RESEARCH

Randomized methods for computing joint eigenvalues, with applications to multiparameter eigenvalue problems and root finding

Haoze He[1] · Daniel Kressner[1] · Bor Plestenjak[2,3]

Abstract

It is well known that a family of $n \times n$ commuting matrices can be simultaneously triangularized by a unitary similarity transformation. The diagonal entries of the triangular matrices define the n joint eigenvalues of the family. In this work, we consider the task of numerically computing approximations to such joint eigenvalues for a family of (nearly) commuting matrices. This task arises, for example, in solvers for multiparameter eigenvalue problems and systems of multivariate polynomials, which are our main motivations. We propose and analyze a simple approach that computes eigenvalues as one-sided or two-sided Rayleigh quotients from eigenvectors of a random linear combination of the matrices in the family. We provide some analysis and numerous numerical examples, showing that such randomized approaches can compute semisimple joint eigenvalues accurately and lead to improved performance of existing solvers.

Keywords Commuting matrices · Joint eigenvalue · Rayleigh quotient · Randomized numerical linear algebra · Multiparameter eigenvalue problem · Polynomial system

✉ Bor Plestenjak
   bor.plestenjak@fmf.uni-lj.si

   Haoze He
   haoze.he@epfl.ch

   Daniel Kressner
   daniel.kressner@epfl.ch

[1] École Polytechnique Fédérale de Lausanne (EPFL), Institute of Mathematics, 1015 Lausanne, Switzerland

[2] Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

[3] Institute of Mathematics, Physics and Mechanics, Jadranska 19, 1000 Ljubljana, Slovenia

 Springer

Mathematics Subject Classification (2010) 65F15 · 15A27 · 68W20 · 15A69 · 65H04

1 Introduction

Let $\mathcal{A} = \{A_1, \ldots, A_d\}$ be a commuting family of $n \times n$ complex matrices, i.e., $A_j A_k = A_k A_j$ for all $1 \leq j, k \leq d$. It is well known that there exists a unitary matrix U such that all matrices $U^* A_1 U, \ldots, U^* A_d U$ are upper triangular, see, e.g., [23, Theorem 2.3.3]. The n d-tuples containing the diagonal elements of $U^* A_1 U, \ldots, U^* A_d U$ are called the *joint eigenvalues* of $\mathcal{A}$. In other words, if we partition $U = [u_1 \cdots u_n]$, then

$$\boldsymbol{\lambda}^{(i)} = (\lambda_1^{(i)}, \ldots, \lambda_d^{(i)}) = (u_i^* A_1 u_i, \ldots, u_i^* A_d u_i), \qquad i = 1, \ldots, n, \qquad (1)$$

are the joint eigenvalues. For every joint eigenvalue $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ of $\mathcal{A}$ there exists a nonzero *common eigenvector* $x \in \mathbb{C}^n$, such that $A_i x = \lambda_i x$ for $i = 1, \ldots, d$.

The task of computing joint eigenvalues of a commuting family arises in a variety of applications. Our main motivation are numerical methods for multiparameter eigenvalue problems as well as multivariate root finding problems.

In this work, we will investigate solvers for joint eigenvalue problems that utilize a random linear combination

$$A(\mu) = \mu_1 A_1 + \mu_2 A_2 + \cdots + \mu_d A_d, \qquad (2)$$

where $\mu = [\mu_1 \cdots \mu_d]^T$ is a random vector from the uniform distribution on the unit sphere in $\mathbb{C}^d$. Assuming that $A(\mu)$ can be diagonalized, we use its eigenvector matrix X to diagonalize $A_1, \ldots, A_d$. More precisely, we extract the joint eigenvalues $\boldsymbol{\lambda}^{(i)} = (\lambda_1^{(i)}, \ldots, \lambda_d^{(i)})$ from the diagonal elements of $X^{-1} A_1 X, \ldots, X^{-1} A_d X$.

The idea of using a random linear combination like (2) is not new and has been used several times. For example, in [8, 31, 35, 47] the unitary transformation U from the Schur decomposition $A(\mu) = U^* R U$ is used to simultaneously transform all matrices from $\mathcal{A}$ to *block* upper triangular form. Using the Schur decomposition, however, inevitably requires eigenvalue clustering, that is, the Schur decomposition of $A(\mu)$ needs to be reordered so that multiple eigenvalues are grouped together. Eigenvalue clustering is a numerically subtle task sensitive to roundoff error. On the other hand, using eigenvectors for (partial) diagonalization does not require clustering and is thus simpler to use. Although unitary transformations appear to be numerically more appealing, in practice one often gets equally good or even better results by using eigenvectors for, e.g., solving multivariate root finding problems.

For commuting *Hermitian* matrices $A_1, \ldots, A_d$, there is a unitary matrix U that jointly diagonalizes all matrices. For this significantly simpler situation, compared to the general non-Hermitian case, randomized methods based on (2) have recently been analyzed in [18], establishing favorable robustness and stability properties.

The rest of this work is organized as follows. In Section 2, we provide some basic theory on commuting families and eigenvalue condition numbers. In Section 3 we present our main algorithm (Algorithm 1) that uses a random linear combination to extract eigenvectors, followed by one- and two-sided Rayleigh quotients to extract joint

eigenvalues. Section 4 contains our main theoretical results, asymptotic error bounds of the approximations returned by Algorithm 1. A probabilistic analysis is performed to account for randomness of the linear combinations and to better understand the effect of perturbations. Section 5 contains several numerical examples with synthetic data to understand different aspects of our proposed approach and the error bounds, highlighting that semisimple joint eigenvalues can be computed accurately with high probability. In Sections 6 and 7 we present applications to multiparameter eigenvalue problems and solving systems of multivariate polynomials, respectively.

2 Preliminaries

2.1 Basic properties of commuting families

A family of matrices $\mathcal{A} = \{A_k \in \mathbb{C}^{n \times n}\}_{k=1}^d$ is *simultaneously diagonalizable* [23, Definition 1.3.20] if there exists an invertible matrix X such that all matrices $X^{-1}A_1 X, \ldots, X^{-1}A_d X$ are diagonal. For a simultaneously diagonalizable family $\mathcal{A}$, it clearly holds that every A_k is diagonalizable and $\mathcal{A}$ is a commuting family. It turns out that the converse also holds.

Theorem 2.1 ([23, Theorem 1.3.21], [11, Corollary 3, p. 224]) *A family of matrices $\mathcal{A}$ is simultaneously diagonalizable if and only if $\mathcal{A}$ is a commuting family and each A_k is diagonalizable.*

We recall from (1) the definition of joint eigenvalues $\boldsymbol{\lambda}^{(1)}, \ldots, \boldsymbol{\lambda}^{(n)}$ for an $n \times n$ commuting family of d matrices. Equivalently, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ is a joint eigenvalue if the eigenspaces $\ker(A_k - \lambda_k I_n)$ have a nontrivial intersection. This observation leads to the following generalized notion of eigenspaces.

Definition 2.2 ([3, Section 3.6], [28]) *Given a joint eigenvalue $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in \mathbb{C}^d$ of a commuting family $\mathcal{A}$, the* common (right) eigenspace *belonging to $\boldsymbol{\lambda}$ is defined as*

$$\ker(\mathcal{A} - \boldsymbol{\lambda} I_n) := \bigcap_{k=1}^d \ker(A_k - \lambda_k I_n),$$

whereas the common *left* eigenspace *belonging to $\boldsymbol{\lambda}$ is defined as*

$$\ker\left((\mathcal{A} - \boldsymbol{\lambda} I_n)^*\right) := \bigcap_{k=1}^d \ker\left((A_k - \lambda_k I_n)^*\right).$$

The *algebraic multiplicity* of $\boldsymbol{\lambda}$ is defined as the number of times that $\boldsymbol{\lambda}$ appears in the list of n joint eigenvalues $\boldsymbol{\lambda}^{(1)}, \ldots, \boldsymbol{\lambda}^{(n)}$; see [28] and [12, Chapter 2] for an equivalent definition using *root subspaces*. The *geometric multiplicity* of $\boldsymbol{\lambda}$ is defined as $\dim(\ker(\mathcal{A} - \boldsymbol{\lambda} I_n))$, the dimension of the common eigenspace. We say that $\boldsymbol{\lambda}$ is *simple* when its algebraic multiplicity is one and $\boldsymbol{\lambda}$ is *semisimple* if its algebraic and geometric multiplicities are equal.

Like for standard eigenvalue problems, the geometric multiplicity is bounded by the algebraic multiplicity. This fact can be derived from choosing a matrix U that transforms all matrices to triangular form such that its first columns contain a basis of the common eigenspace $\ker(\mathcal{A} - \lambda \boldsymbol{I}_n)$. We will use a similar argument to show the following transformation, which will play an important role in our analysis in Section 4.

Lemma 2.3 *Given a commuting family $\mathcal{A}$, let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)$ be a semisimple joint eigenvalue of multiplicity p. Then there exist invertible matrices $X = [X_1\ X_2] \in \mathbb{C}^{n\times n}$, $Y = [Y_1\ Y_2] \in \mathbb{C}^{n\times n}$ such that*

$$X_1^* X_1 = I_p, \qquad Y^* X = I_n, \qquad Y^* A_k X = \begin{bmatrix} \lambda_k I_p & 0 \\ 0 & A_{22}^{(k)} \end{bmatrix}, \quad k = 1, \dots, d, \quad (3)$$

for some $A_{22}^{(k)} \in \mathbb{C}^{(n-p)\times(n-p)}$.

To show Lemma 2.3, we need the following simple but important observation. From the definition (1) of the joint eigenvalues $\boldsymbol{\lambda}^{(i)} = (\lambda_1^{(i)}, \dots, \lambda_d^{(i)})$, $i = 1, \dots, n$, it follows that that the eigenvalues of the linear combination $A(\mu)$ from (2) are given by

$$\lambda_i(\mu) = \mu_1 \lambda_1^{(i)} + \mu_2 \lambda_2^{(i)} + \cdots + \mu_d \lambda_d^{(i)}, \quad i = 1, \dots, n.$$

We then have the following result from [18, Lemma 1].

Lemma 2.4 *Let a commuting family $\mathcal{A}$, joint eigenvalues $\boldsymbol{\lambda}^{(i)}$ and $\lambda_i(\mu)$ be defined as above. Then the following statement holds generically (with respect to $\mu \in \mathbb{C}^d$) for all $i, j = 1, \dots, d$: The equality $\lambda_i(\mu) = \lambda_j(\mu)$ holds if and only if $\boldsymbol{\lambda}^{(i)} = \boldsymbol{\lambda}^{(j)}$.*

As a corollary of Lemma 2.4 it follows for a simultaneously diagonalizable family $\mathcal{A}$ that $\ker(\mathcal{A} - \lambda \boldsymbol{I}_n) = \ker(A(\mu) - \lambda(\mu) I_n)$ holds for generic $\mu \in \mathbb{C}^d$, i.e., generic linear combinations preserve common eigenspaces. We can now prove Lemma 2.3.

Proof of Lemma 2.3 Let the columns of $U_1 = [u_1 \cdots u_p] \in \mathbb{C}^{n\times p}$ be an orthonormal basis of the common eigenspace $\ker(\mathcal{A} - \lambda \boldsymbol{I}_n)$, which, by assumption, has dimension p. Setting a unitary $U = [U_1\ U_2] \in \mathbb{C}^{n\times n}$ yields

$$T_k = U^* A_k U = \begin{bmatrix} \lambda_k I_p & A_{12}^{(k)} \\ 0 & A_{22}^{(k)} \end{bmatrix}, \quad k = 1, \dots, d, \quad T(\mu) = U^* A(\mu) U = \begin{bmatrix} \lambda(\mu) I_p & A_{12}(\mu) \\ 0 & A_{22}(\mu) \end{bmatrix},$$

for every $\mu \in \mathbb{C}^d$. Because $\boldsymbol{\lambda}$ has multiplicity p, it cannot be a joint eigenvalue for the commuting family $A_{22}^{(1)}, \dots, A_{22}^{(d)}$. By Lemma 2.4, there exists $\mu \in \mathbb{C}^d$ such that $\lambda(\mu)$ is not an eigenvalue of $A_{22}(\mu)$. Standard arguments [42, Theorem 1.5, Chapter V] imply that there exists $R \in \mathbb{C}^{p\times(n-p)}$ such that

$$\overline{T}(\mu) := \overline{Y}^* T(\mu) \overline{X} = \begin{bmatrix} \lambda(\mu) I_p & 0 \\ 0 & A_{22}(\mu) \end{bmatrix} \quad \text{for} \quad \overline{X} = \begin{bmatrix} I_p & R \\ 0 & I_{n-p} \end{bmatrix}, \quad \overline{Y} = \begin{bmatrix} I_p & 0 \\ -R^* & I_{n-p} \end{bmatrix}.$$

Note that $\overline{Y}^*\overline{X} = I_n$. Because $\overline{T}(\mu)$ commutes with every $\overline{Y}^*T_k\overline{X}$ and $\lambda(\mu)$ is not an eigenvalue of $A_{22}(\mu)$, it follows that

$$\overline{Y}^*T_k\overline{X} = \begin{bmatrix} \lambda_k I_p & 0 \\ 0 & A_{22}^{(k)} \end{bmatrix}, \quad k = 1, \ldots d.$$

The proof is completed by setting $X = U\overline{X}$ and $Y = U\overline{Y}$.   □

A consequence of Lemma 2.3 is the following result that gives another equivalent characterization of simultaneous diagonalizability (in addition to Theorem 2.1).

Theorem 2.5 *A commuting family $\mathcal{A} = \{A_k \in \mathbb{C}^{n \times n}\}_{k=1}^d$ is simultaneously diagonalizable if and only if all joint eigenvalues of $\mathcal{A}$ are semisimple.*

Proof Let X be an invertible matrix such that each $D_k = X^{-1}A_kX$ is diagonal. Since similarity transformations preserve joint eigenvalues, the semisimplicity of every joint eigenvalue $\boldsymbol{\lambda}$ follows from $\dim(\ker(\mathcal{D} - \boldsymbol{\lambda} I_n)) = \dim(\ker(\mathcal{A} - \boldsymbol{\lambda} I_n))$, where $\mathcal{D} = \{D_1, \ldots, D_p\}$.

If all joint eigenvalues of $\mathcal{A}$ are semisimple, the same property holds for the diagonal blocks $\{A_{22}^{(k)}\}_{k=1}^d$ in (3). In turn, the proof in the other direction follows from the repeated application of Lemma 2.3.   □

We now relate the diagonalizability of the family to the diagonalizability of a linear combination.

Theorem 2.6 *Let $\mathcal{A} = \{A_k \in \mathbb{C}^{n \times n}\}_{k=1}^d$ be a simultaneously diagonalizable commuting family. Then the following statement holds generically (with respect to $\mu \in \mathbb{C}^d$): If an invertible matrix X diagonalizes $A(\mu)$, then X diagonalizes A_k for every $k = 1, \ldots, d$.*

Proof The proof proceeds along the lines of the proof of [18, Theorem 2]. Without loss of generality, we may assume that the first p columns of X span the right eigenspace belonging to an eigenvalue $\lambda_1(\mu)$ of $A(\mu)$. Then

$$X^{-1}A(\mu)X = \begin{bmatrix} \lambda_1(\mu)I_p & 0 \\ 0 & A_{22}(\mu) \end{bmatrix} \quad \text{and} \quad X^{-1}A_kX = \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ A_{21}^{(k)} & A_{22}^{(k)} \end{bmatrix}, \quad k = 1, \ldots, d.$$

Since $A(\mu)$ and A_k commute, we have $A_{12}^{(k)}(A_{22}(\mu) - \lambda_1(\mu)I_{n-p}) = 0$, $(A_{22}(\mu) - \lambda_1(\mu)I_{n-p})A_{21}^{(k)} = 0$, and $A_{22}(\mu)A_{22}^{(k)} = A_{22}^{(k)}A_{22}(\mu)$. Since $A_{22}(\mu) - \lambda_1(\mu)I_{n-p}$ is invertible, $A_{12}^{(k)} = 0$ and $A_{21}^{(k)} = 0$. It follows from Lemma 2.4 that $A_{11}^{(k)}$ is a multiple of the identity matrix for a generic μ. We continue the proof by considering $A_{22}(\mu)$ and $A_{22}^{(k)}$ in the same way.   □

If all joint eigenvalues of $\mathcal{A}$ are simple, then the common eigenvectors are uniquely determined (up to scaling). Therefore, for generic $\mu \in \mathbb{C}^d$, we get essentially the same eigenvector matrix X by diagonalizing $A(\mu)$.

If the diagonalizability assumption on $\mathcal{A}$ in Theorem 2.6 is not satisfied, then, by Theorem 2.1, $\mathcal{A}$ contains at least one nondiagonalizable matrix. It can be shown that this implies that $A(\mu)$ is nondiagonalizable as well for almost every μ. Note that the canonical structure of a general commuting family $\mathcal{A}$ can be quite complex, see, e.g., [12, 27, 43], and the matrices A_k do not necessarily have the same Jordan structure. We will touch on this case again in Section 3.

2.2 Eigenvalue condition numbers

Errors, due to roundoff or other sources, inevitably destroy commutativity properties. Hence, in practical situations one usually deals with a perturbed, *nearly* commuting family of matrices $\widetilde{\mathcal{A}} = \{\widetilde{A}_1, \ldots, \widetilde{A}_d\}$, with $\widetilde{A}_i = A_i + \epsilon E_i$ for $i = 1, \ldots, d$. To study the impact of these perturbations on eigenvalues, we first recall the notion of eigenvalue condition numbers.

For a simple eigenvalue λ of a matrix $A \in \mathbb{C}^{n \times n}$ it is well known [13, 16, 23] that λ is differentiable with respect to A and that the corresponding eigenvalue $\widetilde{\lambda}(\epsilon)$ of the perturbed matrix $A + \epsilon E$ admits the expansion

$$\widetilde{\lambda} = \lambda + (y^* E x)\epsilon + \mathcal{O}(\epsilon^2), \tag{4}$$

where x and y are right/left eigenvectors belonging to λ with the normalization chosen such that $y^* x = 1$ and $\|x\|_2 = 1$. Here and in the following, $\|\cdot\|_2$ denotes the 2-norm of a vector or matrix. Motivated by (4), the condition number of λ is defined as follows.

Definition 2.7 *With the notation introduced above, $\kappa(\lambda) = \|y\|_2$ is called the eigenvalue condition number of a simple eigenvalue λ.*

The matrix $P_\lambda = x y^*$ is the spectral projection [2, 25, 41] corresponding to the simple eigenvalue λ. Noting that $\kappa(\lambda) = \|P_\lambda\|_2$ motivates the following extension to semisimple eigenvalues.

Definition 2.8 *The* spectral projection *corresponding to a semisimple eigenvalue λ of a matrix A is the matrix P_λ such that P_λ is a projection, the column space of P_λ is the right eigenspace $\ker(A - \lambda I_n)$ and the row space of P_λ is the left eigenspace $\ker((A - \lambda I_n)^*)$. The eigenvalue condition number of λ is defined as $\kappa(\lambda) = \|P_\lambda\|_2$.*

Note that LAPACK [2] also uses norms of spectral projections to define condition numbers for semisimple eigenvalues and clusters of eigenvalues.

Similar to (3), if λ is a semisimple eigenvalue of multiplicity p of a matrix $A \in \mathbb{C}^{n \times n}$, then there exist invertible matrices $X = [X_1 \ X_2] \in \mathbb{C}^{n \times n}$, $Y = [Y_1 \ Y_2] \in \mathbb{C}^{n \times n}$ satisfying

$$X_1^* X_1 = I_p, \quad Y^* X = I_n, \quad Y^* A X = \begin{bmatrix} \lambda I_p & 0 \\ 0 & A_{22} \end{bmatrix}$$

for some $A_{22} \in \mathbb{C}^{(n-p) \times (n-p)}$. It follows that $X_1 Y_1^*$ is the spectral projection corresponding to λ and, hence, $\kappa(\lambda) = \|Y_1\|_2$.

The following result extends the expansion (4) to semisimple eigenvalues.

Theorem 2.9 *Given $A \in \mathbb{C}^{n \times n}$, let λ be a semisimple eigenvalue of multiplicity p and let X_1 and Y_1 be bases of $\ker(A - \lambda I_n)$ and $\ker((A - \lambda I_n)^*)$, respectively, such that $Y_1^* X_1 = I_p$. Then $A + \epsilon E$ has p eigenvalues satisfying the expansion*

$$\widetilde{\lambda}_i = \lambda_i + \eta_i \epsilon + o(\epsilon), \quad i = 1, \ldots, p,$$

where $\eta_1, \ldots, \eta_p$ are the eigenvalues of $Y_1^* E X_1$.

Proof This is a particular case covered by Theorem 6 in [30] on eigenvalues of analytic matrix functions of the form $L(\lambda, \epsilon)$, where we take $L(\lambda, \epsilon) = A - \lambda I + \epsilon E$. □

When imposing the normalization $X_1^* X_1 = I_p$, it follows that $|\eta_i| \leq \|Y_1^* E X_1\|_2 \leq \|Y_1\|_2 \|E\|_2$, which provides another justification for defining $\|Y_1\|_2$ to be the eigenvalue condition number.

By extending the notion of spectral projections to a commuting family $\mathcal{A}$, an extension of the condition number to semisimple eigenvalues follows naturally.

Definition 2.10 *The* spectral projection *corresponding to a semisimple joint eigenvalue $\boldsymbol{\lambda}$ of a commuting family $\mathcal{A}$ is the matrix $P_{\boldsymbol{\lambda}}$ such that $P_{\boldsymbol{\lambda}}$ is a projection, the column space of $P_{\boldsymbol{\lambda}}$ is the right common eigenspace $\ker(\mathcal{A} - \boldsymbol{\lambda} \boldsymbol{I}_n)$ and the row space of $P_{\boldsymbol{\lambda}}$ is the left common eigenspace $\ker((\mathcal{A} - \boldsymbol{\lambda} \boldsymbol{I}_n)^*)$ (see Definition 2.2). The eigenvalue condition number of $\boldsymbol{\lambda}$ is defined as $\kappa(\boldsymbol{\lambda}) = \|P_{\boldsymbol{\lambda}}\|_2$.*

The block diagonal decomposition (3) implies again that $\kappa(\boldsymbol{\lambda}) = \|Y_1\|_2$.

3 Rayleigh quotients

Algorithm 1 summarizes our simple approach for approximating joint eigenvalues of nearly commuting family. The two types of Rayleigh quotients used for this purpose are discussed in more detail below.

Algorithm 1 Randomized Joint Eigenvalue Approximation

Input: A nearly commuting family $\widetilde{\mathcal{A}} = \{\widetilde{A}_1, \ldots, \widetilde{A}_d\}$, opt $\in$ {RQ1, RQ2}.
Output: Approximations of joint eigenvalues of $\widetilde{\mathcal{A}}$.
1: Draw $\mu \in \mathbb{C}^d$ from the uniform distribution on the unit sphere.
2: Compute $\widetilde{A}(\mu) = \mu_1 \widetilde{A}_1 + \cdots + \mu_d \widetilde{A}_d$.
3: Compute invertible matrices $\widetilde{X}, \widetilde{Y}$ such that the columns of $\widetilde{X}$ have norm 1,
    $\widetilde{Y}^* \widetilde{X} = I_n$, and $\widetilde{Y}^* \widetilde{A}(\mu) \widetilde{X}$ is diagonal.
4: **if** opt = RQ1
5:    **return** $\widetilde{\lambda}_{\text{RQ1}}^{(i)} = (\widetilde{x}_i^* \widetilde{A}_1 \widetilde{x}_i, \ldots, \widetilde{x}_i^* \widetilde{A}_d \widetilde{x}_i), \quad i = 1, \ldots, n.$
6: **else if** opt = RQ2
7:    **return** $\widetilde{\lambda}_{\text{RQ2}}^{(i)} = (\widetilde{y}_i^* \widetilde{A}_1 \widetilde{x}_i, \ldots, \widetilde{y}_i^* \widetilde{A}_d \widetilde{x}_i), \quad i = 1, \ldots, n.$
8: **end if**

The random vector $\mu \in \mathbb{C}^d$ in Line 1 of Algorithm 1 can be generated by drawing a complex Gaussian random vector g (that is, the real and imaginary parts of its entries are i.i.d $\mathcal{N}(0, 1/2)$) and setting $\mu = g/\|g\|_2$. Note that the matrix $\widetilde{Y} = \widetilde{X}^{-*}$ contains

the left eigenvectors of $\widetilde{A}(\mu)$. The normalization conditions in Line 3 correspond to what is computed by standard eigenvalue solvers, such as CGEEVX in LAPACK [2]. In the following, we provide details on the key quantities $\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ1}}^{(i)}$ and $\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ2}}^{(i)}$ of Algorithm 1.

One-sided Rayleigh quotient From each normalized common (right) eigenvector x_i of $\mathcal{A}$, we can compute the *(one-sided) Rayleigh quotient*

$$\boldsymbol{\lambda}_{\mathrm{RQ1}}^{(i)}(x_i, \mathcal{A}) := (x_i^* A_1 x_i, \ldots, x_i^* A_d x_i). \tag{5}$$

In the absence of noise, the quantity $\boldsymbol{\lambda}_{\mathrm{RQ1}}^{(i)}(x_i, \mathcal{A})$ computed in Line 5 of Algorithm 1 equals, by Theorem 2.6, the joint eigenvalue $\boldsymbol{\lambda}^{(i)}$ of $\mathcal{A}$ with probability 1. In the presence of errors and noise, if $\widetilde{x}_i = x_i + \epsilon \Delta x_i$ is an approximation to the exact common eigenvector x_i and we apply it to $\widetilde{A}_k = A_k + \epsilon E_k$, then for each component $\lambda_k^{(i)}$ of $\boldsymbol{\lambda}^{(i)}$,

$$\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ1}}^{(i)}(\widetilde{x}_i, \widetilde{\mathcal{A}})_k - \lambda_k^{(i)} = \widetilde{x}_i^*(A_k - \lambda_k^{(i)} I)\widetilde{x}_i = \left(\widetilde{x}_i^*(A_k - \lambda_k^{(i)} I)\Delta x_i + \widetilde{x}_i^* E_k \widetilde{x}_i \right)\epsilon.$$

A rough error estimate is

$$|\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ1}}^{(i)}(\widetilde{x}_i, \widetilde{\mathcal{A}})_k - \lambda_k^{(i)}| \leq \left(\|(A_k - \lambda_k^{(i)} I)\Delta x_i\|_2 + \|E_k\|_2 \right)\epsilon. \tag{6}$$

In Section 4, we will derive a more refined bound.

Two-sided Rayleigh quotient From each pair of left and right common eigenvectors y_i, x_i of $\mathcal{A}$, we can compute the *two-sided Rayleigh quotient*

$$\boldsymbol{\lambda}_{\mathrm{RQ2}}^{(i)}(x_i, y_i, \mathcal{A}) := \left(\frac{y_i^* A_1 x_i}{y_i^* x_i}, \ldots, \frac{y_i^* A_d x_i}{y_i^* x_i} \right) = \left(y_i^* A_1 x_i, \ldots, y_i^* A_d x_i \right), \tag{7}$$

using our normalization assumptions on X and Y. In the absence of noise, the quantity $\boldsymbol{\lambda}_{\mathrm{RQ2}}^{(i)}(x_i, y_i, \mathcal{A})$ computed in Line 7 of Algorithm 1 is again the joint eigenvalue $\boldsymbol{\lambda}^{(i)}$ of $\mathcal{A}$ with probability 1.

If $\widetilde{x}_i = x_i + \epsilon \Delta x_i$, $\widetilde{y}_i = y_i + \epsilon \Delta y_i$ are approximations to the right/left common eigenvectors and $\widetilde{A}_k = A_k + \epsilon E_k$, then for each component $\lambda_k^{(i)}$ of $\boldsymbol{\lambda}^{(i)}$,

$$\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ2}}^{(i)}(\widetilde{x}_i, \widetilde{y}_i, \widetilde{\mathcal{A}})_k - \lambda_k^{(i)} = \frac{\epsilon}{\widetilde{y}_i^* \widetilde{x}_i}(\widetilde{y}_i^* E_k \widetilde{x}_i) + \frac{\epsilon^2}{\widetilde{y}_i^* \widetilde{x}_i} \Delta y_i^*(A_k - \lambda_k^{(i)} I)\Delta x_i,$$

yielding a rough estimate

$$|\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ2}}(\widetilde{x}_i, \widetilde{y}_i, \widetilde{A}_k)| \leq \|E_k\|_2 \|y_i\|_2 \epsilon + |\Delta y_i^*(A_k - \lambda_k^{(i)} I)\Delta x_i|\epsilon^2. \tag{8}$$

Comparing (8) to (6), we see that the error of the two-sided Rayleigh quotient depends on the condition number of a joint eigenvalue. On the other hand, the second term in (8) involves Δx_i and Δy_i. For sufficiently small ϵ, if we know that $\|\Delta x_i\|_2$ and

$\|\Delta y_i\|_2$ are independent of ϵ, this term should be smaller than the first term in (6). We will present a more detailed comparison between $\widetilde{\lambda}_{\text{RQ2}}^{(i)}$ and $\widetilde{\lambda}_{\text{RQ1}}^{(i)}$ in Section 4.

If $\mathcal{A}$ is not simultaneously diagonalizable, it follows from the discussion at the end of Section 2.1 that $A(\mu)$ is not diagonalizable. In such case there is no invertible matrix X of right eigenvectors for $A(\mu)$. Under perturbations, such X almost surely exists but can be expected to be extremely ill-conditioned, so there is little hope of getting highly accurate approximations for all eigenvalues of $\mathcal{A}$ from (7). However, we will show in Example 5.5 that even in such an extreme situation one still obtains meaningful approximations. We also note that when (7) cannot be computed numerically, it is more reliable to resort to (5), which is well defined even for a singular X.

4 Rayleigh quotient approximation error

In this section, we analyze the quality of the Rayleigh quotient approximations produced by Algorithm 1 applied to a nearly commuting family $\widetilde{\mathcal{A}}$. As in Algorithm 1, we will assume that the eigenvector matrix $\widetilde{X}$ of $\widetilde{A}(\mu)$ is normalized such that its columns have norm 1. This scaling gives a quasi-optimal condition number for X; see [9, Theorem 2]. In the (unlikely) case that $\widetilde{A}(\mu)$ has a multiple eigenvalue, we will assume that this eigenvalue is semisimple and X contains an orthonormal basis of the corresponding eigenspace.

Our analysis consists of two parts. In Section 4.1 we will derive deterministic bounds that depend on the particular choice of μ and make no assumption on the perturbation. In Sections 4.2 and 4.3 we will derive probabilistic bounds assuming that μ is a random vector from the uniform distribution on the unit sphere in $\mathbb{C}^d$ and considering a random model for the perturbations.

4.1 Structural bounds

Lemma 4.1 below provides error bounds for the one- and two-sided Rayleigh quotients returned by Algorithm 1. We recall, see Lemma 2.3 and the discussion in Section 2.1, that we can choose the right/left eigenbases $X_1, Y_1 \in \mathbb{C}^{n \times p}$ associated with a semisimple joint eigenvalue λ such that there are invertible matrices $X = [X_1 \ X_2] \in \mathbb{C}^{n \times n}$, $Y = [Y_1 \ Y_2] \in \mathbb{C}^{n \times n}$ satisfying

$$X_1^* X_1 = I_p, \qquad Y^* X = I_n, \qquad Y^* A_k X = \begin{bmatrix} \lambda_k I_p & 0 \\ 0 & A_{22}^{(k)} \end{bmatrix}, \quad k = 1, \dots, d, \qquad (9)$$

for some $A_{22}^{(k)} \in \mathbb{C}^{(n-p) \times (n-p)}$.

Lemma 4.1 *For a commuting family $\mathcal{A} = \{A_k \in \mathbb{C}^{n \times n}\}_{k=1}^d$, let $\lambda = (\lambda_1, \dots, \lambda_d)$ be a semisimple joint eigenvalue of multiplicity p, and let $X = [X_1 \ X_2] \in \mathbb{C}^{n \times n}$, $Y = [Y_1 \ Y_2] \in \mathbb{C}^{n \times n}$ be such that (9) is satisfied. Suppose that $\mu \in \mathbb{C}^d$, $\|\mu\|_2 = 1$, is such that $\lambda(\mu)$ is not an eigenvalue of $A_{22}(\mu)$. Consider a perturbed family $\widetilde{\mathcal{A}} = \{A_k + \epsilon E_k\}_{k=1}^d$ with $\|E_1\|_F^2 + \cdots + \|E_d\|_F^2 = 1$ and $\epsilon > 0$, such that the p*

eigenvalues of $\widetilde{A}(\mu)$ closest to $\lambda(\mu)$ remain semisimple. Suppose that $\widetilde{X}_1 = [\widetilde{x}_1 \cdots \widetilde{x}_p]$ and $\widetilde{Y}_1 = [\widetilde{y}_1 \cdots \widetilde{y}_p]$ contain the corresponding right/left eigenvectors of $\widetilde{A}(\mu)$, normalized such that $\widetilde{Y}_1^* \widetilde{X}_1 = I_p$ and the columns of $\widetilde{X}_1$ have norm 1. Then the inequalities

$$|\widetilde{x}_i^* \widetilde{A}_k \widetilde{x}_i - \lambda_k| \le (\|E_k\|_2 + \|X_2\|_2 \|Y_2\|_2 \|D_k(\mu)\|_2)\,\epsilon + \mathcal{O}(\epsilon^2), \qquad (10)$$

$$|\widetilde{y}_i^* \widetilde{A}_k \widetilde{x}_i - \lambda_k| \le \|\widetilde{y}_i\|_2 \|E_k\|_2\,\epsilon + \mathcal{O}(\epsilon^2) \qquad\qquad (11)$$

hold for every $i = 1, \ldots, p$ and $k = 1, \ldots, d$, where

$$D_k(\mu) = \left(A_{22}^{(k)} - \lambda_k I_{n-p}\right)\left(A_{22}(\mu) - \lambda(\mu) I_{n-p}\right)^{-1}. \qquad (12)$$

Proof By definition, the columns of $\widetilde{X}_1$ form a basis for an invariant subspace of $\widetilde{A}(\mu)$. The perturbation result of Lemma A.2 provides another basis $\widehat{X}_1$ of the same subspace such that $\widehat{X}_1 = X_1 + \epsilon \Delta X_1 + \mathcal{O}(\epsilon^2)$ with

$$\Delta X_1 = -X_2(A_{22}(\mu) - \lambda(\mu) I_{n-p})^{-1} Y_2^* E(\mu) X_1. \qquad (13)$$

Because $\widehat{X}_1$ and $\widetilde{X}_1$ span the same subspace, there is an invertible matrix C_1 such that $\widetilde{X}_1 = \widehat{X}_1 C_1$. Moreover, there is a basis $\widehat{Y}_1$ of the space spanned by the columns of $\widetilde{Y}_1$ such that $\widehat{Y}_1^* = Y_1^* + \epsilon \Delta Y_1^* + \mathcal{O}(\epsilon^2)$ with

$$\Delta Y_1^* = -Y_1^* E(\mu) X_2(A_{22}(\mu) - \lambda(\mu) I_{n-p})^{-1} Y_2^*$$

and $\widehat{Y}_1^* \widehat{X}_1 = I_p + \mathcal{O}(\epsilon^2)$. Note that the latter relation implies $\widetilde{Y}_1 = \widehat{Y}_1(C_1^{-*} + \mathcal{O}(\epsilon^2))$.

Letting e_i denote the ith unit vector of length p and using that $\|\widetilde{x}_i\|_2 = 1$, we obtain for the one-sided Rayleigh quotient that

$$
\begin{aligned}
\widetilde{x}_i^* \widetilde{A}_k \widetilde{x}_i - \lambda_k &= \epsilon \widetilde{x}_i^* E_k \widetilde{x}_i + \widetilde{x}_i^*(A_k - \lambda_k I)\widetilde{x}_i \\
&= \epsilon \widetilde{x}_i^* E_k \widetilde{x}_i + \widetilde{x}_i^*(A_k - \lambda_k I)(X_1 + \epsilon \Delta X_1) C_1 e_i + \mathcal{O}(\epsilon^2) \\
&= \epsilon \widetilde{x}_i^* E_k \widetilde{x}_i + \epsilon \widetilde{x}_i^*(A_k - \lambda_k I)\Delta X_1 C_1 e_i + \mathcal{O}(\epsilon^2) \\
&= \epsilon \widetilde{x}_i^* E_k \widetilde{x}_i - \epsilon \widetilde{x}_i^*(A_k - \lambda_k I)X_2(A_{22}(\mu) - \lambda(\mu) I_{n-p})^{-1} Y_2^* E(\mu)\widetilde{x}_i + \mathcal{O}(\epsilon^2) \\
&= \epsilon \widetilde{x}_i^* E_k \widetilde{x}_i - \epsilon \widetilde{x}_i^* X_2 D_k(\mu) Y_2^* E(\mu)\widetilde{x}_i + \mathcal{O}(\epsilon^2),
\end{aligned}
$$

where we used $(A_k - \lambda_k I)X_1 = 0$ for the third and $(A_k - \lambda_k I)X_2 = X_2\left(A_{22}^{(k)} - \lambda_k I_{n-p}\right)$ for the last equality. It follows from $Y_2^* X_2 = I_{n-p}$ and $\|E(\mu)\|_2 \le 1$ that

$$|\widetilde{x}_i^* \widetilde{A}_k \widetilde{x}_i - \lambda_k^{(i)}| \le (\|E_k\|_2 + \|X_2\|_2 \|Y_2\|_2 \|D_k(\mu)\|_2)\,\epsilon + \mathcal{O}(\epsilon^2),$$

which proves (10).

For the two-sided Rayleigh quotient, we get (11) from

$$\widetilde{y}_i^* \widetilde{A}_k \widetilde{x}_i - \lambda_k^{(i)} = \epsilon \widetilde{y}_i^* E_k \widetilde{x}_i + \widetilde{y}_i^*(A_k - \lambda_k I)\widetilde{x}_i$$

$$= \epsilon \widetilde{y}_i^* E_k \widetilde{x}_i + e_i^* C_1^{-*}(Y_1^* + \epsilon \Delta Y_1^*)(A_k - \lambda_k I)(X_1 + \epsilon \Delta X_1)C_1 e_i + \mathcal{O}(\epsilon^2)$$
$$= \epsilon \widetilde{y}_i^* E_k \widetilde{x}_i + \mathcal{O}(\epsilon^2)$$

where we used $\widetilde{y}_i^* \widetilde{x}_i = 1$, $Y_1^*(A_k - \lambda_k I) = 0$, and $(A_k - \lambda_k I)X_1 = 0$. □

The following theorem turns the error bounds of Lemma 4.1 for individual eigenvalue components into error bounds for the whole vector containing a joint eigenvalue. For one-sided Rayleigh quotients, we additionally assume diagonalizability in order to obtain a convenient expression for $\|D_k(\mu)\|_2$.

Theorem 4.2 *Under the notation and assumptions of Lemma 4.1, the following statements hold:*

1) *The two-sided Rayleigh quotients* $\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ2}}^{(i)} := (\widetilde{y}_i^* \widetilde{A}_1 \widetilde{x}_i, \dots, \widetilde{y}_i^* \widetilde{A}_d \widetilde{x}_i) \in \mathbb{C}^d$ *satisfy*

$$\|\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ2}}^{(i)} - \boldsymbol{\lambda}\|_2 \leq \|\widetilde{y}_i\|_2 \epsilon + \mathcal{O}(\epsilon^2), \quad i = 1, \dots, p. \tag{14}$$

2) *Additionally, assume that $\mathcal{A}$ is diagonalizable and let $X = [X_1 \ X_2]$, $Y = [Y_1 \ Y_2]$ be such that $Y^* A_k X$ is diagonal for $k = 1, \dots, d$. Then the one-sided Rayleigh quotients* $\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ1}}^{(i)} := (\widetilde{x}_i^* \widetilde{A}_1 \widetilde{x}_i, \dots, \widetilde{x}_i^* \widetilde{A}_d \widetilde{x}_i) \in \mathbb{C}^d$ *satisfy*

$$\|\widetilde{\boldsymbol{\lambda}}_{\mathrm{RQ1}}^{(i)} - \boldsymbol{\lambda}\|_2 \leq \left(1 + \frac{\sqrt{d}\|X_2\|_2\|Y_2\|_2}{d(\mu)}\right)\epsilon + \mathcal{O}(\epsilon^2), \quad i = 1, \dots, p, \tag{15}$$

where

$$d(\mu) := \min_{\boldsymbol{\lambda}^{(j)} \neq \boldsymbol{\lambda}} \left(\frac{|\lambda_j(\mu) - \lambda(\mu)|}{\|\boldsymbol{\lambda}^{(j)} - \boldsymbol{\lambda}\|_2}\right),$$

and $\boldsymbol{\lambda}^{(1)}, \dots, \boldsymbol{\lambda}^{(n)}$ denote the joint eigenvalues of $\mathcal{A}$.

Proof The first result (14) is obtained from (11) by summing the squared errors and using $\|E_1\|_F^2 + \cdots + \|E_d\|_F^2 = 1$.

The additional assumption for the second result implies that the matrix $D_k(\mu)$ (12) is diagonal and, hence,

$$\|D_k(\mu)\|_2 = \max_{\boldsymbol{\lambda}^{(i)} \neq \boldsymbol{\lambda}} \left(\frac{|\lambda_k^{(i)} - \lambda_k|}{|\lambda_i(\mu) - \lambda(\mu)|}\right) \leq \max_{\boldsymbol{\lambda}^{(i)} \neq \boldsymbol{\lambda}} \left(\frac{\|\boldsymbol{\lambda}^{(i)} - \boldsymbol{\lambda}\|_2}{|\lambda_i(\mu) - \lambda(\mu)|}\right) = \frac{1}{d(\mu)}.$$

Now, (15) is again obtained from (10) by summing the squared errors. □

The bounds of Theorem 4.2 contain three factors that require explanation and additional considerations:

1. The factor $\|\widetilde{y}_i\|_2$ in (14) is the condition number of the ith eigenvalue of $\widetilde{A}(\mu)$. For $p = 1$, this quantity converges to the condition number of $\boldsymbol{\lambda}$ as $\epsilon \to 0$; see Definition 2.10 in Section 2.2. For $p > 1$, the interpretation of this quantity is much more subtle; see Section 4.3 below.

2. The factor $\|X_2\|_2\|Y_2\|_2$ in (15) quantifies how well $\mathcal{A}$ can be diagonalized because it is bounded by the condition number of the matrix X diagonalizing $\mathcal{A}$. In the context of the bound (10) from Lemma 4.1, X only needs to block diagonalize $\mathcal{A}$ and, in turn, $\|X_2\|_2\|Y_2\|_2$ can be chosen to be significantly smaller. In particular, for $p = 1$ we may assume, without loss of generality, that $x_1 = e_1$ and $y_1 = \begin{bmatrix} 1 \\ v \end{bmatrix}$ for some $v \in \mathbb{C}^{n-1}$. Then we can choose $X_2 = \begin{bmatrix} -v^* \\ I_{n-1} \end{bmatrix}$, for which we have $\|X_2\|_2\|Y_2\|_2 = \|y_1\|_2$. In other words, $\|X_2\|_2\|Y_2\|_2$ also becomes the condition number of λ.

3. For unfortunate choices of μ, the quantity $d(\mu)$ in (15) becomes arbitrarily small. In order to keep $1/d(\mu)$ bounded, it is essential to choose μ randomly; we will analyze this choice in the next section.

4.2 A probabilistic bound for one-sided Rayleigh quotients

In order to investigate the quantity $d(\mu)$ in the error bound (15) for one-sided Rayleigh quotients, we now assume that $\mu \sim \mathrm{Unif}(\mathbb{S}_{\mathbb{C}}^{d-1})$, i.e., μ is from the uniform distribution on the unit sphere in $\mathbb{C}^d$ as chosen in Algorithm 1. For this purpose, we will use anti-concentration results.

Lemma 4.3 ([4, Lemma 5.2]) *Let* $\mu \sim \mathrm{Unif}(\mathbb{S}_{\mathbb{C}}^{d-1})$ *and* $v \in \mathbb{S}_{\mathbb{C}}^{d-1}$. *Then* $\mathrm{Prob}\left(|\mu^* v| \le t/\sqrt{d-1}\right) \le t^2$ *holds for all* $t \in [0, 1]$.

Lemma 4.4 *For* $\mu \sim \mathrm{Unif}(\mathbb{S}_{\mathbb{C}}^{d-1})$, *the quantity* $d(\mu)$ *defined in Theorem 4.2 satisfies*

$$\mathrm{Prob}\big(d(\mu) \ge C\big) \ge 1 - (n-p)(d-1)C^2.$$

Proof From the definition of $d(\mu)$, we obtain that

$$\mathrm{Prob}\big(d(\mu) \ge C\big) = \mathrm{Prob}\Big(\min_{\lambda^{(j)} \ne \lambda} \frac{|\lambda_j(\mu) - \lambda(\mu)|}{\|\lambda^{(j)} - \lambda\|_2} \ge C\Big)$$
$$= \mathrm{Prob}\Big(\min_{\lambda^{(j)} \ne \lambda} \frac{|\mu^*(\lambda^{(j)} - \lambda)|}{\|\lambda^{(j)} - \lambda\|_2} \ge C\Big) \ge 1 - (n-p)(d-1)C^2.$$

The last inequality uses Lemma 4.3 combined with the union bound, noting that at most $n - p$ eigenvalues are different from λ. $\qquad\square$

We now use Lemma 4.4 to turn the result of Theorem 4.2 2) into a probabilistic bound.

Theorem 4.5 *Under the notation and assumptions of Theorem 4.2 2), let* $\mu \sim \mathrm{Unif}(\mathbb{S}_{\mathbb{C}}^{d-1})$. *Then for every* $R > 0$ *it holds for* $i = 1, \ldots, p$ *that*

$$\mathrm{Prob}\big(\big\|\widetilde{\lambda}_{\mathrm{RQ1}}^{(i)} - \lambda\big\|_2 \le (1 + \|X_2\|_2\|Y_2\|_2 R)\epsilon + \mathcal{O}(\epsilon^2)\big) \ge 1 - \frac{(n-p)(d-1)d}{R^2}. \tag{16}$$

Proof Setting $\widetilde{R} = \|X_2\|_2\|Y_2\|_2 R$, the inequality (15) from Theorem 4.2 yields

$$\text{Prob}\big(\big\|\widetilde{\lambda}_{\text{RQ1}}^{(i)} - \lambda\big\|_2 \leq (1 + \widetilde{R})\epsilon + \mathcal{O}(\epsilon^2)\big)$$

$$\geq \text{Prob}\left(\frac{\sqrt{d}\|X_2\|_2\|Y_2\|_2}{d(\mu)} \leq \widetilde{R}\right) = \text{Prob}\left(d(\mu) \geq \frac{\sqrt{d}\|X_2\|_2\|Y_2\|_2}{\widetilde{R}}\right)$$

$$\geq 1 - (n-p)(d-1)\frac{d\|X_2\|_2^2\|Y_2\|_2^2}{\widetilde{R}^2} = 1 - \frac{(n-p)(d-1)d}{R^2},$$

where the last inequality uses Lemma 4.4. $\qquad\square$

The inequality (16) mixes a probabilistic with an asymptotic bound, and its interpretation requires some care. We first draw $\mu \sim \text{Unif}(\mathbb{S}_{\mathbb{C}}^{d-1})$. Then there is a constant C_μ, independent of ϵ but possibly depending on μ, such that the approximation error of the one-sided Rayleigh quotient is bounded by $(1 + \|X_2\|_2\|Y_2\|_2 R)\epsilon + C_\mu\epsilon^2$ for every $\epsilon > 0$ with a failure probability that decreases proportionally to $1/R^2$ when the extra factor R increases.

From the discussion after Theorem 4.5, we know that, for $p = 1$, the approximation error of the two-sided Rayleigh quotient is bounded by $\|y_1\|_2\epsilon + \widetilde{C}_\mu\epsilon^2$. As $\|y_1\|_2 \leq \|X_2\|_2\|Y_2\|_2$, our analysis indicates that the two-sided Rayleigh quotient is more accurate for simple, well isolated eigenvalues. Because of second-order terms, this analysis does not extend to clusters of very close eigenvalues. Nevertheless, also for such cases numerical experiments reveal that the one-sided approach is not more reliable than the two-sided approach; see Example 5.4 below for more details.

4.3 On two-sided Rayleigh quotients for semisimple eigenvalues

As mentioned after Theorem 4.2, the factor $\|\widetilde{y}_i\|_2$, featured in the error bound (14) for two-sided Rayleigh quotients, approaches the eigenvalue condition number in the case of a simple eigenvalue. This quantity becomes much harder to interpret for a semisimple eigenvalue λ of multiplicity $p > 1$, because the perturbations may turn λ into p arbitrarily ill-conditioned perturbed eigenvalues . To better understand this phenomenon, we first note that $\widetilde{y}_i = \widetilde{Y}_1 e_i = \widehat{Y}_1 C_1^{-*} e_i = Y_1 C_1^{-*} e_i + \mathcal{O}(\epsilon)$ implies

$$\|\widetilde{y}_i\|_2 \lesssim \|Y_1\|_2 \cdot \|C_1^{-*} e_i\|_2 \leq \|Y_1\|_2 \cdot \|C_1^{-*}\|_2. \tag{17}$$

We recall that C_1 is the matrix defined in the proof of Theorem 4.2, linking the perturbed invariant subspace $\widehat{X}_1$, defined in (13), to the computed basis $\widetilde{X}_1$ of $\widetilde{A}(\mu)$. Because the columns of X_1 are orthonormal, it follows that every column of C_1 has norm $1 + \mathcal{O}(\epsilon)$. Moreover, by definition, the matrix $\widetilde{Y}_1^*\widetilde{A}(\mu)\widetilde{X}_1$ is diagonal and satisfies

$$\widetilde{Y}_1^*\widetilde{A}(\mu)\widetilde{X}_1 - \lambda(\mu)I_p = \widetilde{Y}_1^*(\widetilde{A}(\mu) - \lambda(\mu)I)\widetilde{X}_1$$
$$= C_1^{-1}\widehat{Y}_1^*(\widetilde{A}(\mu) - \lambda(\mu)I)\widehat{X}_1 C_1 = \epsilon C_1^{-1}Y_1^* E(\mu)X_1 C_1 + \mathcal{O}(\epsilon^2).$$

Hence, in first order, C_1 is a matrix that diagonalizes $Y_1^* E(\mu) X_1$ and its columns have norm 1.

Generically, the p eigenvalues of $Y_1^* E(\mu) X_1$ are mutually distinct eigenvalues and therefore this matrix is almost surely diagonalizable. Still, the matrix C_1 can become arbitrarily ill-conditioned. To be able to say more about $\|C_1^{-*}\|_2$, we need to impose some assumptions on the perturbations. In the following, we will consider random perturbations. Specifically, we will consider the case that $Y_1^* E(\mu) X_1$ is a (scaled) complex Gaussian random matrix.

In the following, we will consider a $p \times p$ complex Gaussian random matrix $G_p = (g_{ij})$, scaled such that the real and imaginary parts of its entries g_{ij} are i.i.d. $\mathcal{N}(0, 1/2p)$. The scaling by $2p$ ensures that the norm does not grow with p.

Lemma 4.6 ([5, Lemma 2.2]) *The $p \times p$ complex Gaussian matrix G_p defined above satisfies*

$$\mathrm{Prob}\big(\|G_p\|_2 > (2\sqrt{2} + t)\big) \leq 2\exp(-pt^2)$$

for every $t > 0$ and $R > 0$,

The following result is a corollary of Theorem 1.5 in [5] (obtained by setting $A = 0$ and $\delta = 1$ in the statement of that theorem).

Lemma 4.7 *Let $\lambda_1, \ldots, \lambda_p \in \mathbb{C}$ denote the eigenvalues of a complex Gaussian matrix G_p. Then, for every measurable open set $B \subset \mathbb{C}$, the condition number of λ_i satisfies*

$$\mathbb{E} \int_{\lambda_i \in B} \kappa(\lambda_i)^2 \leq \frac{p^2 \mathrm{vol}(B)}{\pi}.$$

Combining the above two results allows us to control the eigenvector matrix condition number of G_p.

Theorem 4.8 *Let X be a $p \times p$ matrix containing the eigenvectors of a $p \times p$ complex Gaussian matrix G_p defined above, scaled such that the columns of X have unit norm. Then*

$$\mathrm{Prob}\left(\kappa_2(X) \geq R\right) \leq 2\exp(-pt^2) + \frac{p^3(2\sqrt{2} + t)^2}{R^2}$$

holds for every $t > 0$ and $R > 0$.

Proof Letting B be the ball of radius ρ centered at the origin, we can conclude from Lemma 4.7 that

$$\mathbb{E} \int_{\lambda_i \in B} \kappa(\lambda_i)^2 \leq p^2 \rho^2. \tag{18}$$

By [5, Lemma 3.1], $\kappa_2(X)^2 \leq p(\kappa(\lambda_1)^2 + \cdots + \kappa(\lambda_p)^2)$. Applying Markov's inequality to (18) gives

$$\mathrm{Prob}\left(\kappa_2(X)^2/p \geq s^2 \cap \|G_p\|_2 \leq \rho\right) \leq p^2 \rho^2/s^2,$$

or, equivalently,

$$\mathrm{Prob}\left(\kappa_2(X) \geq R \cap \|G_p\|_2 \leq \rho\right) \leq p^3 \rho^2/R^2,$$

where we used that $\|G_p\|_2 \leq \rho$ implies $|\lambda_i| \leq \rho$. The proof is completed by using

$$\text{Prob}\left(\kappa_2(X) \geq R\right) \leq \text{Prob}\left(\kappa_2(X) \geq R \cap \|G_p\|_2 \leq \rho\right) + \text{Prob}\left(\|G_p\|_2 \geq \rho\right)$$

and applying Lemma 4.6 with $\rho = 2\sqrt{2} + t$. $\qquad\qquad\square$

Theorem 4.8 shows that the eigenvector matrix of a $p \times p$ complex Gaussian matrix has a condition number $\mathcal{O}(p^{3/2})$ with high probability, implying that it is extremely unlikely to encounter very large condition numbers. Combined with (17), this *indicates* that the condition numbers $\|\widetilde{y}_i\|_2$ of the perturbed eigenvalues are likely not much larger than the condition number $\|Y_1\|_2$ of the unperturbed semisimple eigenvalue. This expectation is matched by the numerical experiments reported in Section 5.

5 Numerical experiments for synthetic data

In this section, we report numerical experiments using synthetic data that highlight various theoretical aspects of this work. For this purpose, we generate pairs of commuting matrices A_1 and A_2 with known joint eigenvalues by performing a similarity transformation of diagonal matrices. The transformation is effected by a random $n \times n$ matrix X with normalized columns and prescribed condition number $\kappa_2(X) = \kappa$. In this regard, we generate $\widetilde{X} = Q_1 D Q_2$, where $D = \text{diag}(1, \widetilde{\kappa}^{1/(n-1)}, \widetilde{\kappa}^{2/(n-1)}, \ldots, \widetilde{\kappa})$ and Q_1, Q_2 are random orthogonal $n \times n$ matrices. The matrix X is obtained from $\widetilde{X}$ by normalizing its columns, and $\widetilde{\kappa}$ is adjusted such that $\kappa_2(X) = \kappa$. We generate nearly commuting matrices $\widetilde{A}_i = A_i + \epsilon\sqrt{2}/2 E_i$ for $i = 1, 2$ by adding Gaussian random matrices E_i, scaled such that $\|E_i\|_F = 1$. To compute eigenvalue approximations, we apply Algorithm 1 repeatedly with $N = 10^4$ realizations of the random vector μ. All numerical experiments were carried out in Matlab 2023a. The code and data for numerical examples in this paper are available at https://github.com/borplestenjak/RandomJointEig.

Example 5.1 We take $A_i = X D_i X^{-1}$ with $D_1 = \text{diag}(1, 1, 1, 2, 2, 2, 3)$, $D_2 = \text{diag}(1, 2, 3, 1, 2, 3, 3)$, and $\kappa_2(X) = 10^2$. Both matrices A_1 and A_2 have multiple eigenvalues, but the joint eigenvalues are all simple. The condition number of the joint eigenvalue $\boldsymbol{\lambda}^{(1)} = (1, 1)$ is 21.6, $\|A_1\|_2 = 44.8$ and $\|A_2\|_2 = 38.6$, representing a well-conditioned problem. We generated nearly commuting pairs $\widetilde{A}_1, \widetilde{A}_2$ with noise levels $\epsilon = 0, 10^{-14}, 10^{-12}, 10^{-10}$. Even for $\epsilon = 0$, the matrices are inevitably affected by roundoff error on the level $\epsilon_0 = (\|A_1\|_2^2 + \|A_2\|_2^2)^{1/2} u \approx 6.6 \cdot 10^{-15}$, where u is the unit roundoff in double precision.

Figure 1 shows the distribution of the approximation errors obtained when applying Algorithm 1 $N = 10^4$ times to $\widetilde{A}_1, \widetilde{A}_2$. As $\boldsymbol{\lambda}^{(1)}$ is simple, Theorem 4.2 indicates that the distribution of the errors for the two-sided Rayleigh quotient is flat because the first-order term does not depend on μ. While this is indeed true for $\epsilon = 10^{-12}$ and $\epsilon = 10^{-10}$, the effects of roundoff error, committed during the computation
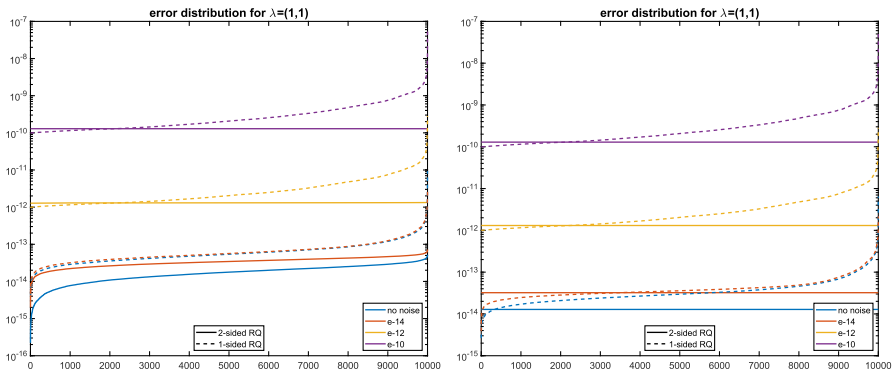
Fig. 1 Distribution of absolute errors for the eigenvalue $\lambda^{(1)}$ from Example 5.1 using double precision (left) and extended precision (right) to compute the eigenvalues of $\widetilde{A}(\mu)$ using one-sided (dashed lines) and two-sided (solid lines) Rayleigh quotients

of the Rayleigh quotients, dominate for $\epsilon = 10^{-14}$ and $\epsilon = 0$, leading to non-flat distributions. It can be seen from the right plot of Fig. 1 that these effects disappear when the computations are carried out in quadruple precision, using the Advanpix Multiprecision Computing Toolbox [1], before rounding the results back to double precision.

Figure 2 compares the approximation errors with the first-order terms of the bounds (14) and (15) from Theorem 4.2. It can be observed that the bounds match the true errors within 1-2 orders of magnitude.

Additional statistics are reported in Table 1. As predicted by Theorem 4.2, the median error is proportional to ϵ. The last two columns show that the error of the two-sided Rayleigh quotient is often better and nearly never more than five times worse than the error of the one-sided Rayleigh quotient.

We have also computed eigenvalue approximation errors for the other joint eigenvalues and the observations are similar to the ones reported for $\lambda^{(1)} = (1, 1)$.
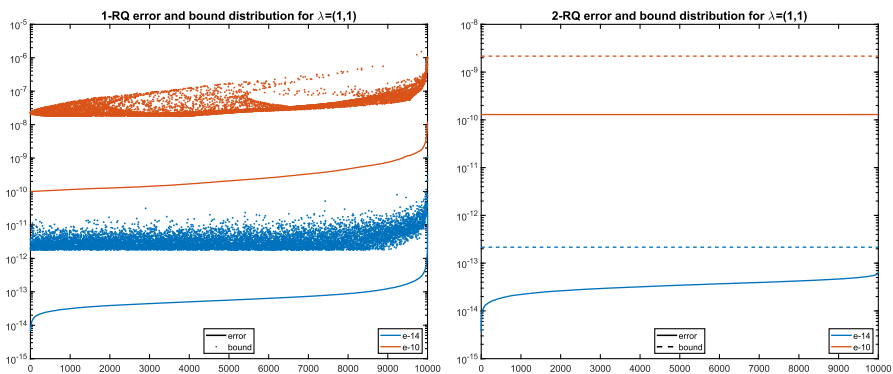
Fig. 2 Absolute errors for the eigenvalue $\lambda^{(1)}$ from Example 5.1 using one-sided Rayleigh quotient vs. bound (15) (left) and two-sided Rayleigh quotient vs. bound (14) (right)

Table 1 Median approximation errors, bounds (15) and (14), and the empirical probability that b (error of two-sided Rayleigh quotient) is (significantly) smaller than a (error of one-sided Rayleigh quotient) for the eigenvalue $\boldsymbol{\lambda}^{(1)}$ from Example 5.1

noise ϵ	1RQ: $\boldsymbol{\lambda}^{(1)} = (1, 1)$		2RQ: $\boldsymbol{\lambda}^{(1)} = (1, 1)$		empirical prob.	
	error (a)	bound (15)	error (b)	bound (14)	$\mathbb{P}(b < a)$	$\mathbb{P}(b < 5a)$
0	$5.4 \cdot 10^{-14}$	$3.5 \cdot 10^{-13}$	$1.8 \cdot 10^{-14}$	$1.4 \cdot 10^{-13}$	0.9609	1.0000
10^{-14}	$5.6 \cdot 10^{-14}$	$3.3 \cdot 10^{-12}$	$3.4 \cdot 10^{-14}$	$2.2 \cdot 10^{-13}$	0.8433	0.9995
10^{-12}	$2.0 \cdot 10^{-12}$	$2.7 \cdot 10^{-10}$	$1.3 \cdot 10^{-12}$	$2.2 \cdot 10^{-11}$	0.7777	1.0000
10^{-10}	$2.1 \cdot 10^{-10}$	$2.3 \cdot 10^{-8}$	$1.3 \cdot 10^{-10}$	$2.2 \cdot 10^{-9}$	0.7865	1.0000

Example 5.2 Using the same matrices from the previous example we now test the probabilistic bounds from Section 4.2. For the eigenvalue $\boldsymbol{\lambda}^{(1)} = (1, 1)$, it follows from Lemma 4.4 that

$$\text{Prob}(d(\mu) < 1/R) < 6/R^2. \tag{19}$$

We computed $d(\mu) := \min_{i=2,\dots,7} \left(\frac{|\lambda_i(\mu) - \lambda_1(\mu)|}{\|\boldsymbol{\lambda}^{(i)} - \lambda_1\|} \right)$ for $N = 10^6$ realizations of μ and computed the empirical probabilities for $\text{Prob}(d(\mu) < 1/R)$ for various values of R. The left plot of Fig. 3 reveals that the bound (19) is quite tight.

We also computed the empirical probabilities for the result of Theorem 4.5 concerning the one-sided Rayleigh quotient approximation error. For this purpose, we added noise of the levels $\epsilon \in \{10^{-12}, 10^{-11}, 10^{-10}, 10^{-9}\}$ to create nearly commuting pairs $\widetilde{A}_1, \widetilde{A}_2$. In line with (16), we compare the empirically measured probabilities with

$$\text{Prob}(\|\widetilde{\boldsymbol{\lambda}}_{\text{RQ1}}^{(i)} - \boldsymbol{\lambda}^{(1)}\|_2 > (1 + R)\epsilon) < 12\kappa_2(X)/R^2. \tag{20}$$

The results are reported in Fig. 3 (right). This time, the empirical probabilities remain proportional to R^{-2}, but the bound is not as tight as on the left-hand side. As expected
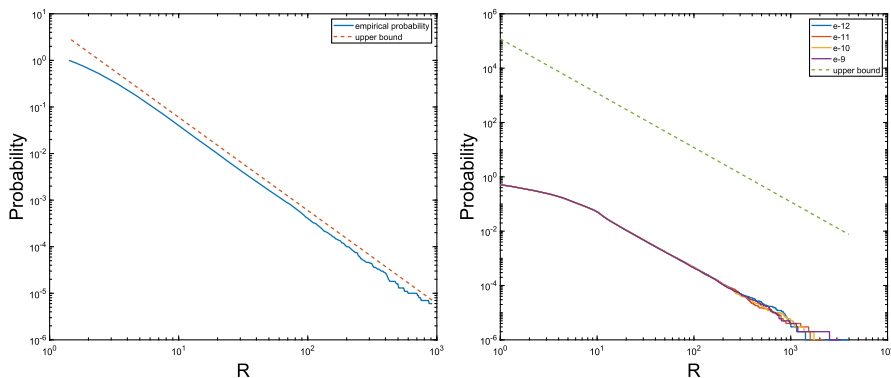

Fig. 3 Empirical probabilities compared to the bounds for (19) (left) and (20) (right) for the data from Example 5.2

from Theorem 4.5, the level of noise only has a negligible impact on the failure probability of the bound.

Example 5.3 We take $A_i = X D_i X^{-1}$ with D_i as in Example 5.1, but now $X = P \cdot \mathrm{diag}(I_2, Z)$, where P is a 7×7 Gaussian random matrix and Z is a 5×5 random matrix with $\kappa_2(Z) = 10^4$. While the joint eigenvalues remain the same, the different choice of X implies that the condition numbers of $\boldsymbol{\lambda}^{(1)} = (1, 1)$ and $\boldsymbol{\lambda}^{(2)} = (1, 2)$ are much smaller than those of the other five eigenvalues. In particular, the condition numbers of $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}^{(4)}$ are 2.1 and $4.1 \cdot 10^3$, respectively. We have $\|A_1\|_2 = 2.6 \cdot 10^3$, $\|A_2\|_2 = 7.1 \cdot 10^3$ and, hence, the noise created by roundoff error is on the level $\epsilon_0 = 8.4 \cdot 10^{-13}$. We repeated the experiments described from Example 5.1 with noise levels $\epsilon = 0, 10^{-12}, 10^{-10}, 10^{-8}$; see Fig. 4. For $\boldsymbol{\lambda}^{(4)}$ the one-sided and two-sided Rayleigh quotients return comparable accuracy, in line with what has been observed in Example 5.1. On contrast, the two-sided Rayleigh quotient is significantly more accurate for $\boldsymbol{\lambda}^{(1)}$. This reflects the fact that the error bound from Theorem 4.2 for the two-sided Rayleigh quotient features $\kappa(\boldsymbol{\lambda}^{(1)}) = 2.1$, while the error bound for the one-sided Rayleigh quotient features $\|X_2\|_2 \|Y_2\|_2 \approx 10^4$. Additional statistics, demonstrating the advantages of the two-sided version, is reported in Table 2.

Example 5.4 The purpose of this example is to investigate the behavior in the presence of (nearly) multiple semisimple eigenvalues. We use $A_i = X D_i X^{-1}$ with $D_1 = \mathrm{diag}(1, 1 + \delta, 1 - \delta, 2, 2, 2, 3)$, $D_2 = \mathrm{diag}(1, 1 - \delta, 1 + \delta, 1, 2, 3, 3)$ for $\delta \geq 0$, and a random X such that $\kappa_2(X) = 10^4$. For every combination of $\delta = 10^{-4}, 10^{-6}, \ldots, 10^{-14}$ and noise level $\epsilon = 10^{-4}, 10^{-6}, \ldots, 10^{-14}$, we compute the approximation errors for the eigenvalue $\boldsymbol{\lambda}^{(1)} = (1, 1)$, which is not well-isolated when δ is small. The results in Table 3 can be roughly clustered into the following three groups depending on (ϵ, δ):

a) For $\epsilon \leq 10^{-4}\delta$, because of $\kappa_2(X) = 10^4$, one can consider $\boldsymbol{\lambda}^{(1)}$ a simple eigenvalue and apply Theorem 4.2 with $p = 1$. Indeed, the errors in this setting behave similarly to the ones observed in Examples 5.1 and 5.3.
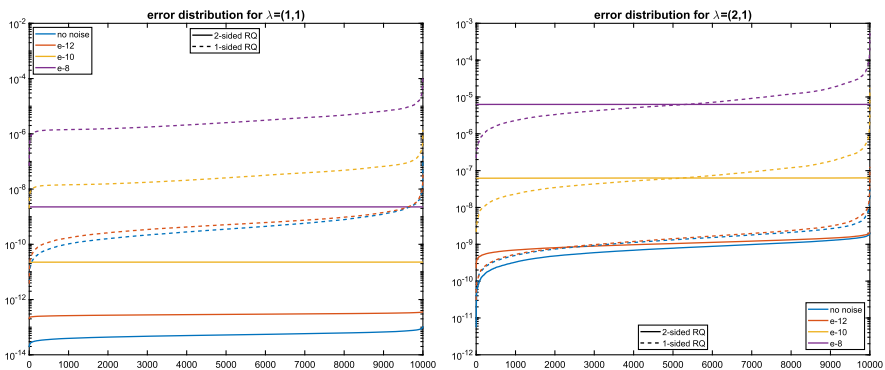

Fig. 4 Distribution of absolute errors for the eigenvalues $\boldsymbol{\lambda}^{(1)} = (1, 1)$ (left) and $\boldsymbol{\lambda}^{(4)} = (2, 1)$ (right) from Example 5.3 using one-sided (dashed lines) and two-sided (solid lines) Rayleigh quotients

Table 2 Median approximation errors, bounds (15) and (14), and the empirical probability that b (error of two-sided Rayleigh quotient) is (significantly) smaller than a (error of one-sided Rayleigh quotient) for the eigenvalues $\lambda^{(1)}$ and $\lambda^{(4)}$ from Example 5.3

	1RQ: $\lambda^{(1)} = (1, 1)$		2RQ: $\lambda^{(1)} = (1, 1)$		empirical prob.	
noise ϵ	error (a)	bound (15)	error (b)	bound (14)	$\mathbb{P}(b < a)$	$\mathbb{P}(b < 5a)$
0	$3.5 \cdot 10^{-10}$	$5.2 \cdot 10^{-8}$	$5.2 \cdot 10^{-14}$	$1.8 \cdot 10^{-12}$	1.0000	1.0000
10^{-12}	$5.1 \cdot 10^{-10}$	$9.6 \cdot 10^{-8}$	$2.9 \cdot 10^{-13}$	$2.1 \cdot 10^{-12}$	1.0000	1.0000
10^{-10}	$2.5 \cdot 10^{-8}$	$3.4 \cdot 10^{-6}$	$2.3 \cdot 10^{-11}$	$2.1 \cdot 10^{-10}$	1.0000	1.0000
10^{-8}	$2.5 \cdot 10^{-6}$	$5.4 \cdot 10^{-4}$	$2.3 \cdot 10^{-9}$	$2.1 \cdot 10^{-8}$	1.0000	1.0000
	1RQ: $\lambda^{(4)} = (2, 1)$		2RQ: $\lambda^{(4)} = (2, 1)$		empirical prob.	
noise ϵ	error (a)	bound (15)	error (b)	bound (14)	$\mathbb{P}(b < a)$	$\mathbb{P}(b < 5a)$
0	$1.3 \cdot 10^{-9}$	$1.4 \cdot 10^{-8}$	$7.8 \cdot 10^{-10}$	$3.5 \cdot 10^{-9}$	0.7519	0.9886
10^{-12}	$1.4 \cdot 10^{-9}$	$1.6 \cdot 10^{-8}$	$1.0 \cdot 10^{-9}$	$4.2 \cdot 10^{-9}$	0.6580	0.9831
10^{-10}	$6.1 \cdot 10^{-8}$	$1.6 \cdot 10^{-6}$	$6.3 \cdot 10^{-8}$	$4.2 \cdot 10^{-7}$	0.4784	0.9715
10^{-8}	$6.0 \cdot 10^{-6}$	$6.1 \cdot 10^{-4}$	$6.3 \cdot 10^{-6}$	$4.2 \cdot 10^{-5}$	0.4671	0.9709

b) For $\epsilon = 10^{-2}\delta$ and $\epsilon = \delta$, $\lambda^{(1)}$ is in a cluster of three joint eigenvalues. In this case, both ϵ and δ contribute to the error.

c) For $\epsilon \geq 10^2\delta$, the perturbation error dominates and one can consider $\lambda^{(1)}$ as a semisimple eigenvalue and apply Theorem 4.2 with $p = 3$.

As δ decreases, the empirical probability for $\mathbb{P}(b < a)$ is seen to decrease and is close to 0.5 for pairs from group c). Still, also for these pairs $\mathbb{P}(b < 5a)$ remains very close to 1. In turn, the two-sided Rayleigh quotient approximation remains competitive also for tight clusters of semisimple eigenvalues.

Figure 5 shows the error distribution of the errors. For $\delta = 10^{-6}$, the curves look similar to the ones before for well-separated simple eigenvalues when $\epsilon \leq 10^{-10}$. The curves for $\epsilon = 10^{-8}$ and $\epsilon = 10^{-6}$ are shaped differently because ϵ is too large and $\lambda^{(1)}$ does not behave anymore as a simple eigenvalue. For $\delta = 10^{-12}$, all curves have this shape. We checked that this behaviour continues down to $\delta = 0$. In summary, it can be observed that multiple semisimple eigenvalues are computed with roughly the same accuracy as simple eigenvalues.

Example 5.5 We now consider $A_i = X B_i X^{-1}$ with random X, such that $\kappa_2(X) = 10$, and

$$B_1 = \begin{bmatrix} 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & 1 & & & \\ & & & 2 & & \\ & & & & 3 & \\ & & & & & 4 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1 & 1 & & & & \\ & 1 & 1 & & & \\ & & 1 & & & \\ & & & 4 & & \\ & & & & 3 & \\ & & & & & 2 \end{bmatrix}.$$

In exact arithmetic, the eigenvalue $\lambda^{(1)} = (1, 1)$ is multiple and *not* semisimple. In the presence of roundoff error, this multiplicity is broken but we can expect the condition

Table 3 Statistics of approximation errors for eigenvalue λ_1 from Example 5.4

	$\delta = 10^{-4}$	$\delta = 10^{-6}$	$\delta = 10^{-8}$	$\delta = 10^{-10}$	$\delta = 10^{-12}$	$\delta = 10^{-14}$
noise ϵ	Median of a (error of one-sided Rayleigh quotient)					
10^{-14}	$3.1 \cdot 10^{-10}$	$3.4 \cdot 10^{-10}$	$3.4 \cdot 10^{-10}$	$2.3 \cdot 10^{-10}$	$7.5 \cdot 10^{-12}$	$2.0 \cdot 10^{-12}$
10^{-12}	$3.9 \cdot 10^{-10}$	$3.2 \cdot 10^{-10}$	$4.6 \cdot 10^{-10}$	$2.4 \cdot 10^{-10}$	$7.2 \cdot 10^{-12}$	$7.5 \cdot 10^{-12}$
10^{-10}	$2.9 \cdot 10^{-8}$	$2.8 \cdot 10^{-8}$	$2.5 \cdot 10^{-8}$	$5.4 \cdot 10^{-10}$	$5.8 \cdot 10^{-10}$	$5.8 \cdot 10^{-10}$
10^{-8}	$2.8 \cdot 10^{-6}$	$2.6 \cdot 10^{-6}$	$5.3 \cdot 10^{-8}$	$5.7 \cdot 10^{-8}$	$5.7 \cdot 10^{-8}$	$5.7 \cdot 10^{-8}$
10^{-6}	$2.5 \cdot 10^{-5}$	$5.4 \cdot 10^{-6}$	$5.8 \cdot 10^{-6}$	$5.8 \cdot 10^{-6}$	$5.8 \cdot 10^{-6}$	$5.8 \cdot 10^{-6}$
noise ϵ	Median of b (error of two-sided Rayleigh quotient)					
10^{-14}	$7.7 \cdot 10^{-11}$	$6.4 \cdot 10^{-11}$	$1.1 \cdot 10^{-10}$	$5.7 \cdot 10^{-11}$	$4.4 \cdot 10^{-12}$	$1.1 \cdot 10^{-12}$
10^{-12}	$1.1 \cdot 10^{-10}$	$7.5 \cdot 10^{-11}$	$1.5 \cdot 10^{-10}$	$6.7 \cdot 10^{-11}$	$4.7 \cdot 10^{-12}$	$6.0 \cdot 10^{-12}$
10^{-10}	$7.2 \cdot 10^{-9}$	$7.2 \cdot 10^{-9}$	$9.5 \cdot 10^{-9}$	$5.7 \cdot 10^{-10}$	$5.7 \cdot 10^{-10}$	$5.6 \cdot 10^{-10}$
10^{-8}	$7.2 \cdot 10^{-7}$	$9.5 \cdot 10^{-7}$	$5.6 \cdot 10^{-8}$	$5.5 \cdot 10^{-8}$	$5.5 \cdot 10^{-8}$	$5.5 \cdot 10^{-8}$
10^{-6}	$9.5 \cdot 10^{-5}$	$5.6 \cdot 10^{-6}$	$5.5 \cdot 10^{-6}$	$5.5 \cdot 10^{-6}$	$5.5 \cdot 10^{-6}$	$5.5 \cdot 10^{-6}$
noise ϵ	empirical probability $\mathbb{P}(b < a)$					
10^{-14}	0.9495	0.9712	0.9334	0.9524	0.8039	0.7068
10^{-12}	0.9474	0.9506	0.8922	0.9403	0.7069	0.5855
10^{-10}	0.9736	0.9735	0.8373	0.4561	0.4817	0.4952
10^{-8}	0.9699	0.8446	0.4436	0.4902	0.4902	0.4903
10^{-6}	0.8413	0.4558	0.4997	0.5005	0.5005	0.5005
noise ϵ	empirical probability $\mathbb{P}(b < 5a)$					
10^{-14}	0.9998	0.9999	0.9994	0.9987	0.9973	0.9887
10^{-12}	1.9996	0.9994	0.9986	0.9991	0.9945	0.9718
10^{-10}	1.0000	1.0000	0.9999	0.9916	0.9767	0.9767
10^{-8}	1.0000	1.0000	0.9944	0.9798	0.9791	0.9793
10^{-6}	1.0000	0.9946	0.9829	0.9821	0.9820	0.9820
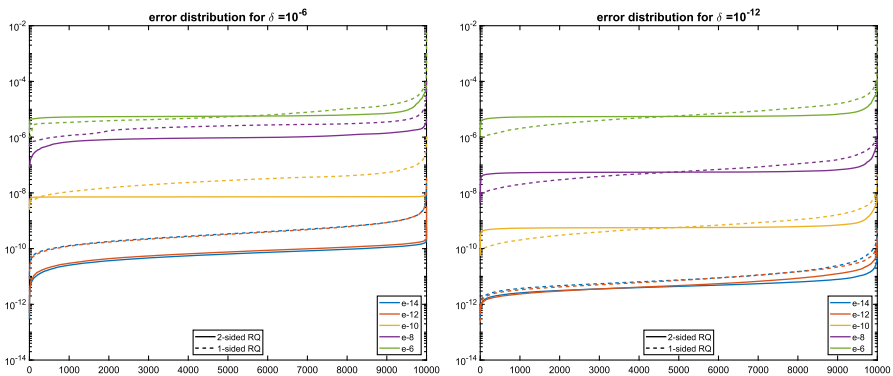
Fig. 5 Distribution of absolute errors for eigenvalue $\lambda^{(1)}$ from Example 5.4 for $\delta = 10^{-6}$ (left) and $\delta = 10^{-12}$ (right), using one-sided (dashed lines) and two-sided (solid lines) Rayleigh quotients

numbers of the eigenvalue $\boldsymbol{\lambda}^{(1)}$ and the eigenvector matrix to become very high. In particular, the results of Theorem 4.2 do not apply in this situation.

As $\widetilde{A}(\mu)$ is diagonalizable, one can still attempt to apply Algorithm 1. Figure 6 shows that the error for the triple eigenvalue $\boldsymbol{\lambda}^{(1)}$ is on the level of $\epsilon^{1/3}$, the same level of error that can be expected when perturbing an eigenvalue in a Jordan block of size three in a single matrix. On the other hand, the error for $\boldsymbol{\lambda}^{(4)} = (2, 4)$ behaves in the same way as the errors for simple eigenvalues in the other examples.

6 Multiparameter eigenvalue problems

This section is concerned with one of the applications that motivated this work. A *d-parameter eigenvalue problem* has the form

$$A_{i0}x_i = \lambda_1 A_{i1}x_i + \cdots + \lambda_d A_{id}x_i, \quad i = 1, \ldots, d, \tag{21}$$

where A_{ij} is an $n_i \times n_i$ complex matrix and $x_i \neq 0$ for $i = 1, \ldots, d$. When (21) is satisfied, a d-tuple $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d) \in \mathbb{C}^d$ is called an *eigenvalue* and $x_1 \otimes \cdots \otimes x_d$ is a corresponding *eigenvector*. Generically, a multiparameter eigenvalue problem (21) has $N = n_1 \cdots n_d$ eigenvalues that are roots of a system of d multivariate characteristic polynomials $p_i(\lambda_1, \ldots, \lambda_d) := \det(A_{i0} - \lambda_1 A_{i1} - \cdots - \lambda_d A_{id}) = 0$ for $i = 1, \ldots, d$.

The problem (21) is closely related to a system of d generalized eigenvalue problems

$$\Delta_i z = \lambda_i \Delta_0 z, \quad i = 1, \ldots, d, \tag{22}$$

with $z = x_1 \otimes \cdots \otimes x_d$ and the $N \times N$ matrices

$$\Delta_0 = \begin{vmatrix} A_{11} & \cdots & A_{1d} \\ \vdots & & \vdots \\ A_{d1} & \cdots & A_{dd} \end{vmatrix}_\otimes = \sum_{\sigma \in S_d} \text{sgn}(\sigma)\, A_{1\sigma_1} \otimes A_{2\sigma_2} \otimes \cdots \otimes A_{d\sigma_d},$$

$$\Delta_i = \begin{vmatrix} A_{11} & \cdots & A_{1,i-1} & A_{10} & A_{1,i+1} & \cdots & A_{1d} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ A_{d1} & \cdots & A_{d,i-1} & A_{d0} & A_{d,i+1} & \cdots & A_{dd} \end{vmatrix}_\otimes, \quad i = 1, \ldots, d,$$

which are called *operator determinants* [3]. In the following we will assume that Δ_0 is invertible, in which case (21) is called *regular* and the matrices $\Gamma_i := \Delta_0^{-1}\Delta_i$ for $i = 1, \ldots, d$ commute. If N is not too large, then a standard approach to solve (21) is to explicitly compute the matrices $\Gamma_1, \ldots, \Gamma_d$ and then solve the joint eigenvalue problem, see, e.g., [19, 39]. In the following, we will discuss randomization-based methods.

The right-definite case If all matrices A_{ij} defining (21) are real and symmetric, and Δ_0 is positive definite, the problem (21) is called *right-definite*. This allows us to perform the Cholesky decomposition $\Delta_0 = VV^T$ and transform (22) into the joint eigenvalue problem $D_i w = \lambda_i w$, $i = 1, \ldots, d$, where $D_i = V^{-1}\Delta_i V^{-T}$ is
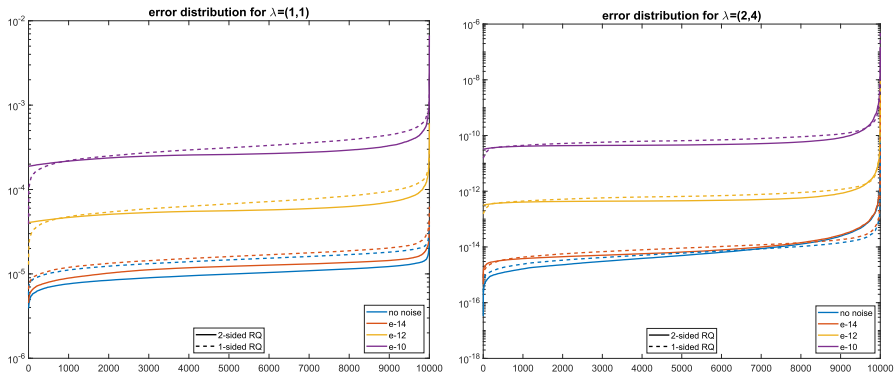
Fig. 6 Distribution of absolute errors for eigenvalues $\lambda^{(1)}$ (left) and $\lambda^{(4)}$ (right) from Example 5.5, using one- sided (dashed lines) and two-sided (solid lines) Rayleigh quotients

symmetric and $w = V^T z$. In exact arithmetic, the matrices $D_1, \ldots, D_d$ commute and can be simultaneously diagonalized by an orthogonal matrix Q; see also [39]. In turn, the right-definite case allows one to apply the randomized joint diagonalization method from [18], which diagonalizes a random linear combination $D(\mu)$ to produce the orthogonal transformation Q. In [18], this approach has been shown to be robust to noise. When Δ_0 is ill-conditioned, the recent randomized approach for simultaneous diagonalization by congruence in [17] can be more reliable.

The non-definite regular case In many applications, Δ_0 is invertible but not symmetric definite. As the matrices $\Gamma_i = \Delta_0^{-1} \Delta_i$ commute, they admit a simultaneous Schur factorization, that is, they can be triangularized simultaneously by the same unitary similarity transformation. The matrix pencils $(\Delta_1, \Delta_0), \ldots, (\Delta_d, \Delta_0)$ admit a simultaneous generalized Schur factorization, that is, there are unitary matrices Q and Z such that every $Q\Delta_i Z$ is upper triangular for $i = 0, \ldots, d$. The approach in [19, Algorithm 2.3] for two-parameter eigenvalue problems first applies the QZ algorithm to the pencil (Δ_1, Δ_0), reorders the generalized Schur form so that (nearly) multiple eigenvalues are grouped together, and then applies Q and Z to Δ_2. This yields block upper triangular matrices and the eigenvalues are extracted from the diagonal blocks of $Q\Delta_0 Z, Q\Delta_1 Z, Q\Delta_2 Z$. A problem with this approach is that the correct reordering of the generalized Schur form is difficult to ensure in the presence of noise. Also, this procedure has to be repeatedly applied for problems with more than two parameters.

An alternative to Schur forms is to use eigenvectors. Let us take a random linear combination $\Delta(\mu) = \sum_{i=1}^{d} \mu_i \Delta_i$. Then we can compute the left and right eigenvector matrices either of $\Delta_0^{-1} \Delta(\mu)$ or of the pencil $(\Delta(\mu), \Delta_0)$, and apply the resulting transformation to the other generalized eigenvalue problems from (22). The use of pencils avoids the multiplication with Δ_0^{-1} and corresponds to the use of generalized Rayleigh quotients. Specifically, the *generalized one-sided Rayleigh quotients* take the form

$$\rho(x_i, \Delta_k, \Delta_0) = \frac{x_i^* \Delta_k x_i}{x_i^* \Delta_0 x_i}, \quad k = 1, \ldots, d, \ i = 1, \ldots, N,$$

while the *generalized two-sided Rayleigh quotients* are given by

$$\rho(x_i, y_i, \Delta_k, \Delta_0) = \frac{y_i^* \Delta_k x_i}{y_i^* \Delta_0 x_i}, \quad k = 1, \ldots, d, \ i = 1, \ldots, N.$$

The singular case It can happen that every linear combination of the matrices $\Delta_0, \Delta_1, \ldots, \Delta_d$ is singular, but (21) still has finitely many solutions that are roots of d multivariate characteristic polynomials. Such problems are called *singular multiparameter eigenvalue problems*, resulting in a system of d singular matrix pencils (22). The classical approach to deal with a singular pencil proceeds by extracting the regular part by the staircase algorithm [45] and using the QZ algorithm. Recently, randomized methods for singular pencils have been proposed that do not require the use of the staircase method [21, 22, 29]. Unfortunately, none of these methods has been so far generalized to a family of singular pencils (22). The only numerical method available is a generalized staircase-type algorithm [34]. The regular problem extracted by this method can be addressed with the algorithm proposed in this work.

6.1 Solution of multiparameter eigenvalue problems with Algorithm 1

We suggest to solve a multiparameter eigenvalue problem in the following way. First, compute eigenvalues and, more importantly, eigenvectors of $\Gamma(\mu) = \Delta_0^{-1} \Delta(\mu)$, where $\Delta(\mu)$ is a random linear combination of matrices $\Delta_1, \ldots, \Delta_d$. For each right and left eigenvector z and w we can then compute the corresponding eigenvalue $(\lambda_1, \ldots, \lambda_d)$ from two-sided Rayleigh quotients because z and w are common eigenvectors for all matrices $\Gamma_1, \ldots, \Gamma_d$.

We note that for the above we do not have to compute all operator determinants $\Delta_0, \ldots, \Delta_d$ explicitly. This is important because these matrices can be very large and consume a lot of memory. We explicitly need just Δ_0 and $\Delta(\mu)$, which we can compute from the matrices in (21) and values $\mu_1, \ldots, \mu_d$. Once we have an eigenvector z we can think of it as a vectorization of a d-dimensional tensor $\mathcal{Z} \in \mathbb{C}^{n_1 \times \cdots \times n_d}$, i.e., $z = \text{vec}(\mathcal{Z})$. Multiplication by Δ_0 (and similarly with $\Delta_1, \ldots, \Delta_d$) in the computation of Rayleigh quotients can be then efficiently performed as $\Delta_0 z = \text{vec}(\mathcal{W})$, where

$$\mathcal{W} = \sum_{\sigma \in S_d} \text{sgn}(\sigma) \, A_{1\sigma_1} \times_1 A_{2\sigma_2} \times_2 \cdots A_{d\sigma_d} \times_d \mathcal{Z}$$

and $\times_1, \ldots, \times_d$ are multiplications of tensor by matrices in directions $1, \ldots, d$, for details see, e.g., [26].

6.2 Numerical examples

Example 6.1 Recently [10], it was exposed that the numerical solver for multiparameter eigenvalue problems in [35] sometimes fails to find all eigenvalues of random three-parameter eigenvalue problems. The observed failures can be attributed to the method for computing joint eigenvalues in [35], which is based on the generalized Schur form followed by a simple clustering. In the new release [36] this is replaced
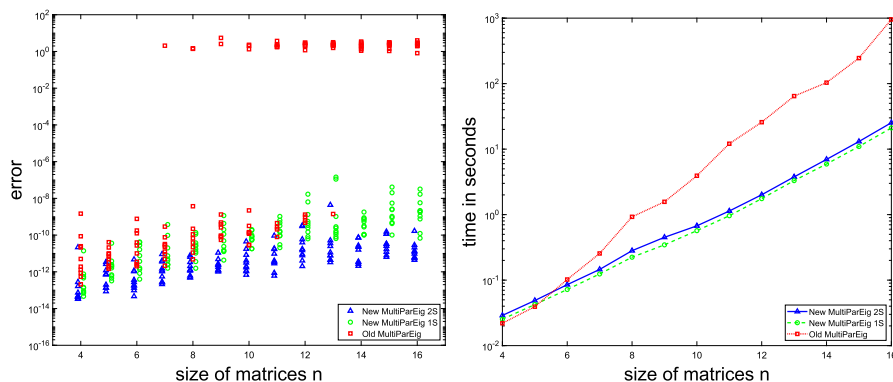
Fig. 7 Comparison of old and new solver for three-parameter eigenvalue problems from [35] and [36] on problems from Example 6.1. The left figure presents the maximal error of computed eigenpairs, while the right figure compares computational times

by the method based on Algorithm 1 from this paper. The new solver is much faster and computes eigenvalues of the problems reported in [10] more accurately.

To demonstrate this, we constructed three-parameter eigenvalue problems in a similar way as in [10]. We take $n \times n$ matrices

$$A_{ij} = Q_{ij} D_{ij} Q_{ij}^T + \delta_{ij} I_n, \quad i = 1, 2, 3, \; j = 0, 1, 2, 3, \tag{23}$$

where Q_{ij} is a random orthogonal matrix (constructed from the QR decomposition of a Gaussian random matrix), D_{ij} is a diagonal matrix whose diagonal elements are random values uniformly distributed in $[-\frac{1}{2n}, \frac{1}{2n}]$, and δ_{ij} is the Kronecker delta. For each size $n = 4, \ldots, 16$ we generated 10 different problems of form (23) and for each computed the largest error of obtained eigenpairs. We compared `threepareig`, which is a Matlab solver for three-parameter eigenvalue problems in MultiParEig, from [35] and the updated version [36]. The results are presented in Fig. 7. If we compare the two-sided (2S) and one-sided (1S) versions, we see that the two-sided version is slightly more expensive but much more accurate.

Example 6.2 Wave propagation in an elastic layer that is in contact with another medium of infinite extent at one of the surfaces can be modelled by a nonlinear eigenvalue problem of the form

$$(-k^2 E_0 + (ik)E_1 - E_2 + \omega^2 M + \sqrt{\kappa_1^2 - k^2} R_1 + \sqrt{\kappa_2^2 - k^2} R_2)u = 0, \tag{24}$$

where $E_0, E_1, E_2, M, R_1, R_2$ are $n \times n$ matrices, whose solutions (ω, k) give dispersion curves, where ω is the frequency and k is the wavenumber.

Recently [15], it was shown that for a given ω we can get the corresponding solutions k of (24) by solving a four-parameter eigenvalue problem

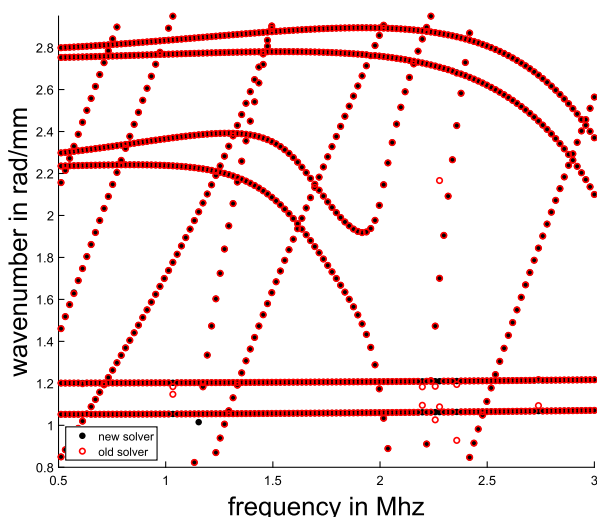$$\left(-E_2 + \omega^2 M + ik E_1 + i\eta_1 R_1 + i\eta_2 R_2 + \xi E_0 \right) u = 0$$

Fig. 8 Points on dispersion curves related to the problem of wave propagation of a brass plate coupled to infinite Teflon from [15], computed by the old and updated solver for multiparameter eigenvalue problems

$$\left(\begin{bmatrix} 0 & -\eta_1^2 \\ 1 & 0 \end{bmatrix} + i\eta_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \xi \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}\right) x_1 = 0$$

$$\left(\begin{bmatrix} 0 & -\eta_2^2 \\ 1 & 0 \end{bmatrix} + i\eta_2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \xi \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix}\right) x_2 = 0$$

$$\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + ik \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \xi \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) x_3 = 0,$$

where $\eta_1^2 = \kappa_1^2 - k^2$, $\eta_2^2 = \kappa_2^2 - k^2$, $\xi = -k^2$, and $x_1, x_2, x_3 \neq 0$. This leads to Δ-matrices of size $8n \times 8n$ such that $\Delta_0^{-1}\Delta_1, \ldots, \Delta_0^{-1}\Delta_4$ commute, for more details, see [15].

For a particular problem of a brass plate coupled to infinite Teflon, using the same matrices of size 45×45 in (24) as in [15, Section 6.2], we computed the points on the dispersion curves for 150 frequencies up to 3Mhz using the old solver for multiparameter eigenvalue problems in [35] and the new solver in [36] that uses two-sided Rayleigh quotients from eigenvectors of a random linear combination $\Delta_0^{-1}(\mu_1\Delta_1 + \cdots + \mu_4\Delta_4)$. The comparison is in Fig. 8, where we see that the old algorithm could not compute all the points on the two horizontal dispersion curves accurately.

7 Roots of polynomial systems

Another important source of problems leading to joint eigenvalue problems for a commuting family of matrices are eigenvector-based methods for solving multivariate polynomial equations; see, e.g., [8, 32, 40]. In this section, we provide a brief introduction to such an approach and demonstrate the accuracy of the proposed two-sided

Rayleigh quotient method for joint eigenvalue problems that arise in polynomial root finding.

Let $\mathcal{P}^s$ denote the set of polynomials over $\mathbb{C}$ in s variables $x_1, \ldots, x_s$. Given a system P of n polynomials $p_i \in \mathcal{P}^s$, $i = 1, \ldots, n$, a fundamental task of computer algebra and numerical polynomial algebra is to compute their common roots, or the so-called algebraic variety $V := \{x \in \mathbb{C}^s : p_i(x) = 0, i = 1, \ldots, n\}$ of P. For this purpose, we need to study the vector space $\mathcal{R}[\mathcal{I}] := \mathcal{P}^s / \mathcal{I}$ of the *quotient ring* or the *residue class ring modulo the ideal* $\mathcal{I}$ generated by P, which is defined as:

$$\mathcal{I} = \langle p_1, \ldots, p_n \rangle := \{c_i \, p_i : c_i \in \mathcal{P}^s\}.$$

When the algebraic variety V is 0-dimensional, i.e., V only consists of finitely many points, the vector space $\mathcal{R}[\mathcal{I}]$ has the dimension equal to the number of common roots (counting their multiplicities). This allows us to perform computation of linear algebra if a basis $\mathcal{B}$ of $\mathcal{R}[\mathcal{I}]$ is provided. Let M_g denote the matrix associated with the linear operation of polynomial multiplication with a given polynomial $g \in \mathcal{P}^s$ in the vector space $\mathcal{R}[\mathcal{I}]$ with respect to the basis $\mathcal{B}$. It turns out that the family $\{M_g : g \in \mathcal{P}^s\}$ of all the possible multiplication matrices is a commuting family since polynomial multiplication is commutative [40]. For illustration purposes, we assume all the roots $z_1, \ldots, z_m$ are simple. It can be shown that the joint eigenvectors for this commuting family are the Lagrange polynomials, which evaluate to 1 on exactly one z_i and 0 on the other z_j, $j \neq i$, and that the corresponding eigenvalue for the matrix M_g is $g(z_i)$. Refer to [14, Proposition 6.1] for further details on Lagrange polynomials defined on the common roots of a given polynomial system. Hence, the joint eigenvalues of the commuting family

$$\mathcal{M} := \{M_{x_1}, \ldots, M_{x_s}\}$$

are the common roots [40, Theorem 2.27]. These multiplication matrices can be obtained through symbolic computation of the Gröbner basis [7], direct computation of the normal forms from Macaulay matrices [44], and by leveraging the shift-invariant structure of the numerical basis of the null space of Macaulay matrices [46]. For further details on numerical polynomial algebra, see, e.g., the classical textbook [40].

We note that a similar approach with block Macaulay matrices that also requires joint eigenvalues of commuting matrices, was recently developed for rectangular multiparameter eigenvalue problems, for details see, e.g., [47]. A rectangular multiparameter eigenvalue problem is different from (21), a generic linear form is

$$B_0 x = \lambda_1 B_1 x + \cdots + \lambda_d B_d x, \tag{25}$$

where $B_0, \ldots, B_d$ are rectangular matrices of size $(n + d - 1) \times n$. If $\lambda_1, \ldots, \lambda_d$ satisfy (25) for a nonzero x, then $\lambda = (\lambda_1, \ldots, \lambda_d)$ is an eigenvalue. It is also possible to solve (25) with tools for standard multiparameter eigenvalue problems of the form (21), for details see [20].

Example 7.1 As we have already observed the advantage of the two-sided Rayleigh quotient over the one-sided one in Section 5, we only compare the two-sided Rayleigh quotient with Schur decomposition of one random linear combination (RSchur) used

Table 4 Polynomial system accuracy comparison

	random	rose	katsura7	redeco8
Rschur	1.31e-11	1.98e-08	1.74e-10	3.09e-12
schur	5.58e-12	1.12e-08	8.19e-10	2.75e-12
RQ2	9.72e-14	9.26e-09	1.61e-11	3.52e-12

in [44] and Schur decomposition of the first matrix (`Schur`) used in [46]. Our comparison focuses on the commuting families generated from random polynomial systems of 3 variables with maximum degree 10 (resulting in 3 1000×1000 matrices named `random`)[1], the polynomial system arising in the computation of general economic equilibrium models [38] (resulting in 4 136×136 matrices named `rose`), the polynomial system in magnetism in physics [6] (resulting in 7 128×128 matrices named `katsura7`) and another system in economics [33, p. 148] (resulting in 9 64×64 matrices named `redeco8`).[2] The commuting family in `random` is generated using normal forms from Macaulay matrices [44], while in `rose`, `katsura7` and `redeco8` the null space of Macaulay matrices [46] is used. The error is measured by the residual in terms of the roots of the corresponding polynomial systems, averaged over 100 runs. Note that in the original implementations in [46] and [44], there is an option for eigenvalue clustering. However, we observed significant errors due to clustering for `RSchur`, so we excluded those results from our comparison. We also observed that generating the multiplication matrices is significantly more expensive than computing the joint eigenvalues. Therefore, only a comparison of the accuracy is included.

The results presented in Table 4 illustrate that the two-sided Rayleigh quotient method can achieve superior accuracy for both synthetic polynomial systems and polynomial systems from real-world applications.

Example 7.2 It is well known that we can compute roots of a univariate polynomial as eigenvalues of its companion matrix. In a similar way it is possible to linearize a system of two bivariate polynomials into a two-parameter eigenvalue problem whose eigenvalues agree with the roots, see, e.g., [37]. We start from a system of two bivariate polynomials $p_i(x_1, x_2) = 0$, $i = 1, 2$, and construct matrices A_{ij} of a so-called *determinantal representation* such that $\det(A_{i0} - x_1 A_{i1} - x_2 A_{i2}) = p_i(x_1, x_2)$. The eigenvalues of the obtained two-parameter eigenvalue problem then give the roots of the initial polynomial system.

Recently [14], the stability of this approach and of several other methods from Section 7 was analyzed. Although the main result in [14] is pessimistic and shows that such methods for computing roots of systems of bivariate polynomials are unstable, we can show that by replacing the old solver for multiparameter eigenvalue problems with the new two-sided Rayleigh quotient method we get more accurate results. The results are presented in Fig. 9, where we recreate the example in [14, Figure 1.(e)].

---

[1] We use the implementation in [44], available at https://github.com/simontelen/NormalForms, to get the multiplication matrices.

[2] Examples `rose`, `katsura7` and `redeco8` are documented in PHCpack [48] and matlab codes for these examples are documented in [46], available at https://gitlab.esat.kuleuven.be/Christof.Vermeersch/macaulaylab-public.

Fig. 9 Comparison of accuracy obtained with the old and new solver for two-parameter eigenvalue problems applied to polynomials in [14, Example 1.(e)]
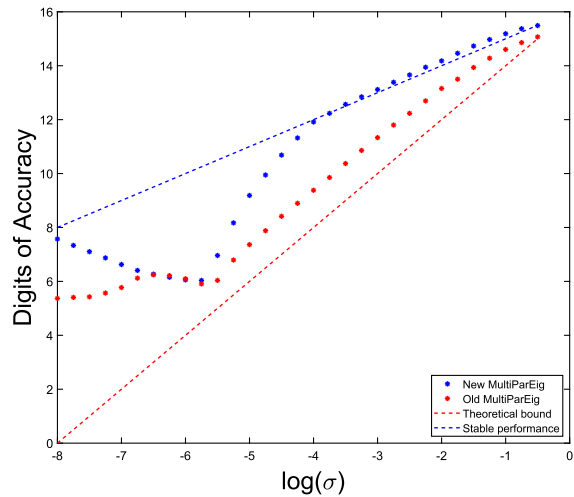
The polynomials in this example have the form

$$p_i(x_1, x_2) = (x_i - \tfrac{1}{3})^2 + \sigma \sum_{j=1}^{2} q_{ij}(x_j - \tfrac{1}{3}), \quad i = 1, 2, \tag{26}$$

where $\sigma > 0$ and q_{ij} are entries of a random 2×2 real orthogonal matrix. The system has a root at $(\tfrac{1}{3}, \tfrac{1}{3})$. Like in [14], we computed roots of 1000 random systems of the form (26) for each value of σ and we report the median error of the root $(\tfrac{1}{3}, \tfrac{1}{3})$. As shown in [14], for small $\sigma > 0$, the condition number of $(\tfrac{1}{3}, \tfrac{1}{3})$ as a root of a polynomial system is $\mathcal{O}(\sigma^{-1})$, while the condition number of $(\tfrac{1}{3}, \tfrac{1}{3})$ as an eigenvalue of the related two-parameter eigenvalue problem is $\mathcal{O}(\sigma^{-2})$. Because of that, while the accuracy of a stable method for computing roots of bivariate polynomials should be around the blue line on Fig. 9, the accuracy of all methods from Section 7 is expected to be much worse and close to the red line. We see that the new solver gives much more accurate results that are even close to the blue line of stability if σ is not too small.

8 Conclusions

We proposed a simple numerical approach to compute joint eigenvalues of a family of (nearly) commuting matrices, combining eigenvectors of a random linear combination with Rayleigh quotients. Our main results show that this approach, in particular the use of two-sided Rayleigh quotients, accurately computes well-conditioned semisimple joint eigenvalues with high probability. It still works satisfactorily in the presence of defective eigenvalues. Numerous numerical examples show that the method can be efficiently used in solvers for multiparameter eigenvalue problems and roots of systems of multivariate polynomials. The method is included in the new release of MultiParEig, a Matlab toolbox for multiparameter eigenvalue problems [36].

A Perturbations of invariant subspaces

This appendix contains the eigenspace perturbation expansions needed for the analysis in Section 4.1. These expansions are direct consequences of existing perturbation results in, e.g., [24] and [42, Section 5].

Lemma A.1 *Consider a matrix $A \in \mathbb{C}^{n \times n}$ transformed to a block diagonal form*

$$Y^* A X = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, \quad A_{11} \in \mathbb{C}^{p \times p}, \ A_{22} \in \mathbb{C}^{(n-p) \times (n-p)},$$

such that $Y^* X = I$ and the spectra of A_{11} and A_{22} are disjoint. Partition $X = [X_1 \ X_2] \in \mathbb{C}^{n \times n}$ and $Y = [Y_1 \ Y_2] \in \mathbb{C}^{n \times n}$ accordingly. Then, given a perturbed matrix $\widetilde{A} = A + \epsilon E$ with $\epsilon > 0$ sufficiently small, there exist bases $\widetilde{X}_1, \widetilde{Y}_1 \in \mathbb{C}^{n \times p}$ for right/left invariant subspaces of $\widetilde{A}$ such that

$$\widetilde{X}_1 = X_1 - \epsilon X_2 \mathbb{T}^{-1}(Y_2^* E X_1) + \mathcal{O}(\epsilon^2),$$
$$\widetilde{Y}_1^* = Y_1^* - \epsilon \mathbb{T}_*^{-1}(X_2^* E^* Y_1)^* Y_2^* + \mathcal{O}(\epsilon^2),$$

where $\mathbb{T} : Z \mapsto A_{22} Z - Z A_{11}$, $\mathbb{T}_* : Z \mapsto A_{22}^* Z - Z A_{11}^*$, and $\widetilde{Y}_1^* \widetilde{X}_1 = I_p + \mathcal{O}(\epsilon^2)$.

Proof Define

$$F := Y^* E X = [Y_1 \ Y_2]^* E [X_1 \ X_2] = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix}.$$

Applying [24, Corollary 2.4] to the matrix $B := Y^* A X$ shows that there is a basis $\begin{bmatrix} I_p \\ Z_E \end{bmatrix}$ for a right invariant subspace of $B + \epsilon F$ such that $Z_E = -\epsilon \mathbb{T}^{-1}(F_{21}) + \mathcal{O}(\epsilon^2)$. Because the matrices $A + \epsilon E$ and $B + \epsilon F$ are similar, it follows that

$$\widetilde{X}_1 = X \begin{bmatrix} I_p \\ Z_E \end{bmatrix} = X_1 - \epsilon X_2 \mathbb{T}^{-1}(Y_2^* E X_1) + \mathcal{O}(\epsilon^2)$$

is a basis for a right invariant subspace of $A + \epsilon E$.

By applying the same argument to A^* we obtain the analogous result for $\widetilde{Y}_1$. Because of $Y_1^* X_2 = 0$ and $Y_2^* X_1 = 0$, the perturbation expansions for $\widetilde{X}_1, \widetilde{Y}_1$ imply that $\widetilde{Y}_1^* \widetilde{X}_1 = I_p + \mathcal{O}(\epsilon^2)$. □

The following result is a special case of Lemma A.1, which coincides with [16, Theorem 2] for $p = 1$.

Lemma A.2 *Under the notation and assumptions of Lemma A.1, suppose that $A_{11} = \lambda I_p$, that is, X_1, Y_1 are bases for right/left eigenspaces of A. Then there exist bases $\widetilde{X}_1, \widetilde{Y}_1 \in \mathbb{C}^{n \times p}$ for right/left invariant subspace of $\widetilde{A} = A + \epsilon E$ such that*

$$\widetilde{X}_1 = X_1 - \epsilon X_2 (A_{22} - \lambda I_{n-p})^{-1} Y_2^* E X_1 + \mathcal{O}\left(\epsilon^2\right), \tag{27}$$

$$\widetilde{Y}_1^* = Y_1^* - \epsilon Y_1^* E X_2 (A_{22} - \lambda I_{n-p})^{-1} Y_2^* + \mathcal{O}\left(\epsilon^2\right), \tag{28}$$

and $\widetilde{Y}_1^* \widetilde{X}_1 = I_p + \mathcal{O}(\epsilon^2)$.

Proof The result follows from Lemma A.1 by noting that the operators $\mathbb{T}$ and $\mathbb{T}_*$ from Lemma A.1 simplify to $A_{22} - \lambda I_{n-p}$ and $A_{22}^* - \overline{\lambda} I_{n-p}$, respectively. □

Code availability The code and data for numerical examples in this paper are available at https://github.com/borplestenjak/RandomJointEig.

Declarations

Conflict of interests Not applicable.

References

1. Advanpix LLC.: Multiprecision Computing Toolbox for Matlab. https://www.advanpix.com, (2024). Version 5.2.9.15553
2. Anderson, E., Bai, Z., Bischof, C., Blackford, L.S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A., Sorensen, D.: LAPACK Users' Guide. Society for Industrial and Applied Mathematics, third edition (1999)
3. Atkinson, F.V.: Multiparameter Eigenvalue Problems. Mathematics in Science and Engineering, Vol. 82. Academic Press, New York-London. Volume I: Matrices and compact operators (1972)
4. Banks, J., Garza-Vargas, J., Srivastava, N.: Global convergence of Hessenberg shifted QR III: Approximate Ritz values via shifted inverse iteration (2022). arXiv:2205.06804
5. Banks, J., Kulkarni, A., Mukherjee, S., Srivastava, N.: Gaussian regularization of the pseudospectrum and Davies' conjecture. Comm. Pure Appl. Math. **74**(10), 2114–2132 (2021)
6. Boege, W., Gebauer, R., Kredel, H.: Some examples for solving systems of algebraic equations by calculating Groebner bases. J. Symbolic Comput. **2**(1), 83–98 (1986)
7. Corless, R.M.: Gröbner bases and matrix eigenproblems. SIGSAM Bull. **30**(4), 26–32 (1996)
8. Corless, R.M., Gianni, P.M., Trager, B.M.: A reordered Schur factorization method for zero-dimensional polynomial systems with multiple roots. In Proceedings of the 1997 International Symposium on Symbolic and Algebraic Computation (Kihei, HI), pages 133–140. ACM, New York (1997)

9. Demmel, J.: The condition number of equivalence transformations that block diagonalize matrix pencils. SIAM J. Numer. Anal. **20**(3), 599–610 (1983)
10. Eisenmann, H.: A Newton method for solving locally definite multiparameter eigenvalue problems by multiindex (2024). arXiv:2404.04194
11. Gantmacher, F.R.: The theory of matrices. Vols. 1, 2. Chelsea Publishing Co., New York. Translated by K. A. Hirsch (1959)
12. Gohberg, I., Lancaster, P., Rodman, L.: Invariant subspaces of matrices with applications, volume 51 of Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Reprint of the 1986 original (2006)
13. Golub, G.H., Van Loan, C.F.: Matrix computations, 4th edn. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD (2013)
14. Graf, E., Townsend, A.: Numerical instability of algebraic rootfinding methods (2024). arXiv:2408.02805
15. Gravenkamp, H., Plestenjak, B., Kiefer, D.A., Elias, J.: Computation of leaky waves in layered structures coupled to unbounded media by exploiting multiparameter eigenvalue problems. J. Sound Vib. **596**, 118716 (2025)
16. Greenbaum, A., Li, R.-C., Overton, M.L.: First-order perturbation theory for eigenvalues and eigenvectors. SIAM Rev. **62**(2), 463–482 (2020)
17. He, H., Kressner, D.: A randomized algorithm for simultaneously diagonalizing symmetric matrices by congruence (2024). arXiv:2402.16557
18. He, H., Kressner, D.: Randomized joint diagonalization of symmetric matrices. SIAM J. Matrix Anal. Appl. **45**(1), 661–684 (2024)
19. Hochstenbach, M.E., Košir, T., Plestenjak, B.: A Jacobi-Davidson type method for the two-parameter eigenvalue problem. SIAM J. Matrix Anal. Appl. 26(2),477–497 (2004/05)
20. Hochstenbach, M.E., Košir, T., Plestenjak, B.: Numerical methods for rectangular multiparameter eigenvalue problems, with applications to finding optimal ARMA and LTI models. Numer. Linear Algebra Appl., page e2540 (2023)
21. Hochstenbach, M.E., Mehl, C., Plestenjak, B.: Solving singular generalized eigenvalue problems by a rank-completing perturbation. SIAM J. Matrix Anal. Appl. **40**(3), 1022–1046 (2019)
22. Hochstenbach, M.E., Mehl, C., Plestenjak, B.: Solving singular generalized eigenvalue problems. Part II: Projection and augmentation. SIAM J. Matrix Anal. Appl. **44**(4),1589–1618 (2023)
23. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge; New York, 2nd edition (2013)
24. Karow, M., Kressner, D.: On a perturbation bound for invariant subspaces of matrices. SIAM J. Matrix Anal. Appl. **35**(2), 599–618 (2014)
25. Kato, T.: Perturbation theory for linear operators. Classics in Mathematics. Springer, Berlin, Heidelberg (1995)
26. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
27. Košir, T.: On the structure of commutative matrices. Linear Algebra Appl. **187**, 163–182 (1993)
28. Košir, T., Plestenjak, B.: On stability of invariant subspaces of commuting matrices. Linear Algebra Appl. **342**, 133–147 (2002)
29. Kressner, D., Šain Glibić, I.: Singular quadratic eigenvalue problems: linearization and weak condition numbers. BIT **63**(1), 18 (2023)
30. Lancaster, P., Markus, A.S., Zhou, F.: Perturbation theory for analytic matrix functions: the semisimple case. SIAM J. Matrix Anal. Appl. **25**(3), 606–626 (2003)
31. Manocha, D., Demmel, J.: Algorithms for intersecting parametric and algebraic curves I: simple intersections. ACM Trans. Graph. **13**(1), 73–100 (1994)
32. Möller, H.M., Stetter, H.J.: Multivariate polynomial equations with multiple zeros solved by matrix eigenproblems. Numer. Math. **70**(3), 311–329 (1995)
33. Morgan, A.: Solving polynomial systems using continuation for engineering and scientific problems, volume 57 of Classics in Applied Mathematics. SIAM, Philadelphia, PA (2009)
34. Muhič, A., Plestenjak, B.: On the quadratic two-parameter eigenvalue problem and its linearization. Linear Algebra Appl. **432**(10), 2529–2542 (2010)
35. Plestenjak, B.: MultiParEig 2.7. https://www.mathworks.com/matlabcentral/fileexchange/47844-multipareig (2022). MATLAB Central File Exchange
36. Plestenjak, B.: MultiParEig 2.8. https://www.mathworks.com/matlabcentral/fileexchange/47844-multipareig (2024). MATLAB Central File Exchange

37. Plestenjak, B., Hochstenbach, M.E.: Roots of bivariate polynomial systems via determinantal representations. SIAM J. Sci. Comput. **38**(2), A765–A788 (2016)
38. Shoven, J.: Applied general equilibrium modelling. IMF Staff. Pap. **30**, 394–419 (1983)
39. Slivnik, T., Tomšič, G.: A numerical method for the solution of two-parameter eigenvalue problems. J. Comput. Appl. Math. **15**(1), 109–115 (1986)
40. Stetter, H.J.: Numerical polynomial algebra. SIAM, Philadelphia, PA (2004)
41. Stewart, G.W.: Error and perturbation bounds for subspaces associated with certain eigenvalue problems. SIAM Rev. **15**, 727–764 (1973)
42. Stewart, G.W., Sun, J.G.: Matrix perturbation theory. Computer Science and Scientific Computing. Academic Press Inc, Boston, MA (1990)
43. Suprunenko, D., Tyshkevich, R.I.: Commutative matrices. Academic paperbacks. Mathematics. Academic Press, New York (1968)
44. Telen, S., Van Barel, M.: A stabilized normal form algorithm for generic systems of polynomial equations. J. Comput. Appl. Math. **342**, 119–132 (2018)
45. Van Dooren, P.: The computation of Kronecker's canonical form of a singular pencil. Linear Algebra Appl. **27**, 103–140 (1979)
46. Vermeersch, C.: The (Block) Macaulay Matrix: Solving Systems of Multivariate Polynomial Equations and Multiparameter Eigenvalue Problems. Phd thesis, Katholieke Universiteit Leuven (2023)
47. Vermeersch, C., De Moor, B.: Two complementary block Macaulay matrix algorithms to solve multiparameter eigenvalue problems. Linear Algebra Appl. **654**, 177–209 (2022)
48. Verschelde, J.: Algorithm 795: PHCpack: a general-purpose solver for polynomial systems by homotopy continuation. ACM Trans. Math. Softw. **25**(2), 251–276 (1999)