require the solutions of singular value optimization problems of the form (17). We solve the inner maximization problems using quasi-Newton

methods numerically. Any stationary point of the inner maximization problem in (17) is indeed a global maximizer, as long as the multiplicity and linear independence assumptions hold. We depend on the global technique described in [21] for the solutions of the outer minimization problems. The technique in [21] exploits the smoothness properties of singular value functions.

6.1. Polynomials with two prescribed eigenvalues

Consider the case when $\mathbb{S} = \{\lambda_1, \lambda_2\}$ and $r = 2$. This special case is worked out in detail in Section 5. The singular value optimization formula

$$\tau_2(\mathbb{S}) = \inf_{\mu\in\mathbb{S}^2} \sup_{\gamma} \sigma_{-2}\left(\begin{bmatrix} P(\mu_1) & 0 \\ \gamma P[\mu_1,\mu_2] & P(\mu_2)\end{bmatrix}\right) \quad \text{where } P[\mu_1,\mu_2] = \begin{cases} \frac{P(\mu_1)-P(\mu_2)}{\mu_1-\mu_2} & \text{if } \mu_1 \neq \mu_2 \\ P'(\mu_1) & \text{if } \mu_1 = \mu_2 \end{cases} \tag{27}$$

is shown to hold, if an optimal pair (μ_*, γ_*) with $\mu_* = (\mu_{*1}, \mu_{*2})$ is such that $\gamma_* \neq 0$ and $P[\mu_{*1}, \mu_{*2}]$ is full-rank.

We calculate this distance based on the formula (27) for the prescribed eigenvalues $\mathbb{S} = \{-1-i, -0.7+0.4i\}$. The computed distance is given by $\tau_2(\mathbb{S}) = 0.96328$. In Fig. 2 on the top, the inner-most curves correspond to the boundary of the ϵ-pseudospectrum for $\epsilon = \tau_2(\mathbb{S})$. One of the prescribed eigenvalues $\lambda_1 = -1 - i$ lies on one of these inner-most curves. In general, it is also possible that both of the prescribed eigenvalues lie strictly inside the ϵ-pseudospectrum for $\epsilon = \tau_2(\mathbb{S})$. This is illustrated at the bottom in Fig. 2 for the prescribed eigenvalues $\mathbb{S} = \{-2-i, 2+3i\}$. The computed distance is now given by $\tau_2(\mathbb{S}) = 1.77449$. The optimal γ_* are 0.65088 and 6.13558 for these two examples, respectively. For both examples, $P[\mu_1, \mu_2]$ is full-rank for all $\mu \in \mathbb{S}^2$.

6.2. Nearest polynomials with multiple eigenvalues

By Corollary 3.4, the distance to a nearest matrix polynomial with a multiple eigenvalue is given by

$$\inf_{\mu\in\mathbb{C}} \sup_{\gamma\in\mathbb{R}} \sigma_{-2}\left(\begin{bmatrix} P(\mu) & 0 \\ \gamma P'(\mu) & P(\mu)\end{bmatrix}\right). \tag{28}$$

This formula remains valid even if the multiplicity and linear independence assumptions are violated, as argued in [22, Remark 21]. For a matrix polynomial of size $n \times n$ and degree m with distinct eigenvalues, the ϵ-pseudospectrum for small ϵ is comprised of nm disjoint connected components. There is one connected component evolving around each eigenvalue as ϵ is increased. The smallest ϵ such that two components of the

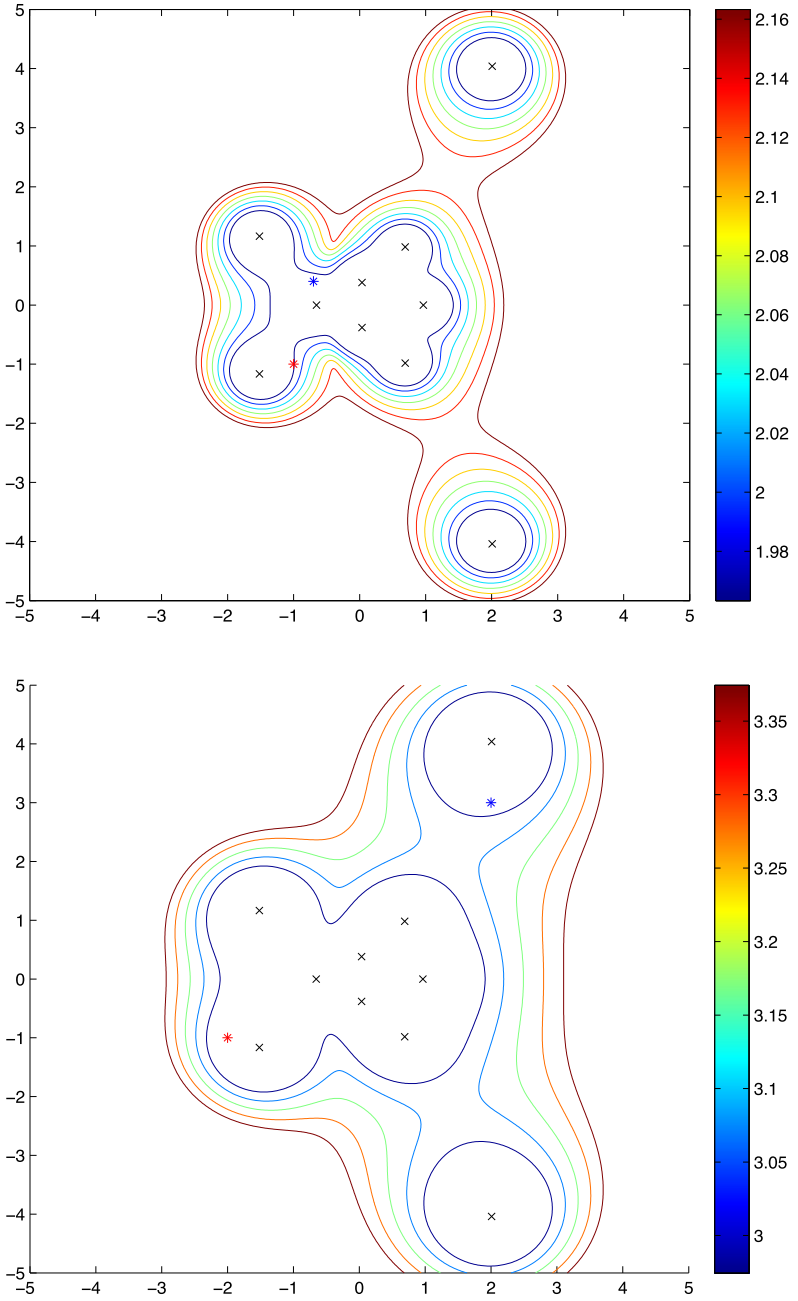

Fig. 2. The distances to two prescribed eigenvalues are depicted on the pseudospectra. In both figures the inner-most curves correspond to the boundary of $\Lambda_\epsilon(P)$ for $\epsilon = \tau_2(\mathbb{S})$. The asterisks represent the prescribed eigenvalues $\mathbb{S} = \{-1 - i, -0.7 + 0.4i\}$ on the top and $\mathbb{S} = \{-2 - i, 2 + 3i\}$ at the bottom.

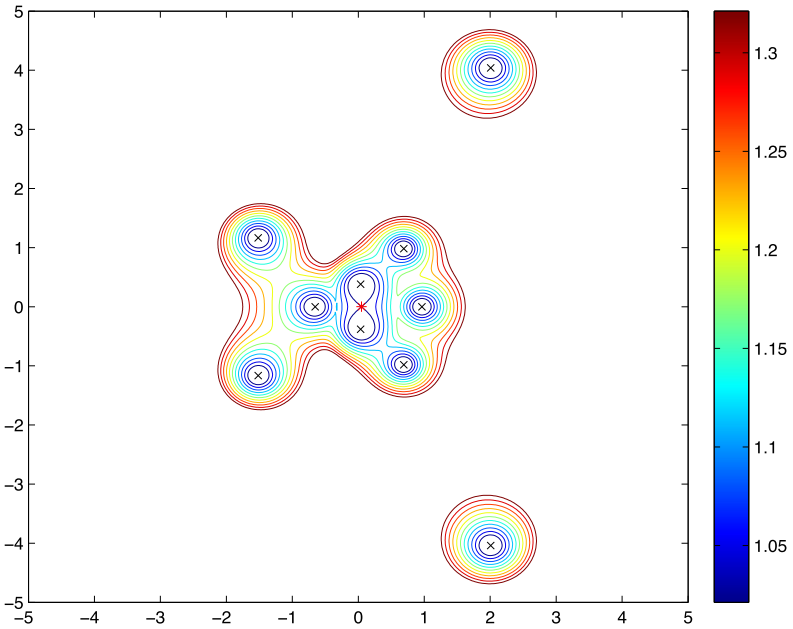

Fig. 3. $\Lambda_\epsilon(P)$ of the quadratic matrix polynomial (26) is illustrated for various ϵ. The inner-most curves represent the boundary of $\Lambda_\epsilon(P)$ for ϵ equal to the distance to a nearest matrix polynomial with a multiple eigenvalue. The asterisk marks the multiple eigenvalue of such a nearest matrix polynomial.

ϵ-pseudospectrum coalesce is equal to (28). This is not an obvious fact: **(i)** for matrices this is established by Alam and Bora [3]; **(ii)** extensions to matrix pencils and matrix polynomials are proven in [2, Theorem 5.1] and [1, Theorem 7.1], respectively.[1]

For the quadratic matrix polynomial (26), we compute this distance as 0.3211 using the formula (28). Two components of the ϵ-pseudospectrum for $\epsilon = 0.3211$ coalesce as expected in theory. This is illustrated in Fig. 3, where the inner-most curves represent the boundary of this ϵ-pseudospectrum. The computed optimal μ value for the outer minimization in (28) is given by $\mu_* = 0.04882$ and corresponds to the multiple eigenvalue of a nearest polynomial. This point, marked by an asterisk in the figure, unsurprisingly turns out to be a point of coalescence of two components of the ϵ-pseudospectrum.

Next, we consider the distance from the quadratic matrix polynomial (26) to a nearest one with a multiple eigenvalue of algebraic multiplicity ≥ 3. The singular value optimization characterization derived in Section 3 for this distance is

$$\inf_{\mu\in\mathbb{C}} \sup_{\gamma_{21},\gamma_{31},\gamma_{32}\in\mathbb{C}} \sigma_{-3}\left(\begin{bmatrix} P(\mu) & 0 & 0 \\ \gamma_{21}P'(\mu) & P(\mu) & 0 \\ \gamma_{21}\gamma_{32}P''(\mu)/2 + \gamma_{31}P'(\mu) & \gamma_{32}P'(\mu) & P(\mu) \end{bmatrix}\right). \tag{29}$$

[1] These extensions to pencils and polynomials are in the general setting when perturbations of all coefficients are admissible.

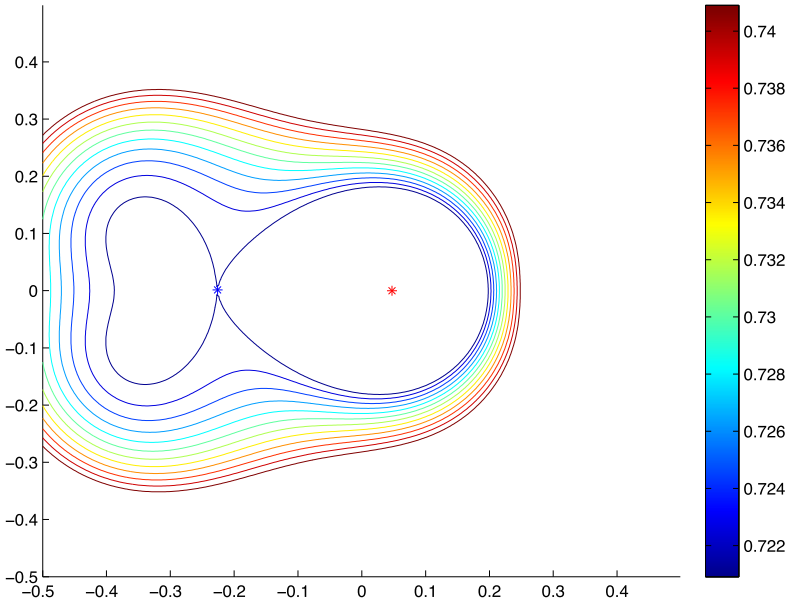

Fig. 4. $\Lambda_{\epsilon,2}(P)$ for the quadratic matrix polynomial (26) is displayed for various ϵ. The inner-most curve represents the boundary of $\Lambda_{\epsilon,2}(P)$ for ϵ equal to the distance to a nearest matrix polynomial with an eigenvalue of algebraic multiplicity at least three. The blue asterisk (located at $z = -0.2260$) marks the triple eigenvalue of such a nearest matrix polynomial. The red asterisk (positioned at $z = 0.04882$) marks the double eigenvalue of a nearest matrix polynomial with a multiple eigenvalue.

In [20], for matrices, it was conjectured that this distance can be posed as the smallest ϵ such that two components of the *ϵ-pseudospectrum of order two* coalesce. The ϵ-pseudospectrum of order two refers to the set consisting of multiple eigenvalues of all matrices within an ϵ neighborhood of the original matrix with respect to the matrix 2-norm. This conjecture is still open. The numerical evidence is in favor of it. The generalization of the ϵ-pseudospectrum of order two for a matrix polynomial P can be defined formally by

$$\begin{aligned}\Lambda_{\epsilon,2}(P) &:= \bigcup_{\|\Delta\|_2 \le \epsilon} \Lambda_2(P+\Delta) \\ &= \left\{ z \in \mathbb{C} \;\middle|\; \sup_{\gamma \in \mathbb{R}} \sigma_{-2}\left(\begin{bmatrix} P(z) & 0 \\ \gamma P'(z) & P(z) \end{bmatrix} \right) \le \epsilon \right\}.\end{aligned}$$

Above $\Lambda_2(P)$ denotes the set of eigenvalues of P of algebraic multiplicity ≥ 2. The computed distance from the polynomial (26) to a nearest one with an eigenvalue of algebraic multiplicity ≥ 3 is 0.6309. In Fig. 4, a plot of $\Lambda_{\epsilon,2}(P)$ is given. Two components of $\Lambda_{\epsilon,2}(P)$ coalesce for $\epsilon = 0.6309$. Furthermore, the optimal μ for the outer minimization in (29), given by $\mu_* = -0.2260$ and marked with a blue asterisk in the figure, appears to be a point of coalescence of two components of $\Lambda_{\epsilon,2}(P)$. The extension of the conjecture in [20] to the matrix polynomial setting seems to hold based on this example.

7. Concluding remarks

We derived a singular value optimization characterization for the distance from a matrix polynomial to a nearest one with a specified number of eigenvalues belonging to a specified set. We restricted ourselves to square matrix polynomials. Extensions to rectangular matrix polynomials are straightforward as long as the leading coefficient matrix is full rank.

There are two important open problems that are not solved fully by this paper. First, it may be more desirable to admit perturbations of all coefficient matrices for some applications. In this case, a lower bound is deduced here, but an exact singular value optimization characterization is not known. Secondly, the results are proven under multiplicity and linear independence assumptions. Our experience with special instances indicates that the singular value optimization formulas remain valid, even when these assumptions are not met. These assumptions in the special case when two eigenvalues are prescribed are discussed here. We plan to address the multiplicity and linear independence assumptions in full generality in future work.

Acknowledgements

We are grateful to Daniel Kressner for reading an initial version and providing helpful suggestions. His suggestions directed us to the simplified divided difference formulas in Section 3.5. We also thank Froilan Dopico, Michael Overton, Francoise Tisseur and an anonymous referee for their valuable feedback. The work of Emre Mengi was supported in part by the European Commission grant PIRG-GA-268355, the TÜBİTAK (the scientific and technological research council of Turkey) carrier grant 109T660, and by the BAGEP program of Turkish Academy of Science.

Appendix A. Proof of attainment of the supremum of the singular value function

Below we establish that the supremum

$$\sup_{\Gamma} \sigma_{-r}\big(\mathcal{Q}(\mu, \Gamma, P)\big) \tag{A.1}$$

is attained if $r \leq n$. The block lower triangular matrix $\mathcal{Q}(\mu, \Gamma, P) \in \mathbb{C}^{nr\times nr}$ above is as in Theorem 3.3. The attainment result here is a generalization of the result presented in the appendix in [16].

Theorem A.1. *Suppose that $P[\mu_k, \mu_l]$ has full rank for each k and l such that $k < l$. We have*

$$\sigma_{-j}\big(\mathcal{Q}(\mu, \Gamma, P)\big) \to 0$$

as at least one of the entries of Γ tends to ∞ in modulus for each $j = 1, \ldots, n$.

Proof. Choose an unbounded entry γ_{lk} of Γ so that $l-k$ is as small as possible. Thus $|\gamma_{ij}|$ is bounded for each i, j such that $i-j<l-k$.

Let us first suppose that none of $\mu_1,\ldots,\mu_r$ is an eigenvalue of $P(\lambda)$. Our approach is based on establishing that the largest n singular values of $\mathcal{Q}(\mu,\Gamma,P)^{-1}$ diverge to ∞ as $|\gamma_{lk}|\to\infty$. Clearly, this is equivalent to the decay of the smallest n singular values of $\mathcal{Q}(\mu,\Gamma,P)$ to zero. In this respect, we claim that $\mathcal{Q}(\mu,\Gamma,P)^{-1}$ is of the form

$$\begin{bmatrix} P(\mu_1)^{-1} & 0 & 0 & \cdots & & 0 \\ X_{21} & P(\mu_2)^{-1} & 0 & & & 0 \\ X_{31} & X_{32} & P(\mu_3)^{-1} & & & \\ & & & \ddots & & \\ & & & & P(\mu_{r-1})^{-1} & 0 \\ X_{r1} & X_{r2} & & & X_{r(r-1)} & P(\mu_r)^{-1} \end{bmatrix}.$$

Above

$$X_{lk}=-\gamma_{lk}P(\mu_l)^{-1}P[\mu_k,\mu_l]P(\mu_k)^{-1}+P_{lk} \tag{A.2}$$

and P_{lk} is a polynomial in γ_{ij} for $j=k,\ldots,l-1$, $i=j+1,\ldots,l$ and $(i,j)\neq(l,k)$. Notice that all these γ_{ij} that P_{lk} depends on are bounded. The validity of (A.2) can be verified by induction on $l-k$. As the base case, we obtain an expression for $X_{(j+1)j}$ for each j by left-multiplying the $(j+1)$th block row of $\mathcal{Q}(\mu,\Gamma,P)^{-1}$ with the jth block column of $\mathcal{Q}(\mu,\Gamma,P)$. This results in

$$\begin{aligned} &X_{(j+1)j}P(\mu_j)+P(\mu_{j+1})^{-1}\big(\gamma_{(j+1)j}P[\mu_j,\mu_{j+1}]\big)=0\\ &\quad\Longrightarrow\quad X_{(j+1)j}=-\gamma_{(j+1)j}P(\mu_{j+1})^{-1}P[\mu_j,\mu_{j+1}]P(\mu_j)^{-1} \end{aligned}$$

as desired. For the inductive case, let us denote the submatrix at the ith block row and jth block column of $\mathcal{Q}(\mu,\Gamma,P)$ by $\mathcal{Q}_{ij}$. Furthermore, let $X_{\ell\ell}:=P(\mu_\ell)^{-1}$. We obtain an expression for X_{ij} with $i>j$ by left-multiplying the ith block row of $\mathcal{Q}(\mu,\Gamma,P)$ with the jth block column of $\mathcal{Q}(\mu,\Gamma,P)^{-1}$. This yields $\sum_{\ell=j}^{i}\mathcal{Q}_{i\ell}X_{\ell j}=0$, that is

$$\begin{aligned} X_{ij}=&-\gamma_{ij}P(\mu_i)^{-1}P[\mu_j,\mu_i]P(\mu_j)^{-1}\\ &-\sum_{(s_0,s_1,\ldots,s_\ell)}\gamma_{s_1s_0}\cdots\gamma_{s_\ell s_{\ell-1}}P(\mu_i)^{-1}P[\mu_{s_0},\ldots,\mu_{s_\ell}]P(\mu_j)^{-1}\\ &-\sum_{\ell=j+1}^{i-1}P(\mu_i)^{-1}\mathcal{Q}_{i\ell}X_{\ell j}. \end{aligned}$$

Above the first summation is over all increasing sequences of integers of length ≥ 3 starting with j and ending with i. In the second summation, by the inductive hypothesis, for each $\ell=j+1,\ldots,i-1$ the matrix $X_{\ell j}$ is a polynomial in γ_{qp} for $p=j,\ldots,\ell-1$,

$q = p+1, \ldots, \ell$ and $(q,p) \neq (\ell, j)$ (i.e., $p \in [j, i-2]$ and $q \in [p+1, i-1]$ meaning $q - p < i - j$). This establishes that

$$X_{ij} = -\gamma_{ij} P(\mu_i)^{-1} P[\mu_j, \mu_i] P(\mu_j)^{-1} + P_{ij}$$

for some P_{ij}, which is a polynomial in γ_{qp} for $p = j, \ldots, i-1$, $q = p+1, \ldots, i$ and $(q,p) \neq (i,j)$. Thus, the validity of (A.2) is confirmed.

It follows from (A.2) that $\sigma_j(X_{lk}) \to \infty$ for $j = 1, \ldots, n$, where we exploit the full rank assumption on $P[\mu_k, \mu_l]$. Finally, we employ the inequality

$$\sigma_j(X_{lk}) \leq \sigma_j\big(\mathcal{Q}(\mu, \Gamma, P)^{-1}\big)$$

to obtain $\sigma_j(\mathcal{Q}(\mu, \Gamma, P)^{-1}) \to \infty$ for each $j = 1, \ldots, n$, as desired.

Now suppose μ_j for some $j \in [1, r]$ is an eigenvalue of $P(\lambda)$. For each $\beta > 0$, there exists $\Delta \in \mathbb{C}^{n \times n}$ such that **(i)** $\|\Delta\|_2 \leq \beta$ and **(ii)** $\widetilde{P}(\lambda) := P(\lambda) + \Delta$ does not contain in its spectrum any μ_k for $k = 1, \ldots, r$. The previous argument applies to $\widetilde{P}$. In particular, the smallest n singular values of $\mathcal{Q}(\mu, \Gamma, \widetilde{P})$ decay to zero as $|\gamma_{lk}| \to \infty$. Thus, for some δ_β, for all γ_{lk} such that $|\gamma_{lk}| > \delta_\beta$, we have

$$\sigma_{-j}\big(\mathcal{Q}(\mu, \Gamma, \widetilde{P})\big) < \beta \quad \Longrightarrow \quad \sigma_{-j}\big(\mathcal{Q}(\mu, \Gamma, P)\big) < 2\beta.$$

This means that $\sigma_{-j}(\mathcal{Q}(\mu, \Gamma, P)) \to 0$ as $|\gamma_{lk}| \to \infty$. □

Observe that the hypothesis that $P[\mu_k, \mu_l]$ is full rank for each $k < l$ is satisfied on a dense subset of the set of complex pairs (μ_k, μ_l). Consequently, the theorem above and the continuity of the singular values guarantee that the supremum (A.1) is attained for each $\mu \in \mathbb{C}^r$ if $r \leq n$.

References

[1] S.S. Ahmad, R. Alam, Pseudospectra, critical points and multiple eigenvalues of matrix polynomials, Linear Algebra Appl. 430 (4) (2009) 1171–1195.
[2] S.S. Ahmad, R. Alam, R. Byers, On pseudospectra, critical points, and multiple eigenvalues of matrix pencils, SIAM J. Matrix Anal. Appl. 31 (4) (2010) 1915–1933.
[3] R. Alam, S. Bora, On sensitivity of eigenvalues and eigendecompositions of matrices, Linear Algebra Appl. 396 (2005) 273–301.
[4] R. Alam, S. Bora, R. Byers, M.L. Overton, Characterization and construction of the nearest defective matrix via coalescence of pseudospectral components, Linear Algebra Appl. 435 (2011) 494–513.
[5] T. Betcke, D. Kressner, Perturbation, extraction and refinement of invariant pairs for matrix polynomials, Linear Algebra Appl. 435 (2011) 514–536.
[6] C. De Boor, Divided differences, Surv. Approx. Theory 1 (2005) 46–69.
[7] L. Boulton, P. Lancaster, P.J. Psarrakos, On pseudospectra of matrix polynomials and their boundaries, Math. Comp. 77 (261) (2008) 313–334.
[8] C. Davis, Explicit functional calculus, Linear Algebra Appl. 6 (1973) 193–199.
[9] J.W. Demmel, A. Edelman, The dimension of matrices (matrix pencils) with given Jordan (Kronecker) canonical forms, Linear Algebra Appl. 230 (1995) 61–87.
[10] F.R. Gantmacher, The Theory of Matrices, vol. 1, Chelsea, 1959.
[11] I. Gohberg, P. Lancaster, L. Rodman, Matrix Polynomials, Academic Press, New York, 1982.

[12] N.J. Higham, Functions of Matrices – Theory and Computation, SIAM, 2008.
[13] N.J. Higham, F. Tisseur, More on pseudospectra for polynomial eigenvalue problems and applications in control theory, Linear Algebra Appl. 351–352 (2002) 435–453.
[14] Kh.D. Ikramov, A.M. Nazari, On the distance to the closest matrix with a triple zero eigenvalue, Mat. Zametki 73 (4) (2003) 545–555.
[15] Kh.D. Ikramov, A.M. Nazari, Justification of a Malyshev-type formula in the nonnormal case, Mat. Zametki 78 (2) (2005) 241–250.
[16] D. Kressner, E. Mengi, I. Nakic, N. Truhar, Generalized eigenvalue problems with specified eigenvalues, IMA J. Numer. Anal. 34 (2014) 408–501.
[17] P. Lancaster, P.J. Psarrakos, On the pseudospectra of matrix polynomials, SIAM J. Matrix Anal. Appl. 27 (2005) 115–129.
[18] C.F. Van Loan, A study of the matrix exponential. Numerical Analysis Report No. 10, University of Manchester, Manchester, UK, August 1975, Reissued as MIMS EPrint 2006.397, Manchester Institute for Mathematical Sciences, The University of Manchester, UK, November 2006.
[19] A.N. Malyshev, A formula for the 2-norm distance from a matrix to the set of matrices with multiple eigenvalues, Numer. Math. 83 (1999) 443–454.
[20] E. Mengi, Locating a nearest matrix with an eigenvalue of prespecified algebraic multiplicity, Numer. Math. 118 (1) (2011) 109–135.
[21] E. Mengi, E.A. Yildirim, M. Kilic, Numerical optimization of eigenvalues of Hermitian matrix functions, SIAM J. Matrix Anal. Appl. 35 (2014) 699–724.
[22] N. Papathanasiou, P.J. Psarrakos, The distance from a matrix polynomial to matrix polynomials with a prescribed multiple eigenvalue, Linear Algebra Appl. 429 (2008) 1453–1477.
[23] S.A. Piyavskii, An algorithm for finding the absolute extremum of a function, USSR Comput. Math. Math. Phys. 12 (1972) 57–67.
[24] P.J. Psarrakos, Distance bounds for prescribed multiple eigenvalues of matrix polynomials, Linear Algebra Appl. 436 (2012) 4107–4119.
[25] L. Qiu, B. Bernhardsson, A. Rantzer, E.J. Davison, P.M. Young, J.C. Doyle, A formula for computation of the real stability radius, Automatica 31 (6) (1995) 879–890.
[26] A. Quarteroni, R. Sacco, F. Saleri, Numerical Mathematics, Texts Appl. Math., Springer, Paris, France, 2007.
[27] A. Ruhe, Properties of a matrix with a very ill-conditioned eigenproblem, Numer. Math. 15 (1970) 57–60.
[28] B. Shubert, A sequential method seeking the global maximum of a function, SIAM J. Numer. Anal. 9 (1972) 379–388.
[29] F. Tisseur, Backward error and condition of polynomial eigenvalue problems, Linear Algebra Appl. 309 (2000) 339–361.
[30] F. Tisseur, N.J. Higham, Structured pseudospectra for polynomial eigenvalue problems with applications, SIAM J. Matrix Anal. Appl. 23 (2001) 187–208.
[31] J.H. Wilkinson, Note on matrices with a very ill-conditioned eigenproblem, Numer. Math. 19 (1972) 176–178.
[32] J.H. Wilkinson, On neighbouring matrices with quadratic elementary divisors, Numer. Math. 44 (1984) 1–21.