LINEARIZABLE EIGENVECTOR NONLINEARITIES *

ROB CLAES†, ELIAS JARLEBRING ‡, KARL MEERBERGEN †, AND PARIKSHIT UPADHYAYA ‡

Abstract. We present a method to linearize, without approximation, a specific class of eigenvalue problems with eigenvector nonlinearities (NEPv), where the nonlinearities are expressed by scalar functions that are defined by a quotient of linear functions of the eigenvector. The exact linearization relies on an equivalent multiparameter problem (MEP) that contains the exact solutions of the NEPv. Due to the characterization of MEPs in terms of a generalized eigenvalue problem this provides a direct way to compute all NEPv solutions for small problems, and it opens up the possibility to develop locally convergent iterative methods for larger problems. Moreover, the linear formulation allows us to easily determine the number of solutions of the NEPv. We propose two numerical schemes that exploit the structure of the linearization: inverse iteration and residual inverse iteration. We show how symmetry in the MEP can be used to improve reliability and reduce computational cost of both methods. Two numerical examples verify the theoretical results and a third example shows the potential of a hybrid scheme that is based on a combination of the two proposed methods.

Key words. nonlinear eigenvalue problem, multiparameter eigenvalue problem, linearization

AMS subject classifications. 65F15, 65H17

1. Introduction.

Consider the following eigenvalue problem

$$(A + \lambda B + f_1(x)C_1 + \cdots + f_m(x)C_m)x = 0 \tag{1.1}$$

with $A, B \in \mathbb{C}^{n\times n}$ and eigenvector nonlinearities defined by $f_i : \mathbb{C}^n \to \mathbb{C}$ and $C_i \in \mathbb{C}^{n\times n}$ for $i = 1, 2, \ldots, m$. The goal is to find $\lambda \in \mathbb{C}$ and $x \in \mathbb{C}^n\backslash\{0\}$ such that the eigenpair (λ, x) solves (1.1). This problem belongs to the class of nonlinear eigenvalue problems with eigenvector nonlinearities or *NEPv*. The functions f_i are invariant of the scaling of the vector such that $f_i(x) = f_i(\alpha x)$ for any $\alpha \in \mathbb{C}\backslash\{0\}$.

Problems with eigenvector nonlinearities arise in a wide range of fields. A common application is in the field of quantum chemistry for electronic structure calculations [32], where the discretization of the Kohn-Sham and Hartree-Fock equations leads to a NEPv. See [30] for a summary of numerical methods and discretization approaches. In these problems, the nonlinearity may depend on a set of eigenvectors, instead of a single eigenvector as in (1.1). The Gross-Pitaevskii equation is a related application in quantum physics, where the goal is to compute the ground state of bosons [3, 15]. A widely used approach for the solution in both these fields is the self-consistent field (SCF) iteration [8, 21] and its variants [26, 27]. SCF is an iterative scheme that involves solving a linear eigenproblem at each step using the iterate from the previous step, until convergence. For convergence results on SCF, see [7, 19, 20, 34, 35].

A different class of applications is in the field of data science. Spectral clustering using the eigenpairs of the p-Laplacian relies on solving a NEPv [6]. In [10], the authors propose an inverse power method for the problem. A Rayleigh quotient minimization approach applied to Fisher linear discriminant analysis is discussed in

*The work by Rob Claes and Karl Meerbergen is supported by the Research Foundation – Flanders (FWO) Grant G0B7818N and the KU Leuven Research Council.

†Department of Computer Science, KU Leuven, Leuven, Belgium, email: {rob.claes, karl.meerbergen}@kuleuven.be

‡Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, SeRC Swedish e-Science Research Center, email: {eliasj, pup}@kth.se

[2]. In [33], an equivalence between a NEPv and a new model for core-periphery detection in network science is shown.

Linearization without approximation in the sense of companion linearizations is a crucial tool for eigenvalue nonlinearities [22]. It is not only used as a procedure in itself, but serves as a basis of projection methods, where the linearization can provide insight in building a projection subspace. In this paper, we will study linearization in this sense, and for simplicity the term *linearization* will be used to refer to this, and not to the linear approximation of a nonlinear problem.

In this paper, we extend this notion of linearization to eigenvector nonlinearities by focusing on a subset of problems that can be written as (1.1) with

$$f_i(x) = \frac{r_i^T x}{s_i^T x}, \quad i = 1, 2, \ldots, m \tag{1.2}$$

and $r_i, s_i \in \mathbb{C}^n$. To illustrate the concept of the new method, consider the following example for $m = 1$:

$$(A + \lambda B + f_1(x)C_1)x = 0 \tag{1.3}$$

$$f_1(x) = \frac{r_1^T x}{s_1^T x}. \tag{1.4}$$

The linearization then consists of writing this problem as a linear multiparameter eigenvalue problem (MEP) as follows:

$$\begin{cases} -Ax = (\lambda B + \mu C_1)x \\ -(A + g_1 r_1^T)x = (\lambda B + \mu(C_1 - g_1 s_1^T))x \end{cases} \tag{1.5}$$

with mild conditions on $g_1 \in \mathbb{C}^n \backslash \{0\}$. Subtracting the first equation from the second results in

$$-g_1 r_1^T x = -\mu g_1 s_1^T x, \tag{1.6}$$

and since $g_1 \neq 0$ it follows that

$$\mu = \frac{r_1^T x}{s_1^T x} = f_1(x). \tag{1.7}$$

Finally, substituting this value of μ in the first equation of (1.5) shows that solutions of the NEPv are contained in the linear MEP (1.5). Note that this was achieved without any approximation.

In this paper we provide theoretical results concerning this class of problems, as well as new algorithms. More specifically, we describe the procedure outlined above for the general case, including a charactarization of a linearization to an MEP, which subsequently can be linearized to a generalized linear eigenvalue problem (GEP) using standard MEP techniques. We first show that, in general, the number of solutions to a problem belonging to this class is bounded. Theory is provided that justifies the method, in particular that the MEP is equivalent to the NEPv under certain conditions. The MEP always contains the NEPv solutions, but depending on certain free parameters (later denoted $g_1, \ldots, g_m$), the converse is not always true. We derive sufficient conditions which are generically satisfied, and show examples that illustrate degenerate cases.

The theory also reveals that it is sufficient to consider MEP solutions with a particular structure, namely that the eigenvector components are identical (referred to as symmetric MEP solutions). It turns out that certain iterative MEP methods preserve the symmetry, if they are initiated with a symmetric initial vector. This allows us to improve the algorithms. In particular, we improve the residual inverse iteration [25] and the inverse iteration applied to the GEP. We provide theory that specializes the characterization of convergence of these general methods. A set of experiments highlights the strengths of each method, and shows how the structure can arise from a PDE-problem.

The remainder of this paper is organized as follows. Section 2 gives an introduction to the theory of MEPs and shows how operator determinants can be used to extract the eigenpairs. This theory is used in section 3 to demonstrate the new linearization method that defines a multiparameter eigenvalue problem containing solutions to the nonlinear problem. Section 4 is devoted to numerical methods, convergence characterizations and simulations.

2. Preliminaries. The proposed method transforms the NEPv to a linear multiparameter eigenvalue problem. This section therefore first summarizes the relevant theory regarding multiparameter eigenvalue problems and their relation to operator determinants based on [1, 17].

We write the multiparameter eigenvalue problem (MEP) as a set of l coupled equations

$$\begin{cases} V_{1,0}x_1 &= (\mu_1 V_{1,1} + \mu_2 V_{1,2} + \cdots + \mu_l V_{1,l})x_1 \\ \quad\vdots & \\ V_{l,0}x_l &= (\mu_1 V_{l,1} + \mu_2 V_{l,2} + \cdots + \mu_l V_{l,l})x_l \end{cases} \tag{2.1}$$

with $V_{i,j} \in \mathbb{C}^{n_i \times n_i}$, $\mu_i \in \mathbb{C}$ and $x_i \in \mathbb{C}^{n_i}$ for $i = 1, 2, \ldots, l$ and $j = 0, 1, \ldots, l$. The goal is to find tuples $(\mu_1, \ldots, \mu_l, x_1, \ldots, x_l)$ with all x_i nonzero that solve (2.1). In general, this problem has $\prod_{i=1}^{l} n_i$ different isolated solution tuples. An important theoretical result is that these tuples can be extracted from a linear generalized eigenvalue problem (GEP). This equivalence between the MEP and the GEP is described by operator determinants. The remainder of this section describes such operator determinants and shows how they are used to construct an equivalent GEP to a given MEP.

In general, an operator determinant of a $k \times k$-array of matrices $D_{i,j}$ is denoted by

$$\begin{vmatrix} D_{1,1} & \cdots & D_{1,k} \\ \vdots & \ddots & \vdots \\ D_{k,1} & \cdots & D_{k,k} \end{vmatrix}_{\otimes} \tag{2.2}$$

where $D_{i,j} \in \mathbb{C}^{n_i \times n_i}$, $i, j = 1, \ldots, k$. The $\otimes$ subscript corresponds to usage of standard formulas for determinants, but where multiplication is replaced by the Kronecker product and the order of operations is chosen consistently. The operator determinant can therefore be defined by an operator cofactor expansion, e.g., expansion by the

first block row yields

$$\begin{aligned}\begin{vmatrix} D_{1,1} & \cdots & D_{1,k} \\ \vdots & \ddots & \vdots \\ D_{k,1} & \cdots & D_{k,k} \end{vmatrix}_{\otimes} :=& D_{1,1} \otimes \begin{vmatrix} D_{2,2} & \cdots & D_{2,k} \\ \vdots & \ddots & \vdots \\ D_{k,2} & \cdots & D_{k,k} \end{vmatrix}_{\otimes} + \dots \\ &+ (-1)^{k+1} D_{1,k} \otimes \begin{vmatrix} D_{2,1} & \cdots & D_{2,k-1} \\ \vdots & \ddots & \vdots \\ D_{k,1} & \cdots & D_{k,k-1} \end{vmatrix}_{\otimes}.\end{aligned} \tag{2.3}$$

Notice that not all properties for standard determinants can be applied to operator determinants. The following theorem lists properties that hold for operator determinants.

THEOREM 2.1. *[1, Theorem 6.2.1] For a determinant (2.2), we have the following properties*

1. *If two columns are interchanged, the sign of the determinant is reversed.*
2. *If two columns are identical, the determinant vanishes.*
3. *The value of the determinant is unchanged if a scalar multiple of one column is added to another column.*

An important difference to standard determinants, is that in general these properties only hold for columns and not for rows.

Now define Δ_0 and Δ_i, $i = 1, \dots, l$ as operator determinants constructed from the matrices of the MEP (2.1):

$$\Delta_0 = \begin{vmatrix} V_{1,1} & V_{1,2} & \cdots & V_{1,l} \\ \vdots & \vdots & & \vdots \\ V_{l,1} & V_{l,2} & \cdots & V_{l,l} \end{vmatrix}_{\otimes}, \tag{2.4}$$

$$\Delta_i = \begin{vmatrix} V_{1,1} & \cdots & V_{1,i-1} & V_{1,0} & V_{1,i+1} & \cdots & V_{1,m+1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ V_{l,1} & \cdots & V_{l,i-1} & V_{l,0} & V_{l,i+1} & \cdots & V_{l,l} \end{vmatrix}_{\otimes}. \tag{2.5}$$

We say that the MEP (2.1) is nonsingular if and only if Δ_0 is nonsingular. The following theorem summarizes the relation between the eigenvalues of the MEP and its operator determinants.

THEOREM 2.2. *[1, Chapter 6] The multiparameter eigenvalue problem (2.1) is equivalent to the following system of generalized eigenvalue problems*

$$\Delta_i z = \mu_i \Delta_0 z, \quad i = 1, \dots, l \tag{2.6}$$

with $z = x_1 \otimes \cdots \otimes x_l$. Moreover if the MEP is nonsingular, all matrices $\Delta_0^{-1}\Delta_i$ commute such that the vector z always exists.

This theorem suggests a way to extract the eigenvalues and eigenvectors of this system. In short, it suffices to solve one eigenvalue problem, e.g. (Δ_1, Δ_0), to find the complete eigentuples $(\mu_1, \dots, \mu_l, x_1, \dots, x_l)$. It is possible to reduce the computational cost by exploiting the structure of the operator determinants [23] or to use a subspace method for larger problems [11, 12, 13]. Subsection 3.2 discusses the system (2.6) in greater detail.

The linearization approach presented in this paper heavily relies on the concept of symmetric decomposable vectors.

DEFINITION 2.3. *A symmetric decomposable vector of rank k is a vector $z \in \mathbb{C}^{n^l}$ that can be decomposed in k linearly independent vectors as $z = \sum_{j=1}^{k} \alpha_j x_j \otimes \cdots \otimes x_j$ with $x_j \in \mathbb{C}^n$.*

For the remainder of this paper, a symmetric eigenvector is thus defined as a symmetric decomposable vector of rank 1 that is a solution to (2.6), and λ is its corresponding symmetric eigenvalue.

We also use the term *symmetric* in the context of solutions to MEPs.

DEFINITION 2.4. *A solution $(\mu_1, \ldots, \mu_l, x_1, \ldots, x_l)$ of the MEP (2.1) is called symmetric if*

$$x_i = \alpha_i x_1, \quad \forall i \in \{2, \ldots, l\}, \quad \alpha_i \in \mathbb{C} \setminus \{0\}. \tag{2.7}$$

If the assumptions of Theorem 2.2 are satisfied, then a symmetric solution of the MEP (2.1), for example, $(\mu_1, \ldots, \mu_l, x_1, \ldots, x_1)$ corresponds to an a rank one symmetric eigenvector $z = x_1 \otimes x_1 \otimes \cdots \otimes x_1$ of the equations in (2.6).

3. Linearization of NEPv. Before introducing the linearization of the NEPv (1.1)-(1.2), we characterize the number of solutions of this problem.

THEOREM 3.1. *The NEPv (1.1)-(1.2) has at most $N_s = \binom{n+m}{m+1}$ isolated solutions.*

Proof. First, rewrite the NEPv (1.1)-(1.2) as

$$(\lambda_0 A + \lambda_1 B + \lambda_2 C_1 + \cdots + \lambda_{m+1} C_{m+1})x = 0 \tag{3.1a}$$

$$\lambda_0 r_i^T x - \lambda_{i+1} s_i^T x = 0 \qquad i = 1, \ldots, m \tag{3.1b}$$

with $\lambda_k \in \mathbb{C}$ for $k = 0, \ldots, m+1$. This is a system of $n+m$ bihomogenous equations of degree $(1,1)$ on the multiplicative projective space $\mathbb{P}^{m+1} \times \mathbb{P}^{n-1}$. Here, a projective space $\mathbb{P}^k$ is defined by complex coordinates $(Z_1, \ldots, Z_{k+1}) \in \mathbb{C}^{k+1} \setminus \{0\}$ such that $(Z_1, \ldots, Z_{k+1})$ and $(\gamma Z_1, \ldots, \gamma Z_{k+1})$ correspond to the same point in the projective space for every $\gamma \in \mathbb{C} \setminus \{0\}$. By the multi-homogenous Bézout theorem [31], the number of solutions is bounded from above by the coefficient of $t_1^{m+1} t_2^{n-1}$ in the expansion of $(t_1 + t_2)^{n+m}$, which is the binomial coefficient $\binom{n+m}{m+1}$. The finite solutions of the NEPv correspond to the solutions of (3.1) by the equalities $\frac{\lambda_{i+1}}{\lambda_0} = \frac{r_i^T x}{s_i^T x} = f_i(x)$ for $i = 1, \ldots, m$ if $\lambda_0 \neq 0$ and $s_i^T x \neq 0$, which means that (1.1)-(1.2) cannot have more than $N_s = \binom{n+m}{m+1}$ finite solutions. ∎

It is important to emphasize that for generic problems of type (1.1)-(1.2), the upper bound N_s is almost always reached. Degenerate cases for which the system (3.1) has less than N_s solutions, form a set of measure zero. It also holds generically that $\lambda_0 \neq 0$ and $s_i^T x \neq 0$, which means that all N_s solutions of (3.1) correspond to finite isolated solutions of (1.1)-(1.2) and the upper bound is reached.

3.1. Linearization of NEPv to MEP. We will show that the original problem NEPv (1.1) can be cast to an equivalent MEP. We say that we linearize the NEPv to

this MEP

$$-Ax_1 = (\lambda B + \mu_1 C_1 + \cdots + \mu_m C_m)x_1 \tag{3.2a}$$

$$-(A + g_1 r_1^T)x_2 = (\lambda B + \mu_1 (C_1 - g_1 s_1^T) + \cdots + \mu_m C_m)x_2 \tag{3.2b}$$

$$\vdots$$

$$-(A + g_m r_m^T)x_{m+1} = (\lambda B + \mu_1 C_1 + \cdots + \mu_m (C_m - g_m s_m^T))x_{m+1} \tag{3.2c}$$

where $g_1, \ldots, g_m \neq 0$ can be arbitrary vectors under mild conditions that are further discussed in subsection 3.3.

THEOREM 3.2. *Assume (λ, x) is a solution to the NEPv (1.1), then*

$$(\lambda, \mu_1, \ldots, \mu_m, x_1, \ldots, x_{m+1}) = (\lambda, f_1(x), \ldots, f_m(x), x, \ldots, x) \tag{3.3}$$

is a solution to (3.2).

Proof. Define the variables $x_1, \ldots, x_{m+1}$ and $\mu_1, \ldots, \mu_m$ from the solution to (1.1) as follows: $\mu_1 := f_1(x)$, ..., $\mu_m := f_m(x)$ and $x_1 = \cdots = x_{m+1} = x$. Equation (3.2a) is satisfied due to (1.1). Equation (3.2b) is satisfied by forming the sum of (1.1) and the following identity multiplied by g_1

$$r_1^T x = \mu_1 s_1^T x, \tag{3.4}$$

which follows from the definition of $\mu_1 = f_1(x) = \frac{r_1^T x}{s_1^T x}$. The other equations in (3.2) follow from an analogous construction. ∎

The converse of the theorem above is illustrated by two theorems that follow. The first theorem illustrates that every symmetric solution to the MEP leads to a solution to the NEPv under an assumption on x, which characterizes potential spurious solutions introduced by our approach.

THEOREM 3.3. *Assume $(\lambda, \mu_1, \ldots, \mu_m, x_1, \ldots, x_{m+1})$ is a solution to (3.2). If $x_1 = x_2 = \cdots = x_{m+1} = x$ then $(\lambda, x_1) = (\lambda, x)$ is a solution to (1.1)-(1.2), if $s_j^T x \neq 0$, for all $j = 1, \ldots, m$.*

Proof. By subtracting equation (3.2b) from (3.2a) we find that

$$g_1 r_1^T x = \mu_1 g_1 s_1^T x. \tag{3.5}$$

By assumption $s_1^T x \neq 0$ and $g_1 \neq 0$, such that $\mu_1 = \frac{r_1^T x}{s_1^T x} = f_1(x)$. By considering differences between the first and kth row in (3.2) we obtain analogously

$$\mu_i = f_i(x), \quad i = 1, \ldots, m. \tag{3.6}$$

The connection with (1.1) is established by using (3.2a) and the formulas for μ in (3.6). ∎

For the next result, first consider the following problem, with the same eigenvalues as (3.2).

$$-A^T y_1 = (\lambda B^T + \mu_1 C_1^T + \cdots + \mu_m C_m^T)y_1 \tag{3.7a}$$

$$-(A^T + r_1 g_1^T)y_2 = (\lambda B^T + \mu_1 (C_1^T - s_1 g_1^T) + \cdots + \mu_m C_m^T)y_2 \tag{3.7b}$$

$$\vdots$$

$$-(A^T + r_m g_m^T)y_{m+1} = (\lambda B^T + \mu_1 C_1^T + \cdots + \mu_m (C_m^T - s_m g_m^T))y_{m+1} \tag{3.7c}$$

A tuple of vectors $(y_1, \ldots, y_{m+1})$ that solves this problem is called a left eigenvector of (3.2). The next theorem shows that even non-symmetric solutions to the MEP lead to a solution to the NEPv under an additional assumption on the g_i vectors.

THEOREM 3.4. *Assume* $(\lambda, \mu_1, \ldots, \mu_m, x_1, \ldots, x_{m+1})$ *is a solution to* (3.2). *Let* $(\lambda, \mu_1, \ldots, \mu_m, y_1, \ldots, y_{m+1})$ *be the solution to the problem given by* (3.7). *If* $s_j^T x_1 \neq 0$ *and* $g_j^T y_{j+1} \neq 0$ *for all* $j = 1, \ldots, m$, *then* (λ, x_1) *is a solution to* (1.1)-(1.2).

Proof. Left-multiplying all the subequations except the first in (3.7) by x_1^T and using (3.2a), we have

$$x_1^T r_j g_j^T y_{j+1} = \mu_j x_1^T s_j g_j^T y_{j+1}, \quad j = 1, \ldots, m. \tag{3.8}$$

The assumptions $s_j^T x_1 \neq 0$ and $g_j^T y_{j+1} \neq 0$ for all $j = 1, \ldots, m$ lead to

$$\mu_j = \frac{r_j^T x_1}{s_j^T x_1}, \quad j = 1, \ldots, m. \tag{3.9}$$

Replacing the above for μ_j in (3.2a), we see that (λ, x_1) is a solution to (1.1). ∎

Theorem 3.4 together with Theorem 3.2 leads to the conclusion that under certain conditions, for every non-symmetric solution $(\lambda, \mu_1, \ldots, \mu_m, x_1 \ldots, x_m)$ to the MEP (3.2), there is a corresponding symmetric solution $(\lambda, \mu_1, \ldots, \mu_m, x_1, \ldots, x_1)$ and both lead to the same solution (λ, x_1) for the NEPv. Since our main interest is to compute solutions to the NEPv, it suffices to restrict our search to the set of symmetric solutions of the MEP. This observation forms the basis of the algorithms in section 4.

3.2. Linearization of NEPv to GEP. The theorems of the preceding subsection show how a solution to the NEPv corresponds to a solution of MEP (3.2), and vice versa. Theorem 2.2 shows an equivalence between a MEP and a system of generalized eigenvalue problems (GEP) involving operator determinants, under certain assumptions. This means that the NEPv can be linearized to a system of GEPs. For completeness, we now specify the structure of these operator determinants arising from the MEP in (3.2).

The operator determinants of (2.6) for the MEP in (3.2) are as follows:

$$\Delta_0 = \begin{vmatrix} B & C_1 & \cdots & C_m \\ B & C_1 - g_1 s_1^T & \cdots & C_m \\ \vdots & \vdots & \ddots & \vdots \\ B & C_1 & \cdots & C_m - g_m s_m^T, \end{vmatrix}_{\otimes} \tag{3.10}$$

and

$$\Delta_i = \begin{vmatrix} B & C_1 & \cdots & -A & \cdots & C_m \\ B & C_1 - g_1 s_1^T & \cdots & -A - g_1 r_1^T & \cdots & C_m \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ B & C_1 & \cdots & \underbrace{-A - g_m r_m^T}_{\text{column } i} & \cdots & C_m - g_m s_m^T \end{vmatrix}_{\otimes} \tag{3.11}$$

for $i = 1, \ldots, m+1$. Note that Δ_i is obtained by taking Δ_0 and only changing the i-th column.

For the linearized NEPv, or the MEP (3.2), the equations in (2.6) lead to an equivalent system of $m+1$ GEPs

$$\Delta_1 z = \lambda \Delta_0 z, \tag{3.12a}$$

$$\Delta_{i+1} z = \mu_i \Delta_0 z, \quad i = 1, \ldots, m. \tag{3.12b}$$

As discussed at the end of section 2, the set of equations in (3.12) forms a natural solution procedure involving two steps. Firstly, the GEP in (3.12a) can be solved to find λ and z. The other values of the eigenvalue tuple can be obtained using the corresponding Rayleigh quotients as

$$\mu_i = \frac{z^H \Delta_{i+1} z}{z^H \Delta_0 z}, \quad i = 1, \ldots, m. \tag{3.13}$$

If z is decomposable as $x_1 \otimes \ldots \otimes x_{m+1}$ and the assumptions of Theorem 3.4 hold, then (λ, x_1) is a solution to the NEPv and we can compute the values of μ_i without using Rayleigh quotients, as follows:

$$\mu_i = \frac{r_i^T x_1}{s_i^T x_1}, \quad i = 1, \ldots, m. \tag{3.14}$$

We illustrate this with a small example with matrices from $\mathbb{R}^{2\times 2}$, where $m = 1$. Let

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad r = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad s = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \tag{3.15}$$

and consider the problem of finding (λ, x) in $\mathbb{C} \times \mathbb{C}^2$ such that

$$\left(A + \lambda B + \frac{r^T x}{s^T x} C \right) x = 0. \tag{3.16}$$

We use $g = \begin{bmatrix} 1 & 3 \end{bmatrix}^T$, which leads to

$$\Delta_0 = \begin{vmatrix} B & C \\ B & (C - gs^T) \end{vmatrix}_{\otimes} = \begin{bmatrix} -4 & -7 & -4 & -6 \\ -18 & -16 & -24 & -16 \\ -6 & -9 & -9 & -14 \\ -36 & -24 & -51 & -36 \end{bmatrix}, \tag{3.17a}$$

$$\Delta_1 = \begin{vmatrix} -A & C \\ -(A + gr^T) & C - gs^T \end{vmatrix}_{\otimes} = \begin{bmatrix} 10 & 9 & 2 & 3 \\ 30 & 22 & 12 & 8 \\ 0 & 0 & 6 & 6 \\ 0 & 0 & 21 & 15 \end{bmatrix}. \tag{3.17b}$$

The four eigenpairs obtained by solving the first GEP in (3.12) are

$$\lambda_1 \approx 5.2462, \quad z_1 \approx \begin{bmatrix} 1 \\ -0.69 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -0.69 \end{bmatrix}, \tag{3.18a}$$

$$\lambda_2 \approx -0.4224, \quad z_2 \approx \begin{bmatrix} 1 \\ -1.47 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -1.47 \end{bmatrix}, \tag{3.18b}$$

$$\lambda_3 \approx -0.4367, \quad z_3 \approx \begin{bmatrix} 1 \\ -14.8 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -14.8 \end{bmatrix}, \tag{3.18c}$$

$$\lambda_4 \approx -1.2500, \quad z_4 \approx \begin{bmatrix} 1 \\ -1.88 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -0.83 \end{bmatrix}. \tag{3.18d}$$

It can be verified by substituting into (1.1) that the first three eigenpairs with symmetric eigenvectors give us solutions to the original NEPv.

However, the fourth eigenpair is a spurious solution in the sense that it does not lead to a solution of the original NEPv, even though it is a solution to the MEP. The fourth eigenvalue can be analytically verified to be $\lambda = -\frac{5}{4}$ with left eigenvector

$$w_4 = \begin{bmatrix} -3 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} -3 \\ 1 \end{bmatrix}. \tag{3.19}$$

Hence, left eigenvectors of the MEP are given by $y_1 = y_2 = (-3, 1)^T$. This implies that $g^T y_1 = g^T y_2 = 0$, which is a violation of the assumptions of Theorem 3.4, which would otherwise state that $-5/4$ is an eigenvalue of the NEPv. This explains why the fourth eigenpair corresponds to a spurious solution.

For the NEPv in (3.16), we have $n = 2$ and $m = 1$. Theorem 3.1 then implies that there are at most three isolated solutions. On the other hand, the associated GEP can have upto four distinct eigenpairs for some choices of g. Hence, for those choices of g, there has to be at least one spurious solution among the eigenpairs of the GEP, implying that the assumptions of Theorem 2.2 and Theorem 3.4 do not hold true. For the case when the assumptions of the two theorems do hold, the following two lemmas illustrate the correspondence between simple eigenpairs of the GEP and solutions of the NEPv.

LEMMA 3.5. *Under the assumptions of Theorem 2.2 and Theorem 3.4, eigenvectors of simple eigenvalues of (3.12) are rank one symmetric.*

Proof. Consider a simple eigenpair (λ, z_1) of (3.12) such that $z_1 = x_1 \otimes x_2 \otimes \cdots \otimes x_{m+1}$. If the assumptions of Theorem 2.2 and Theorem 3.4 hold, then (λ, x_1) is a solution to (3.2) and hence, λ is not a spurious eigenvalue. Then, Theorem 3.2, dictates that $(\lambda, f_1(x_1), \ldots, f_m(x_1), x_1, \ldots, x_1)$ is a solution to (3.2). Therefore, (λ, z_2) is also an eigenpair where $z_2 = x_1 \otimes x_1 \otimes \cdots \otimes x_1$. Since λ is a simple eigenvalue, z_1 and z_2 are linearly dependent, which implies

$$x_i = \alpha_i x_1, \quad i = 2, \ldots, m+1 \tag{3.20}$$

for $\alpha_i \in \mathbb{C}$. Hence, z_1 is rank one symmetric. ∎

LEMMA 3.6. *If the NEPv (1.1)-(1.2) has N_s distinct eigenvalues, the operator determinant system (3.12) has N_s linearly independent rank one symmetric eigenvectors.*

Proof. If (1.1)-(1.2) has N_s distinct eigenvalues, the operator determinant system (3.12) has N_s symmetric eigenvalues, and N_s corresponding rank one symmetric eigenvectors. Since these eigenvectors belong to distinct eigenvalues of (3.12), they are linearly independent. ∎

Note that the space of symmetric vectors has dimensions N_s, which means that there can be no more than N_s linearly independent rank one symmetric vectors.

3.3. Remarks on the choice of g_i. For a NEPv with m different nonlinear terms f_i, our linearization technique introduces m additional vectors g_i that can be chosen freely. The theorems in the previous section suggest that almost all choices of g_i are effective, in the sense that the resulting MEP can be used to find solutions to the original NEPv. More specifically, Theorem 3.4 says that a sufficient condition for the MEP solutions to correspond to solutions to the original NEPv is that $g_i^T y_{i+1} \neq 0$

for all i. The theory is inconclusive when $g_i^T y_{i+1} = 0$ for some i. As we will illustrate now, this may lead to a MEP with a continuum of solutions, from which the NEPv solutions are difficult to identify. This would make our approach impractical for those choices.

Consider the NEPv (1.1) with $m = 2$ and assume $g_1 = g_2$. We illustrate how choosing $g_1 = g_2$ leads to an MEP with a continuum of solutions. To this end, we first consider the following GEP parameterized with λ, with an additional orthogonality constraint

$$-(A^T + \lambda B^T)y = (\mu_1 C_1^T + \mu_2 C_2^T)y, \tag{3.21a}$$

$$g^T y = 0. \tag{3.21b}$$

Relation (3.21) with a normalization condition forms $n+2$ equations and has $n+3$ free variables. Therefore, assuming that the conditions of the implicit function theorem are satisfied, we can view μ_1, μ_2, y as functions of λ: $\mu_1(\lambda), \mu_2(\lambda), y(\lambda)$.

Any solution (μ_1, μ_2, y) to (3.21) leads to a corresponding symmetric solution $(\lambda, \mu_1, \mu_2, y, y, y)$ to the MEP given by

$$-A^T y_1 = (\lambda B^T + \mu_1 C_1^T + \mu_2 C_2^T)y_1 \tag{3.22a}$$

$$-(A^T + r_1 g^T)y_2 = (\lambda B^T + \mu_1 (C_1^T - s_1 g^T) + \mu_2 C_2^T)y_2 \tag{3.22b}$$

$$-(A^T + r_2 g^T)y_3 = (\lambda B^T + \mu_1 C_1^T + \mu_2 (C_2^T - s_2 g^T))y_3. \tag{3.22c}$$

Note that if $r_i \neq \alpha_i s_i$ for any $\alpha_i \in \mathbb{C}$, then every symmetric solution $(\gamma, \mu_1, \mu_2, y, y, y)$ to (3.22) also leads to a solution (μ_1, μ_2, y) to (3.21) with $\lambda = \gamma$. Observe that the MEP in (3.22) is the linearization of the NEPv with $g_1 = g_2 = g$, but formulated with left eigenvectors as in (3.7). The corresponding right eigenvector problem is

$$-Ax_1 = (\lambda B + \mu_1 C_1 + \mu_2 C_2)x_1 \tag{3.23a}$$

$$-(A + g r_1^T)x_2 = (\lambda B + \mu_1 (C_1 - g s_1^T) + \mu_2 C_2)x_2 \tag{3.23b}$$

$$-(A + g r_2^T)x_3 = (\lambda B + \mu_1 C_1 + \mu_2 (C_2 - g s_2^T))x_3. \tag{3.23c}$$

Due to the parameterization of solutions of (3.21) in terms of λ, equation (3.22) has a continuum of solutions given by the tuple $(\lambda, \mu_1(\lambda), \mu_2(\lambda), y(\lambda), y(\lambda), y(\lambda))$. Consequently, we have a continuum of solutions $(\lambda, \mu_1(\lambda), \mu_2(\lambda), x_1(\lambda), x_2(\lambda), x_3(\lambda))$ to the MEP stemming from the linearization (3.23), which undesirable. Only some of the solutions in the continuum are NEPv solutions and they are not easily characterized or computed. This leads to a singular situation for the linearization (3.12a), i.e., $\det(\Delta_1 - \lambda\Delta_0) \equiv 0$ although only some of the values of λ are NEPv solutions. In practice, we also observed that the singular subspace associated with the GEP had a dimension larger than one. Hence, choosing equal g_1 and g_2 is problematic, and should be avoided.

4. Numerical methods. In the previous section, we showed how solutions of the nonlinear problem are contained in a linear multiparameter eigenvalue problem. The eigenpairs can be extracted by solving a set of coupled generalized eigenvalue problems, but the large size of these problems is a limiting factor. In this section we present two specialized methods that extract symmetric solutions by exploiting the structure of the problem.

4.1. Residual inverse iteration. The method of residual inverse iteration has been proposed for nonlinear eigenvalue problems in [24] and was specialized for nonlinear two-parameter eigenvalue problems in [25]. The notation from [25] is used to

introduce the main properties of residual inverse iteration for the MEP defined by (3.2). Furthermore, we show how to exploit the structure of our problem to reduce computational cost and guarantee local convergence to a symmetric solution. To ease the presentation, the method is first introduced for two-parameter problems and later generalized to problems with an arbitrary number of parameters.

First, rewrite the two-parameter problem using matrix-valued functions as

$$(A + \lambda B + \mu_1 C_1)x_1 := T_1(\lambda, \mu_1)x_1 = 0 \tag{4.1a}$$

$$\left(A + g_1 r_1^T + \lambda B + \mu_1 (C_1 - g_1 s_1^T)\right) x_2 := T_2(\lambda, \mu_1)x_2 = 0. \tag{4.1b}$$

The solutions of this system can be expressed as zeros of

$$F(x_1, x_2, \lambda, \mu) := \begin{bmatrix} T_1(\lambda, \mu)x_1 \\ T_2(\lambda, \mu)x_2 \\ v_1^T x_1 - 1 \\ v_2^T x_2 - 1 \end{bmatrix} \tag{4.2}$$

where the last two rows are normalization constraints. The vectors v_1 and v_2 are chosen not orthogonal to an eigenvector. Residual inverse iteration is derived from the quasi-Newton method applied to (4.2):

$$\tilde{J}_F(x_1^{(k)}, x_2^{(k)}, \lambda^{(k)}, \mu^{(k)}) \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \Delta\lambda \\ \Delta\mu \end{bmatrix} = - \begin{bmatrix} T_1(\lambda^{(k)}, \mu^{(k)})x_1^{(k)} \\ T_2(\lambda^{(k)}, \mu^{(k)})x_2^{(k)} \\ v_1^T x_1^{(k)} - 1 \\ v_2^T x_2^{(k)} - 1 \end{bmatrix} \tag{4.3}$$

with approximated Jacobian $\tilde{J}_F(x_1^{(k)}, x_2^{(k)}, \lambda^{(k)}, \mu^{(k)})$ equal to

$$\begin{bmatrix} T_1(\sigma, \tau) & 0 & \frac{\partial T_1}{\partial \lambda} x_1^{(k)} & \frac{\partial T_1}{\partial \mu} x_1^{(k)} \\ 0 & T_2(\sigma, \tau) & \frac{\partial T_2}{\partial \lambda} x_2^{(k)} & \frac{\partial T_2}{\partial \mu} x_2^{(k)} \\ v_1^T & 0 & 0 & 0 \\ 0 & v_2^T & 0 & 0 \end{bmatrix}. \tag{4.4}$$

Matrices $T_1(\sigma, \tau)$ and $T_2(\sigma, \tau)$ are nonsingular approximations to $T_1(\lambda^{(k)}, \mu^{(k)})$ and $T_2(\lambda^{(k)}, \mu^{(k)})$ respectively. The arguments of the partial derivatives are omitted because they are constant for linear functions.

THEOREM 4.1. *If T_1 and T_2 are linear and $x_1^{(k)}$ and $x_2^{(k)}$ are normalized in every iteration, the quasi-Newton system (4.3)-(4.4) is equivalent to the following update steps:*

$$\begin{bmatrix} v_1^T T_1(\sigma, \tau)^{-1} \frac{\partial T_1}{\partial \lambda} x_1^{(k)} & v_1^T T_1(\sigma, \tau)^{-1} \frac{\partial T_1}{\partial \mu} x_1^{(k)} \\ v_2^T T_2(\sigma, \tau)^{-1} \frac{\partial T_2}{\partial \lambda} x_2^{(k)} & v_2^T T_2(\sigma, \tau)^{-1} \frac{\partial T_2}{\partial \mu} x_2^{(k)} \end{bmatrix} \begin{bmatrix} \Delta\lambda \\ \Delta\mu \end{bmatrix} = - \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \tag{4.5}$$

with $\gamma_i = v_i^T T_i(\sigma, \tau)^{-1} T_i(\lambda^{(k)}, \mu^{(k)}) x_i^{(k)}$ for $i = 1, 2$ and

$$x_1 = x_1^{(k)} - T_1(\sigma, \tau)^{-1} T_1(\lambda^{(k+1)}, \mu^{(k+1)}) x_1^{(k)} \tag{4.6}$$

$$x_2 = x_2^{(k)} - T_2(\sigma, \tau)^{-1} T_2(\lambda^{(k+1)}, \mu^{(k+1)}) x_2^{(k)} \tag{4.7}$$

such that $x_1^{(k+1)} = \frac{x_1}{v_1^T x_1}$ and $x_2^{(k+1)} = \frac{x_2}{v_2^T x_2}$.

Proof. Solving the first two rows of (4.3) for Δx_i results in

$$\Delta x_i = -T_i(\sigma,\tau)^{-1}\left(T_i(\lambda^{(k)},\mu^{(k)}) + \Delta\lambda\frac{\partial T_i}{\partial\lambda} + \Delta\mu\frac{\partial T_i}{\partial\mu}\right)x_i^{(k)} \quad i=1,2. \tag{4.8}$$

If the vectors $x_1^{(k)}$ and $x_2^{(k)}$ are normalized in each iterations, it follows from the last two rows of (4.3) that $v_1^T\Delta x_1 = 0$ and $v_2^T\Delta x_2 = 0$. Solving $v_1^T\Delta x_1 = 0$ and $v_2^T\Delta x_2 = 0$ for $\Delta\lambda$ and $\Delta\mu$ leads to the 2×2 system (4.5). Note that the matrix of this system is the Schur complement of (4.4) with respect to the bottom right block of zeros. Now, since T_i is linear, it follows that

$$\Delta x_i = -T_i(\sigma,\tau)^{-1}T_i(\lambda^{(k+1)},\mu^{(k+1)})x_i^{(k)} \quad i=1,2, \tag{4.9}$$

which is equivalent to (4.6)-(4.7). A final normalization step, $x_i^{(k+1)} = \frac{x_i^{(k+1)}}{v_i^T x_i^{(k+1)}}$ for $i=1,2$, proofs the stated equivalence. ∎

THEOREM 4.2. *(Theorem 5.2 [25]) Let $(\lambda^{(*)},\mu^{(*)})$ be a simple eigenvalue of (4.1) and let $x_1\otimes x_2$ be a corresponding right eigenvector such that $v_1^Tx_1 = v_2^Tx_2 = 1$. If $(\lambda^{(0)},\mu^{(0)})$ is sufficiently close to the eigenvalue $(\lambda^{(*)},\mu^{(*)})$, then residual inverse iteration has linear convergence close to the solution.*

THEOREM 4.3. *If residual inverse iteration is applied to problem (4.1) with $v_1 = v_2 = v$ and $x_1^{(k)} = x_2^{(k)} = x^{(k)}$ such that $v^TT_1(\sigma,\tau)^{-1}g_1 \neq 0$, then $x_1^{(k+1)} = x_2^{(k+1)} = x^{(k+1)}$ for $k = 0,1,2,\ldots$.*

Proof. See the proof of Theorem 4.5 ∎

The result presented in Theorem 4.3 can be used to reduce the computational cost of the residual inverse iteration method for solving NEPv problems. The structure exploiting method for finding symmetric solutions of (4.1) is given in Algorithm 4.1. First, note that in the linear case $\frac{\partial T_i}{\partial\lambda}$ and $\frac{\partial T_i}{\partial\mu}$ are constants such that the values u_i and w_i can be precalculated before the loop to reduce the computational cost. Second, the symmetric method uses only one solution vector, x, which implies that we only need to update one vector.

Algorithm 4.1 Symmetric residual inverse iteration for NEPv with 1 nonlinear term

Start with $\sigma = \lambda^{(0)}, \tau = \mu^{(0)}, v$ and $x^{(0)}$ such that $v^Tx^{(0)} = 1$.
Compute $u_i = v^TT_i(\sigma,\tau)^{-1}\frac{\partial T_i}{\partial\lambda}$ for $i=1,2$.
Compute $w_i = v^TT_i(\sigma,\tau)^{-1}\frac{\partial T_i}{\partial\mu}$ for $i=1,2$.
for k=0,1,2,... **do**
  Compute $\gamma_i = v^TT_i(\sigma,\tau)^{-1}T_i(\lambda^{(k)},\mu^{(k)})x^{(k)}$ for $i=1,2$.
  Solve

$$\begin{bmatrix} u_1^Tx^{(k)} & w_1^Tx^{(k)} \\ u_2^Tx^{(k)} & w_2^Tx^{(k)}\end{bmatrix}\begin{bmatrix}\Delta\lambda\\ \Delta\mu\end{bmatrix} = -\begin{bmatrix}\gamma_1\\ \gamma_2\end{bmatrix}.$$

  Update $\lambda^{(k+1)} = \lambda^{(k)} + \Delta\lambda$ and $\mu^{(k+1)} = \mu^{(k)} + \Delta\mu$.
  $x^{(k+1)} = x^{(k)} - T_1(\sigma,\tau)^{-1}T_1(\lambda^{(k+1)},\mu^{(k+1)})x^{(k)}$
  Normalize $x^{(k+1)} = x^{(k+1)}/(v^Tx^{(k+1)})$.
end for

Generalization to more parameters. The theoretical derivation in the preceding part is deliberately limited to two parameters to keep the notation clear. However, the method can be elegantly generalized to handle more than two parameters. The derivation starts by rewriting the linearization (3.2) as an MEP with $m+1$ parameters in the following form

$$\left(A + \lambda B + \sum_{i=1}^{m} \mu_i C_i\right) x_1 := T_1(\lambda, \boldsymbol{\mu}) x_1 = 0 \tag{4.10a}$$

$$\left(A + g_1 r_1^T + \lambda B + \sum_{i=1}^{m} \mu_i C_i - \mu_1 g_1 s_1^T\right) x_2 := T_2(\lambda, \boldsymbol{\mu}) x_2 = 0 \tag{4.10b}$$

$$\vdots$$

$$\left(A + g_m r_m^T + \lambda B + \sum_{i=1}^{m} \mu_i C_i - \mu_m g_m s_m^T)\right) x_{m+1} := T_{m+1}(\lambda, \boldsymbol{\mu}) x_{m+1} = 0 \tag{4.10c}$$

with $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$. The solutions to this system can be expressed as zeros of

$$F(x_1, \ldots, x_{m+1}, \lambda, \boldsymbol{\mu}) := \begin{bmatrix} T_1(\lambda, \boldsymbol{\mu}) x_1 \\ \vdots \\ T_{m+1}(\lambda, \boldsymbol{\mu}) x_{m+1} \\ v_1^T x_1 - 1 \\ \vdots \\ v_{m+1}^T x_{m+1} - 1 \end{bmatrix}. \tag{4.11}$$

The quasi-Newton method is now applied to (4.11):

$$\tilde{J}_F(x_1^{(k)}, \ldots, x_{m+1}^{(k)}, \lambda^{(k)}, \boldsymbol{\mu}^{(k)}) \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_{m+1} \\ \Delta\lambda \\ \Delta\boldsymbol{\mu} \end{bmatrix} = -F(x_1^{(k)}, \ldots, x_{m+1}^{(k)}, \lambda^{(k)}, \boldsymbol{\mu}^{(k)}) \tag{4.12}$$

with approximated Jacobian $\tilde{J}_F(x_1^{(k)}, \ldots, x_{m+1}^{(k)}, \lambda^{(k)}, \boldsymbol{\mu}^{(k)})$ equal to the block matrix

$$\begin{bmatrix} J_{1,1} & J_{1,2} \\ J_{2,1} & 0 \end{bmatrix} \tag{4.13}$$

where the blocks are defined as

$$J_{1,1} = \begin{bmatrix} T_1(\sigma, \boldsymbol{\tau}) & & \\ & \ddots & \\ & & T_{m+1}(\sigma, \boldsymbol{\tau}) \end{bmatrix}, \tag{4.14}$$

$$J_{1,2} = \begin{bmatrix} \frac{\partial T_1}{\partial \lambda} x_1^{(k)} & \cdots & \frac{\partial T_1}{\partial \mu_m} x_1^{(k)} \\ \vdots & \ddots & \vdots \\ \frac{\partial T_{m+1}}{\partial \lambda} x_{m+1}^{(k)} & \cdots & \frac{\partial T_{m+1}}{\partial \mu_m} x_{m+1}^{(k)} \end{bmatrix} \text{ and} \tag{4.15}$$

$$J_{2,1} = \begin{bmatrix} v_1^T & & \\ & \ddots & \\ & & v_{m+1}^T \end{bmatrix}. \tag{4.16}$$

The matrices $T_i(\sigma, \boldsymbol{\tau})$ are fixed nonsingular approximations to $T_i(\lambda^{(k)}, \boldsymbol{\mu}^{(k)})$.

THEOREM 4.4. *If T_i is linear and $x_i^{(k)}$ is normalized in every iteration for all $i = 1, \ldots, m+1$, then the quasi-Newton system (4.12) is equivalent to the following update steps:*

$$\begin{bmatrix} \nu_{\lambda,1} & \nu_{\mu_1,1} & \cdots & \nu_{\mu_m,1} \\ \nu_{\lambda,2} & \nu_{\mu_1,2} & \cdots & \nu_{\mu_m,2} \\ \vdots & \vdots & \ddots & \vdots \\ \nu_{\lambda,m+1} & \nu_{\mu_1,m+1} & \cdots & \nu_{\mu_m,m+1} \end{bmatrix} \begin{bmatrix} \Delta\lambda \\ \Delta\mu_1 \\ \vdots \\ \Delta\mu_m \end{bmatrix} = - \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{m+1} \end{bmatrix} \tag{4.17}$$

with $\nu_{\phi,i} = v_i^T T_i(\sigma, \boldsymbol{\tau})^{-1} \frac{\partial T_i}{\partial \phi} x_i^{(k)}$ and $\gamma_i = v_i^T T_i(\sigma, \boldsymbol{\tau})^{-1} T_i(\lambda^{(k)}, \boldsymbol{\mu}^{(k)}) x_i^{(k)}$ for $i = 1, \ldots, m+1$ and

$$x_i = x_i^{(k)} - T_i(\sigma, \boldsymbol{\tau})^{-1} T_i(\lambda^{(k+1)}, \boldsymbol{\mu}^{(k+1)}) x_i^{(k)} \tag{4.18}$$

such that $x_i^{(k+1)} = \frac{x_i}{v_i^T x_i}$.

The proof is analogous to the proof of Theorem 4.1 and is omitted for brevity. We will now restate an important property of residual inverse iteration when applied to problem (4.10).

THEOREM 4.5. *If residual inverse iteration is applied to problem (4.10) with $v_1 = \cdots = v_{m+1} = v$ and $x_1^{(k)} = \cdots = x_{m+1}^{(k)} = x^{(k)}$ such that $v^T T_1(\sigma, \boldsymbol{\tau})^{-1} g_i \neq 0$ for $i = 1, \ldots, m$, then $x_1^{(k+1)} = \cdots = x_{m+1}^{(k+1)} = x^{(k+1)}$ for $k = 0, 1, 2, \ldots$.*

Proof. The following proof shows that $x_1^{(k+1)} = x_i^{(k+1)}$ if $x_1^{(k)} = x_i^{(k)}$ for $i = 2, \ldots, m+1$.

The bottom rows of (4.12) imply that $v^T \Delta x_1 = v^T \left(x_1^{(k+1)} - x_1^{(k)} \right) = 0$ and $v^T \Delta x_i = v^T \left(x_i^{(k+1)} - x_i^{(k)} \right) = 0$ if $x_1^{(k)}$ and $x_i^{(k)}$ are normalized. Assuming $x_1^{(k)} = x_i^{(k)}$, implies that

$$v^T x_1^{(k+1)} = v^T x_i^{(k+1)}. \tag{4.19}$$

Now rewrite the first and i-th row of (4.12) as

$$T_1(\sigma, \boldsymbol{\tau}) x_1^{(k+1)} = - \left(\Delta\lambda B + \Delta\mu_1 C_1 + \cdots + \Delta\mu_m C_m - T_1(\sigma, \boldsymbol{\tau}) + T_1(\lambda^{(k)}, \boldsymbol{\mu}^{(k)}) \right) x_1^{(k)} \tag{4.20}$$

$$\begin{aligned} T_i(\sigma, \boldsymbol{\tau}) x_i^{(k+1)} = - \Big(& \Delta\lambda B + \Delta\mu_1 C_1 + \cdots + \Delta\mu_{i-1} \left(C_{i-1} - g_{i-1} s_{i-1}^T \right) + \cdots \\ & + \Delta\mu_m C_m - T_i(\sigma, \boldsymbol{\tau}) + T_i(\lambda^{(k)}, \boldsymbol{\mu}^{(k)}) \Big) x_i^{(k)}. \end{aligned} \tag{4.21}$$

Subtracting (4.20) from (4.21) gives

$$T_1(\sigma, \boldsymbol{\tau})(x_i^{(k+1)} - x_1^{(k+1)}) = g_{i-1}\zeta \tag{4.22}$$

with

$$\zeta = (\Delta\mu_{i-1}s_{i-1}^T + (\mu_{i-1} - \tau_{i-1})s_{i-1}^T)x_1^{(k)} - (r_{i-1}^T + \tau_{i-1}s_{i-1}^T)x_i^{(k+1)}. \tag{4.23}$$

Finally, we left-multiply with $v^T T_1(\sigma, \boldsymbol{\tau})^{-1}$:

$$v^T \left(x_i^{(k+1)} - x_1^{(k+1)}\right) = v^T T_1(\sigma, \boldsymbol{\tau})^{-1} g_{i-1}\zeta. \tag{4.24}$$

From (4.19) we know that the left-hand side is zero. The assumption $v^T T_1(\sigma, \tau)^{-1} g_{i-1} \neq 0$ implies that $\zeta = 0$. Since $T_1(\sigma, \boldsymbol{\tau})$ is nonsingular, (4.22) now leads to the conclusion that $x_1^{(k+1)} = x_i^{(k+1)}$. ∎

The result presented in Theorem 4.5 can again be used to reduce the computational cost of the residual inverse iteration method for solving NEPv problems. The structure exploiting method for finding symmetric solutions of (4.10) is given in Algorithm 4.2.

Algorithm 4.2 Symmetric residual inverse iteration for NEPv with m nonlinear terms

Start with $\sigma = \lambda^{(0)}, \boldsymbol{\tau} = \boldsymbol{\mu}^{(0)}, v$ and $x^{(0)}$ such that $v^T x^{(0)} = 1$.
Compute all $\psi_{\phi,i}^T = v^T T_i(\sigma, \boldsymbol{\tau})^{-1} \frac{\partial T_i}{\partial \phi}$ for $\phi \in (\lambda, \mu_1, \ldots, \mu_m)$ and $i \in (1, \ldots, m+1)$.
for k=0,1,2,... **do**
Compute $\gamma_i = v_i^T T_i(\sigma, \boldsymbol{\tau})^{-1} T_i(\lambda^{(k)}, \boldsymbol{\mu}^{(k)}) x_i^{(k)}$ for $i \in (1, \ldots, m+1)$.
Solve

$$\begin{bmatrix} \psi_{\lambda,1}^T x^{(k)} & \psi_{\mu_1,1}^T x^{(k)} & \cdots & \psi_{\mu_m,1}^T x^{(k)} \\ \psi_{\lambda,2}^T x^{(k)} & \psi_{\mu_1,2}^T x^{(k)} & \cdots & \psi_{\mu_m,2}^T x^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\lambda,m+1}^T x^{(k)} & \psi_{\mu_1,m+1}^T x^{(k)} & \cdots & \psi_{\mu_m,m+1}^T x^{(k)} \end{bmatrix} \begin{bmatrix} \Delta\lambda \\ \Delta\mu_1 \\ \vdots \\ \Delta\mu_m \end{bmatrix} = - \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{m+1} \end{bmatrix}.$$

Update $\lambda^{(k+1)} = \lambda^{(k)} + \Delta\lambda$ and $\mu_i^{(k+1)} = \mu_i^{(k)} + \Delta\mu_i$ for $i \in (1, \ldots, m+1)$.
Set $x^{(k+1)} = x^{(k)} - T_1(\sigma, \boldsymbol{\tau})^{-1} T_1(\lambda^{(k+1)}, \boldsymbol{\mu}^{(k+1)}) x^{(k)}$.
Normalize $x^{(k+1)} = x^{(k+1)}/(v^T x^{(k+1)})$.
end for

4.2. Inverse iteration methods for operator determinants. The method of inverse iteration for the standard eigenvalue problem is known to be globally convergent under very mild conditions, which is beneficial if no prior knowledge of the problem is available. The application of this method to the matrix pencil of operator determinants (Δ_1, Δ_0) allows for the extraction of nonlinear eigenpairs (λ, x). In the remainder of this section we discuss how this method can preserve symmetry in the eigenvector estimates while improving the convergence speed.

The goal is to find eigenpairs (λ, z) of the generalized eigenvalue problem

$$\Delta_1 z = \lambda \Delta_0 z \tag{4.25}$$

with $\Delta_0, \Delta_1 \in \mathbb{C}^{n^{m+1} \times n^{m+1}}$ the operator determinants derived from the NEPv (1.1) with m nonlinear terms and $z = x \otimes \cdots \otimes x \in \mathbb{C}^{n^{m+1}}$. Unless stated otherwise, we assume that the operator determinants Δ_0 and Δ_1 are nonsingular. In the general case,

this problem has n^{m+1} linearly independent eigenvectors, but, for the corresponding NEPv, only the $N_s = \binom{n+m}{m+1}$ symmetric eigenvectors are of interest.

A well known method to solve the eigenvalue problem is inverse iteration for which the iterates are dictated by

$$(\Delta_1 - \sigma\Delta_0)\,\tilde{z}_{k+1} = \Delta_0 z_k \tag{4.26a}$$

$$z_{k+1} = \frac{\tilde{z}_{k+1}}{\|\tilde{z}_{k+1}\|} \tag{4.26b}$$

with $\sigma \in \mathbb{C}$ an estimate of the desired eigenvalue such that $\Delta_1 - \sigma\Delta_0$ is nonsingular. Although this method will generally converge to the eigenvector corresponding to the eigenvalue closest to σ, it is possible to restrict convergence to symmetric eigenvectors only.

LEMMA 4.6. *If the pencil (Δ_1, Δ_0) is regular and non-defective, σ is not an eigenvalue and if z_k is symmetric and decomposable as $z_k = \sum_{j=1}^{l} \alpha_j x_j \otimes \cdots \otimes x_j$ for some index $l \leq N_s$ with x_j an eigenvector of the NEPv, then z_{k+1} in (4.26b) is symmetric too.*

Proof. If $z_k = \sum_{j=1}^{l} \alpha_j x_j \otimes \cdots \otimes x_j$ then it follows from inverse iteration property on eigenvectors that $z_{k+1} = \sum_{j=1}^{l} \frac{\alpha_j}{\lambda_j - \sigma} x_j \otimes \cdots \otimes x_j$ with λ_j the eigenvalue related to the eigenvector x_j. It is then clear that z_{k+1} is also symmetric. ∎

This lemma leads to a more powerful result as described in the following theorem.

THEOREM 4.7. *If the NEPv has N_s distinct solutions, the pencil (Δ_1, Δ_0) is regular and non-defective, σ is not an eigenvalue and if z_k is symmetric, then z_{k+1} is symmetric.*

Proof. If the NEPv has N_s solutions, Lemma 3.6 dictates that if the pencil (Δ_1, Δ_0) is regular and non-defective, there are N_s linearly independent eigenvectors spanning the complete space of symmetric vectors that is of dimension N_s. This means that every symmetric vector z_k can be decomposed as $z_k = \sum_{j=1}^{l} \alpha_j x_j \otimes \cdots \otimes x_j$ for some index $l \leq N_s$ and with x_j an eigenvector of the NEPv. By Lemma 4.6 the vector z_{k+1} is then symmetric too, proving the stated result. ∎

The previous theorem not only shows how symmetric eigenvectors can be extracted using inverse iteration, but also indicates how the convergence rate only depends on those eigenvectors. It is well known that, for general matrices, the convergence rate of inverse iteration is dictated by $\mathcal{O}\left(\left|\frac{\sigma-\lambda_{1*}}{\sigma-\lambda_{2*}}\right|^k\right)$ with λ_{1*} and λ_{2*} respectively the closest and second closest eigenvalue to σ. The following theorem shows how this rate can be specified for symmetric starting vectors.

THEOREM 4.8. *If the NEPv has N_s distinct solutions, the pencil (Δ_1, Δ_0) is regular and non-defective, σ is not an eigenvalue and if z_0 is symmetric, then the convergence rate of inverse iteration (4.26) is dictated by $\mathcal{O}\left(\left|\frac{\sigma-\lambda_1}{\sigma-\lambda_2}\right|^k\right)$ with λ_1 and λ_2 respectively the closest and second closest symmetric eigenvalue to σ, assuming that z_0 has a nonzero component in the corresponding eigenvectors.*

Proof. It follows from Lemma 4.6 and Theorem 4.7 that $z_{k+1} = \sum_{j=1}^{N_s} \frac{\alpha_j}{\lambda_j - \sigma} x_j \otimes \cdots \otimes x_j$ with the index j defined such that $|\lambda_1 - \sigma| \leq |\lambda_2 - \sigma| \leq \cdots \leq |\lambda_{N_s} - \sigma|$. This means that after each iteration the component of the first eigenvector grows by

a factor $\frac{1}{|\sigma-\lambda_1|}$, while the component of the second only grows by a factor $\frac{1}{|\sigma-\lambda_2|}$. The coefficient of $x_1 \otimes \cdots \otimes x_1$ will therefore grow faster than the coefficient of the eigenvector of the second closest symmetric eigenvalue by a factor of $\left|\frac{\sigma-\lambda_2}{\sigma-\lambda_1}\right|$ which leads to a convergence rate of $\mathcal{O}\left(\left|\frac{\sigma-\lambda_1}{\sigma-\lambda_2}\right|^k\right)$. ∎

As a result of this theorem, the convergence of inverse iteration is independent of the nonsymmetric solutions of the linearization if the starting vector is symmetric.

This section concludes with a discussion about the complexity of inverse iteration. If the nonlinear problem of dimension n has m nonlinear terms, each iteration requires solving the linear system given in (4.26) that is of dimension $n^{m+1} \times n^{m+1}$. A standard implementation of inverse iteration would therefore lead to a computational complexity of $\mathcal{O}(n^{3(m+1)})$. However, it is possible to reduce the computational cost by exploiting the structure of the operator determinants [23]. To this end, we first introduce the definition of the vec$(\cdot)$ operator.

DEFINITION 4.9. *The vectorization of a tensor is the bijective mapping that reorders the elements of the tensor $\mathcal{A} \in \mathbb{C}^{n_1 \times \cdots \times n_d}$ in a vector such that*

(4.27)
$$\text{vec}(\mathcal{A}) = [a_{1,1,\ldots,1} \;\; a_{2,1,\ldots,1} \;\; \cdots \;\; a_{n_1,1,\ldots,1} \;\; a_{1,2,1,\ldots,1} \;\; a_{2,2,1,\ldots,1} \;\; \cdots \;\; a_{n_1,n_2,\ldots,n_d}]^T.$$

Consider for simplicity the case where $m = 1$ for which (4.26) is rewritten as

$$\begin{aligned}\left(C \otimes (A + g_1 r_1^T + \sigma B) - (A + \sigma B) \otimes (C - g_1 s_1^T)\right) z_{k+1} \\ = \left(B \otimes (C - g_1 s_1^T) - C \otimes B\right) z_k.\end{aligned} \tag{4.28}$$

This equation is transformed to a Sylvester equation by using the "vec trick" [28]:

$$\begin{aligned}(A + g_1 r_1^T + \sigma B) Z_{k+1} C^T - (C - g_1 s_1^T) Z_{k+1} (A + \sigma B)^T \\ = (C - g_1 s_1^T) Z_k B^T - B Z_k C^T\end{aligned} \tag{4.29}$$

with $\text{vec}(Z_{k+1}) = z_{k+1}$ and $\text{vec}(Z_k) = z_k$. The computational cost reduces from $\mathcal{O}(n^6)$ for the standard implementation to $\mathcal{O}(n^3)$ for the Sylvester equation approach. This computational complexity is based on the Bartels-Stewart method [4] to solve the Sylvester equation. For $m > 1$, equation (4.26) can be transformed to a tensor Sylvester equation by using the following generalization of the vec trick:

$$(T_1 \otimes \cdots \otimes T_{m+1})\text{vec}(\mathcal{Y}) = \text{vec}(\mathcal{Y} \times_1 T_{m+1} \times_2 \cdots \times_{m+1} T_1) \tag{4.30}$$

where $\mathcal{Y}$ is a square tensor of dimension n and order $m+1$, $T_i \in \mathbb{C}^{n \times n}$ for $i = 1, \ldots, m+1$, and the operator $\times_i$ is the i-mode tensor product described by [16]. The numerical methods that solve these tensor Sylvester equations [5, 9, 14, 18, 29] are based on iterative schemes which makes it harder to quantify the computational cost reduction for the Sylvester approach for $m > 1$. A more thorough discussion of the existing methods falls outside of the scope of this paper.

4.3. Numerical experiments. This section shows the convergence properties of three iterative methods: residual inverse iteration (RI) as in [25], residual inverse iteration with symmetry exploitation (RIS) as in Algorithm 4.2 and inverse iteration (II) defined by (4.26). These methods all require initial estimates of the desired eigenpair (λ, x). RI and RIS require additional estimates of the remaining components of the eigenvalue $(\mu_1, \ldots, \mu_m)$ that are calculated using the eigenvector estimate $x^{(0)}$ as

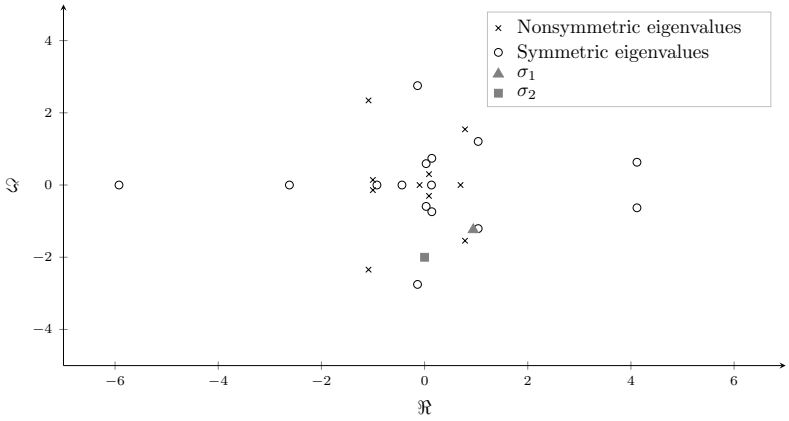


Figure 4.1: Eigenvalues of 2EP in Example 1.

$\tau_i = f_i(x^{(0)})$. Finally, we assume that the normalization vectors are $v = v_1 = v_2 = v_3$ and the starting vectors are $x^{(0)} = x_1^{(0)} = x_2^{(0)} = x_3^{(0)}$ for RI and RIS while the starting vector for II is $z_0 = x^{(0)} \otimes x^{(0)}$ for Example 1 and $z_0 = x^{(0)} \otimes x^{(0)} \otimes x^{(0)}$ for Example 2. The Matlab code that generates the results discussed below, can be found online [1].

4.3.1. Example 1. Consider the following 5×5 problem

$$(A + \lambda B + f(x)C)x = 0 \tag{4.31}$$

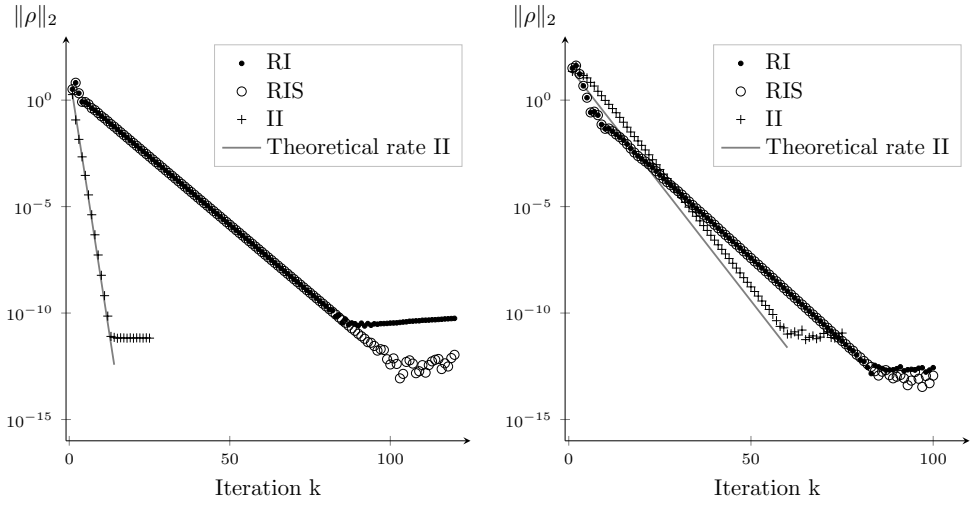
and the corresponding two parameter eigenvalue problem

$$\begin{cases} -Ax_1 = (\lambda B + \mu C)x_1 \\ -(A + gr^T)x_2 = (\lambda B + \mu(C - gs^T))x_2. \end{cases} \tag{4.32}$$

Due to the small problem size, it is feasible to solve the MEP using the explicit operator determinant formulation $\Delta_1 z = \lambda \Delta_0 z$ for which the eigenvalues are given in Figure 4.1. There are 15 eigenvalues related to symmetric eigenvectors and 10 eigenvalues related to nonsymmetric eigenvectors, which is to be expected for a 5×5 nonlinear problem. Using the results from the operator determinant eigenvalues, we can now show the convergence properties of every method.

The first test assumes a priori knowledge about the desired eigenpair: $\sigma = 0.9440 - 1.2382\imath$ is close to a desired eigenvalue and $x^{(0)}$ is calculated as the corresponding eigenvector with a random perturbation of norm 10^{-1}. The starting value σ is indicated in Figure 4.1 as a triangle. Figure 4.2a shows the residual, ρ, of the NEPv and shows the convergence behaviour of the three methods. It is clear that RI and RIS are mathematically equivalent by looking at the first part of the convergence, but the last part shows numerical rounding errors that break the symmetry in RI which causes it to diverge from RIS. The convergence plot for II shows that it stagnates around 10^{-11}, which can be explained by looking at the condition number of Δ_0 and Δ_1 which is $\mathcal{O}(10^4)$. This means that (4.26) cannot be solved with a smaller residual in double precision. Lastly, note that II follows the symmetric convergence speed of $\mathcal{O}\left(0.11^k\right)$ instead of the nonsymmetric estimate $\mathcal{O}\left(0.29^k\right)$.

[1] www.github.com/robclaes/nepv-mep



(a) $\sigma = 0.9440 - 1.2382\imath$ (b) $\sigma = -2\imath$

Figure 4.2: Residual of NEPv of Example 1.

The second test for this example assumes no prior knowledge and aims to find the eigenvalue closest to $\sigma = -2\imath$, depicted as a square in Figure 4.1. All three methods start with the same randomly generated vector $x^{(0)}$, which leads to the convergence result as given in Figure 4.2b. The method of II follows the predicted convergence rate and stagnates at the same level as in the first test. RI and RIS converge linearly to a solution, but closer inspection shows that they converged to $\lambda \approx 0.14 - 0.74\imath$ which is not the eigenvalue nearest to σ. We observe linear convergence for all tested methods even if RI and RIS are only guaranteed to converge locally.

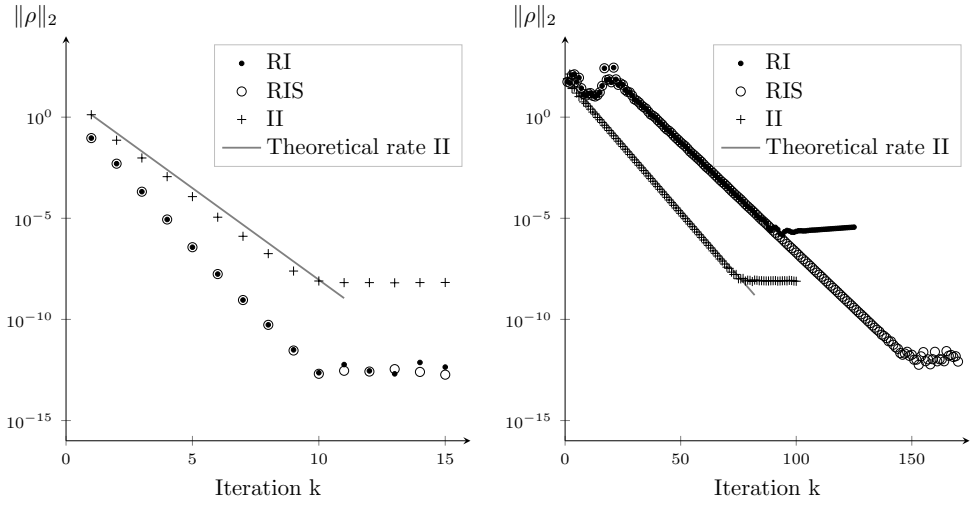
4.3.2. Example 2. Consider the following NEPv with two nonlinear terms

$$\begin{aligned} &(A + \lambda B + f_1(x)C_1 + f_2(x)C_2))\, x = 0 \\ &f_1(x) = \frac{r_1^T x}{s_1^T x}, f_2(x) = \frac{r_2^T x}{s_2^T x} \end{aligned} \tag{4.33}$$

which is linearized as the following 3-parameter eigenvalue problem

$$\begin{cases} -Ax_1 = (\lambda B + \mu_1 C_1 + \mu_2 C_2)x_1 \\ -(A + g_1 r_1^T)x_2 = (\lambda B + \mu_1(C_1 - g_1 s_1^T) + \mu_2 C_2)x_2 \\ -(A + g_2 r_2^T)x_3 = (\lambda B + \mu_1 C_1 + \mu_2(C_2 - g_2 s_2^T))x_3. \end{cases} \tag{4.34}$$

In this example, the matrices $A, B, C_1, C_2 \in \mathbb{R}^{10\times 10}$ and vectors $r_1, r_2, s_1, s_2, g_1, g_2 \in \mathbb{R}^{10}$ are all randomly generated. The size of the NEPv is rather small such that the equivalent generalized eigenvalue problem is still feasible to solve. This way, a generic eigenproblem solver can be used to find all symmetric eigenpairs and to compare them to the result from the iterative methods. Similarly as in Example 1, we present two situations: the first will assume that prior knowledge of the eigenpair is available, the second situation will assume no knowledge of the solution.



(a) $\sigma = -1.2769 - 0.0442\imath$ (b) $\sigma = -4 - 3\imath$

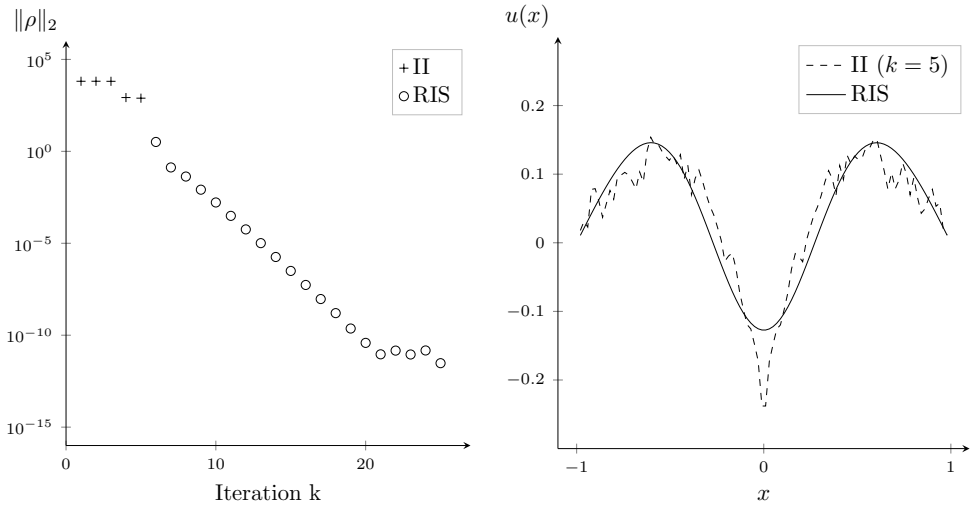
Figure 4.3: Residual of NEPv of Example 2.

First consider the case where prior knowledge of the problem is available to start the iteration. We use the result from the operator determinant eigenvalue problem to pick an eigenvalue $\lambda \approx -1.2830 - 0.0363\imath$ and its corresponding normalized eigenvector x. The starting values for the iterative methods are a perturbation of this eigenpair $\sigma = -1.2769 - 0.0442\imath$ and $x^{(0)} = x + e$ with e a randomly generated vector of norm 10^{-2}. The starting value for the eigenvector is used to determine the values of $\mu_1^{(0)} = f_1(x^{(0)})$ and $\mu_2^{(0)} = f_2(x^{(0)})$. Figure 4.3a shows the residual as a function of the number of iterations. As expected, all three methods converge to the closest eigenpair within a small number of iterations. The method of inverse iterations follows the symmetric convergence rate of $\mathcal{O}(0.12^k)$ instead of the standard rate $\mathcal{O}(0.22^k)$. Again, note that II stagnates at the level around 10^{-9} which can be explained by the same reasoning as in Example 1 since the condition number of Δ_0 and Δ_1 is $\mathcal{O}(10^7)$.

The second part of this example considers the situation in which no prior knowledge is available and we initialize with a target $\sigma = -4 - 3\imath$ and a random starting vector while the initial values for $\mu_1^{(0)}$ and $\mu_2^{(0)}$ are calculated in the same way as in the previous part of this example. The residual norms are plotted in Figure 4.3b. Again, RI and RIS show the same convergence behaviour but, between iterations 80 and 90, RI diverges from RIS again due to numerical rounding errors. However, since RIS did not start sufficiently close to an eigenpair, it converged to the eigenvalue $\lambda \approx -2.6086 + 0.2827\imath$ and not to the closest eigenvalue $\lambda \approx -3.9910 - 2.2599\imath$. The method of inverse iteration did converge, with limited accuracy, to the closest eigenpair and followed the theoretical symmetric convergence rate of $\mathcal{O}(0.74^k)$ which is better than the standard convergence rate of $\mathcal{O}(0.94^k)$.

4.3.3. Example 3. Consider the following differential equation in one spatial dimension

$$u''(x) + \lambda k_1(x)u(x) + f(u)k_2(x)u(x) = 0 \tag{4.35}$$



(a) Residual

(b) Calculated solution

Figure 4.4: Convergence and solution for Example 3.

defined for the domain $x \in [-1, 1]$ with $u(-1) = u(1) = 0$. The function $f(u) = \frac{\alpha(u)}{\beta(u)}$ is defined by $\alpha(u) = \int_{-1}^{1} g(x)u(x)dx$ and $\beta(u) = u'(0)$ with $g(x) = e^{-\gamma x^2}$. For this example, we set $k_1(x) = 1 + \frac{1}{2}\tanh(5x)$, $k_2(x) = 1 + \frac{1}{2}\cos(\pi x)$ and $\gamma = 10$. The solutions of this differential equation can be interpreted as vibrations in a medium with a damping coefficient that depends on both the spatial location and the shape of the vibrations itself. By spatial discretization, the equation can be written as a NEPv:

$$Av + \lambda K_1 v + \frac{a^T v}{b^T v} K_2 v = 0 \tag{4.36}$$

with v the discretization of $u(x)$ and where the derivatives are discretized by finite differences and the integral by the trapezoidal rule. The NEPv (4.36) is of the general form (1.1)-(1.2) and can be linearized to a MEP.

Using 100 discretization points, we will show how inverse iteration and residual inverse iteration can go hand in hand to find desired solutions in a reliable and fast way. From the theory and the first two examples, we know that inverse iteration is a very robust, but more expensive, method while residual inverse iteration requires better starting values but is cheaper. Therefore, we use a hybrid approach for this example: start with a small number of iterations of inverse iteration to improve the starting values and continue with residual inverse iteration until convergence. The result is presented in Figure 4.4 where we use $\sigma = 1$ and a random symmetric starting vector for II. The intermediate approximation after five iterations of II is used as a starting value for RIS. The approximation of $u(x)$ after II is shown together with the final result after RIS in Figure 4.4b. These results show how the proposed methods can be complementary in practical applications: II reliably finds a rough estimate of the desired solution, and RIS cheaply converges to the solution with small residual.

5. Discussion. We introduced a new way of analyzing the NEPv by the concept of linearization for eigenvector nonlinearities. The linearization is derived for a subset of nonlinear problems and introduces some additional parameters that can be chosen freely. We show how to linearize problems of this class and characterize the number of solutions. We also derive conditions that guide us on how to choose the free parameters. Contrary to existing methods for NEPv, the proposed linearization is able to identify any and all solutions, under mild assumptions.

The new linearization technique also allows for the use and adaption of well-understood state-of-the-art methods. We proposed two structure-exploiting numerical schemes: residual inverse iteration and inverse iteration. Residual inverse iteration requires less calculations per iteration but needs accurate starting values to obtain predictable convergence. Inverse iteration is more expensive but is in general globally convergent to the desired solution. Three numerical examples confirmed these theoretical results and showed that the two methods are complementary: a small number of iterations of inverse iteration lead to a good starting value that can be used for residual inverse iteration. A thorough discussion of the implementation of the numerical methods falls outside the scope of this paper, but it is an interesting point for future research.

Acknowledgements. We thank Bor Plestenjak (Univ. Ljubljana) for explanations concerning general MEPs, Lars Karlsson (Univ. Umeå) for comments concerning GEP pencil singularity, and Simon Telen (Max Planck Institute, Leipzig) for insights in the multi-homogenous Bézout theorem.

The work by Rob Claes and Karl Meerbergen is supported by the Research Foundation – Flanders (FWO) Grant G0B7818N and the KU Leuven Research Council.

REFERENCES

[1] F. V. Atkinson, *Multiparameter Eigenvalue Problems. Volume I: Matrices and Compact Operators*, Academic Press, New York, 1972.

[2] Z. Bai, D. Lu, and B. Vandereycken, *Robust Rayleigh quotient minimization and nonlinear eigenvalue problems*, SIAM J. Sci. Comput., 40 (2018), pp. A3495–A3522.

[3] W. Bao and Q. Du, *Computing the ground state solution of Bose-Einstein condensates by a normalized gradient flow*, SIAM J. Sci. Comput., 25 (2004), pp. 1674–1697.

[4] R. H. Bartels and G. W. Stewart, *Solution of the matrix equation $AX + XB = C$*, Comm. ACM., 15 (1972), pp. 820–826.

[5] F. P. A. Beik, F. S. Movahed, and S. Ahmadi-Asl, *On the Krylov subspace methods based on tensor format for positive definite Sylvester tensor equations*, Numer. Linear Algebra Appl., 23 (2016), pp. 444–466.

[6] T. Bühler and M. Hein, *Spectral clustering based on the graph p-Laplacian*, in Proc. 26th Int. Conf. Mach. Learn., 2009, pp. 81–88.

[7] Y. Cai, L. H. Zhang, Z. Bai, and R. C. Li, *On an eigenvector-dependent nonlinear eigenvalue problem*, SIAM J. Matrix Anal. Appl., 39 (2018), pp. 1360–1382.

[8] E. Cancès and C. L. Bris, *On the convergence of SCF algorithms for the Hartree-Fock equations*, M2AN, Math. Model. Numer. Anal., (2000).

[9] Z. Chen and L. Lu, *A projection method and Kronecker product preconditioner for solving Sylvester tensor equations*, Sci. China Math., 55 (2012), pp. 1281–1292.

[10] M. Hein and T. Bühler, *An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse PCA*, in Advances in Neural Information Processing Systems 23, 2010, pp. 847–855.

[11] M. E. Hochstenbach, T. Košir, and B. Plestenjak, *A Jacobi–Davidson type method for the two-parameter eigenvalue problem*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 477–497.

[12] M. E. Hochstenbach, K. Meerbergen, E. Mengi, and B. Plestenjak, *Subspace methods for three-parameter eigenvalue problems*, Numer. Linear Algebra Appl., 26 (2019), p. e2240.

[13] M. E. Hochstenbach and B. Plestenjak, *A Jacobi–Davidson type method for a right definite two-parameter eigenvalue problem*, SIAM J. Matrix Anal. Appl., 24 (2002), pp. 392–410.

[14] B. Huang and C. Ma, *An iterative algorithm to solve the generalized Sylvester tensor equations*, Linear Multilinear Algebra., 68 (2020), pp. 1175–1200.
[15] E. Jarlebring, S. Kvaal, and W. Michiels, *An inverse iteration method for eigenvalue problems with eigenvector nonlinearities*, SIAM J. Sci. Comput., 36 (2014), pp. A1978–A2001.
[16] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
[17] T. Košir, *Finite-dimensional multiparameter spectral theory: The nonderogatory case*, Linear Algebra Appl., 212 (1994), pp. 45–70.
[18] D. Kressner and C. Tobler, *Krylov subspace methods for linear systems with tensor product structure*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 1688–1714.
[19] A. Levitt, *Convergence of gradient-based algorithms for the Hartree-Fock equations*, ESAIM. M2AN, 46 (2012), pp. 1321–1336.
[20] X. Liu, X. Wang, Z. Wen, and Y. Yuan, *On the convergence of the self-consistent field iteration in Kohn–Sham density functional theory*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 546–558.
[21] X. Liu, Z. Wen, X. Wang, M. Ulbrich, and Y. Yuan, *On the analysis of the discretized Kohn–Sham density functional theory*, SIAM J. Numer. Anal., 53 (2015), pp. 1758–1785.
[22] D. S. Mackey, N. Mackey, C. Mehl, and V. Mehrmann, *Structured polynomial eigenvalue problems: Good vibrations from good linearizations*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 1029–1051.
[23] K. Meerbergen and B. Plestenjak, *A Sylvester–Arnoldi type method for the generalized eigenvalue problem with two-by-two operator determinants*, Numer. Linear Algebra Appl., 22 (2015), pp. 1131–1146.
[24] A. Neumaier, *Residual inverse iteration for the nonlinear eigenvalue problem*, SIAM J. Numer. Anal., 22 (1985), pp. 914–923.
[25] B. Plestenjak, *Numerical methods for nonlinear two-parameter eigenvalue problems*, BIT, 56 (2016), pp. 241–262.
[26] P. Pulay, *Convergence acceleration of iterative sequences. The case of SCF iteration*, Chem. Phys. Lett., 73 (1980), pp. 393–398.
[27] T. Rohwedder and R. Schneider, *An analysis for the DIIS acceleration method used in quantum chemistry calculations*, J. Math. Chem., 49 (2011), pp. 1889–1914.
[28] W. E. Roth, *On direct product matrices*, Bullet. Am. Math. Soc., 40 (1934), pp. 461–468.
[29] K. Ruymbeek, K. Meerbergen, and W. Michiels, *Subspace method for multiparameter-eigenvalue problems based on tensor-train representations*, 2020. arxiv 2012.00815.
[30] Y. Saad, J. T. Chelikowsky, and S. M. Shontz, *Numerical methods for electronic structure calculations of materials*, SIAM Rev., 52 (2010), pp. 3–54.
[31] I. R. Shafarevich, *Basic Algebraic Geometry 1*, Springer-Verlag, 2013.
[32] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, 1996.
[33] F. Tudisco and D. J. Higham, *A nonlinear spectral method for core–periphery detection in networks*, SIAM J. Math. Data. Sci., 1 (2019), pp. 269–292.
[34] P. Upadhyaya, E. Jarlebring, and E. H. Rubensson, *A density matrix approach to the convergence of the self-consistent field iteration*, Numer. Algebra Control Optim., (2019).
[35] C. Yang, W. Gao, and J. C. Meza, *On the convergence of the self-consistent field iteration for a class of nonlinear eigenvalue problems*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1773–1788.