Automatic rational approximation and linearization of nonlinear eigenvalue problems

Pieter Lietaert †, Javier Pérez ‡, Bart Vandereycken § and Karl Meerbergen †

We present a method for solving nonlinear eigenvalue problems using rational approximation. The method uses the AAA method by Nakatsukasa, Sète, and Trefethen to approximate the nonlinear eigenvalue problem by a rational eigenvalue problem and is embedded in the state space representation of a rational polynomial by Su and Bai. The advantage of the method, compared to related techniques such as NLEIGS and infinite Arnoldi, is the efficient computation by an automatic procedure. In addition, a set-valued approach is developed that allows building a low degree rational approximation of a nonlinear eigenvalue problem. The method perfectly fits the framework of the Compact rational Krylov methods (CORK and TS-CORK), allowing to efficiently solve large scale nonlinear eigenvalue problems. Numerical examples show that the presented framework is competitive with NLEIGS and usually produces smaller linearizations with the same accuracy but with less effort for the user.

Keywords: Nonlinear eigenvalue problem, Rational interpolation, Rational Krylov method

1. Introduction

The nonlinear eigenvalue problem (NEP) is the problem of finding scalars \( \lambda \in \mathbb{C} \) and nonzero vectors \( x, y \in \mathbb{C}^n \) such that

\[
A(\lambda)x = 0 \quad \text{and} \quad y^*A(\lambda) = 0,
\]

where \( A : \mathbb{C} \to \mathbb{C}^{n \times n} \) is a nonlinear matrix valued function. The scalar \( \lambda \in \mathbb{C} \) is called an eigenvalue and the vectors \( x \) and \( y \) are called, respectively, associated right and left eigenvectors. Usually, one is interested in the eigenvalues in a specific region \( \Sigma \subset \mathbb{C} \). We assume that \( A \) in (1.1) is regular, i.e., there is at least one \( \lambda \in \mathbb{C} \) for which \( \det(A(\lambda)) \neq 0 \).

Numerical methods for computing the eigenvalues of generic nonlinear eigenvalue problems in a region \( \Sigma \) are based on approximation theory. There currently are two classes of methods: those based on contour integrals (Beyn, 2012), and those based on rational and polynomial approximation, e.g. Effenberger & Kressner (2012), infinite Arnoldi (Jarlebring et al., 2012) and NLEIGS (Güttel et al., 2014). The methods based on contour integration rely on Keldysh theorem, where the eigenvalues in \( \Sigma \) are found as the poles of a resolvent using contour integration. There are several variations of this approach, see Beyn (2012). For large scale problems, the Jacobi–Davidson method is combined with contour integration in Effenberger (2013) for computing invariant pairs of (1.1). The second class of methods approximates \( A \) by a polynomial or rational function on \( \Sigma \) and solves the resulting polynomial

---

†Department of Computer Science, KU Leuven, University of Leuven, 3001 Heverlee, Belgium. Supported by KU Leuven Research Council Grant OT/14/074 and BELSPO Grant IAP VII/19.
‡Department of Mathematical Sciences, University of Montana, MT, USA. The research of J. Perez was partially supported by KU Leuven Research Council grant OT/14/074 and the Interuniversity Attraction Pole DYSO, initiated by the Belgian State Science Policy Office.
§Section of Mathematics, University of Geneva, Rue du Lièvre 2-4, 1211 Geneva, Switzerland
or rational eigenvalue problem. Polynomial and rational eigenvalue problems can be solved by Krylov methods through a linearization. The prototype linearization is the companion pencil.

In this paper, we focus on methods that build a rational approximation, reformulate the resulting problem as a linear eigenvalue problem (by the process of linearization) and then use a Krylov method for solving the linearized problem. Roughly speaking, there exist three approaches for rational approximation. The NLEIGS method uses potential theory for the selection of poles and interpolation points, and embeds this within a rational polynomial expressed in rational Newton basis (see Güttel et al., 2014). The second is the infinite Arnoldi method (Jarlebring et al., 2012) that uses the discretization of an infinite dimensional operator that is a linear representation of the nonlinear eigenvalue problem. The discretization of this operator leads to a finite dimensional linear problem that is solved by the Arnoldi method. The third approach expresses a Padé approximation in state-space form and applies a Krylov method to a linearization, see Su & Bai (2011).

The approach of this paper is inspired by Su & Bai (2011) and assumes that the matrix valued function $A$ can be written as

$$A(\lambda) = P(\lambda) + G(\lambda),$$  \hspace{1cm} (1.2)

where $P(\lambda)$ is a matrix polynomial and $G$ is an arbitrary matrix valued function of the form

$$G(\lambda) = \sum_{i=1}^{s} (C_i - \lambda D_i)g_i(\lambda),$$  \hspace{1cm} (1.3)

where $C_i,D_i$ are constant $n \times n$ matrices and $g_i: \mathbb{C} \rightarrow \mathbb{C}$ is a nonlinear function. Using the approach of Su & Bai (2011), each $g_i$ can be approximated by a different rational function with different poles and interpolation points, determined independently from each other. This in contrast to NLEIGS (Güttel et al., 2014), where the same poles and nodes are used for all terms. For efficiency, it is assumed that the number of nonlinear terms, $s$, is modest.

The contribution of this paper is threefold. First, the rational approximations used in this work are obtained by employing the adaptive Antoulas–Anderson (AAA) algorithm introduced by Nakatsukasa, Sète, and Trefethen in Nakatsukasa et al. (2016). This approach presents two key advantages: the AAA algorithm is not domain-dependent, i.e., it works effectively even with sets that may include disconnected regions of irregular shape, possibly unbounded; and once the approximation region has been fixed, it is the algorithm and not the user who chooses the number of poles and zeros of the rational approximation and their values in an adaptive way. Hence, unlike other methods, neither special knowledge of the nonlinear functions nor advance knowledge of complex analysis is required from the user.

The second contribution is an automatic strategy that uses the same poles and interpolation points for (a subset of) all $g_i$, which leads to the same appealing properties as NLEIGS, when $s$ is not so small. Numerical experiments compare with rational approximations obtained using NLEIGS in Güttel et al. (2014).

The third contribution of the paper lies in the observation that the linearization from Su & Bai (2011) fits perfectly in the framework of the CORK (Compact Rational Krylov) method in Van Beeumen et al. (2015) and the two-sided CORK method in Lietaert et al. (2017), which makes the linearization suitable for large scale problems and rational functions of high degree. There is no need to require $C_i,D_i$ to be of low rank but if they are it can be exploited.

The rest of the paper is structured as follows. Section 2 presents the original AAA approximation of a nonlinear function $g$ and its generalization to a set of nonlinear functions $g_1, \ldots, g_s$. In §3, we reformulate the linearization by Su & Bai (2011) as a CORK linearization and show a relation between
the eigenvalues and eigenvector of the linearization and the rational matrix polynomial, including the case of low rank \( C_i - \lambda D_i, i = 1, \ldots, s \). Section 4 illustrates the AAA approaches by solving nonlinear eigenvalue problems and compares with NLEIGS. Section 5 is reserved for the conclusions.

2. Scalar rational approximations by AAA

As explained in the introduction, we intend to use rational approximations for the NEP. As a first step, this requires approximating the scalar functions \( g(\lambda) \) in (1.3), either separately or together, by rational functions. Our approach is based on a recently introduced algorithm from Nakatsukasa et al. (2016) that we review next.

2.1 The AAA algorithm

Let \( g: \mathbb{C} \to \mathbb{C} \) denote a generic nonlinear function that we would like to approximate on \( \Sigma \subset \mathbb{C} \) by a rational function \( r(\lambda) \). The adaptive Antoulas–Anderson (AAA) algorithm from Nakatsukasa et al. (2016), will construct this function \( r(\lambda) \) in barycentric form:

\[
r(\lambda) = \frac{\sum_{j=1}^{m} g(z_j) \omega_j}{\sum_{j=1}^{m} \omega_j}, \quad \omega_j = \frac{\lambda - z_j}{d_m(\lambda)},
\]

(2.1)

Here, \( z_1, \ldots, z_m \) are a set of distinct support points and \( \omega_1, \ldots, \omega_m \) are the weights. Note that, as long as \( \omega_j \neq 0 \), \( \lim_{\lambda \to z_j} r(\lambda) = g(z_j) \). In other words, the rational function (2.1) interpolates the function \( g(\lambda) \) at \( z_1, \ldots, z_m \).

The AAA algorithm computes the support points and the weights iteratively by minimizing the linearized residual of the rational approximation on a sample set \( Z \) of \( M \) points. The set \( Z \) can be seen as a sufficiently fine discretization of the region \( \Sigma \) which means \( M \) is typically quite large, say, \( 10^4 \). At the \( m \)th step of the algorithm, the next support point \( z_m \) is chosen where the residual \( g(\lambda) - n_{m-1}(\lambda)/d_{m-1}(\lambda) \) attains its maximum absolute value. Then, denoting

\[
Z^{(m)} = \{Z_1^{(m)}, \ldots, Z_{M_m}^{(m)}\} := Z/\{z_1, \ldots, z_m\}
\]

and

\[
G^{(m)} = \{G_1^{(m)}, \ldots, G_{M_m}^{(m)}\} := g(Z^{(m)}),
\]

it computes the vector of weights \( \omega = [\omega_1 \ldots \omega_m]^T \) with \( \|\omega\|_2 = 1 \) that minimizes the 2-norm of the linearized residual

\[
\begin{bmatrix}
g(Z_1^{(m)})d_m(Z_1^{(m)}) - n_m(Z_1^{(m)}) \\
g(Z_2^{(m)})d_m(Z_2^{(m)}) - n_m(Z_2^{(m)}) \\
\vdots \\
g(Z_{M_m}^{(m)})d_m(Z_{M_m}^{(m)}) - n_m(Z_{M_m}^{(m)})
\end{bmatrix} =
\begin{bmatrix}
G_1^{(m)} - g(z_1) & \cdots & \frac{G_1^{(m)} - g(z_m)}{Z_1^{(m)} - z_1} \\
\vdots & \ddots & \vdots \\
\frac{G_{M_m}^{(m)} - g(z_1)}{Z_{M_m}^{(m)} - z_1} & \cdots & \frac{G_{M_m}^{(m)} - g(z_m)}{Z_{M_m}^{(m)} - z_m}
\end{bmatrix}
\begin{bmatrix}
\omega_1 \\
\vdots \\
\omega_m
\end{bmatrix}.
\]

(2.2)

This can be done by using the SVD on the Loewner matrix above. The procedure terminates when the norm of the residual (2.2) is less than a user defined tolerance, for example, \( 10^{-13} \). For further details of the AAA algorithm, including the removal of numerical Froissart doublets, we refer to Nakatsukasa et al. (2016) and our (modified) implementation in App. 5.
A key feature of AAA is its flexibility in selecting the domain of approximation (through the set $Z$), unlike other methods that are domain-dependent. Furthermore, the user only needs to supply this domain and a tolerance. Then, the poles and zeros of the rational interpolant (2.1) are found automatically by the algorithm. In many cases, AAA succeeds in computing a rational interpolant that is close to the optimal one in min-max sense. However, the algorithm can fail on difficult functions; see Filip et al. (2017) for examples. In the numerical experiments, however, we did not see such pathological behavior and AAA performed adequately.

In Section 3, it will be convenient to write the rational functions from AAA that are in barycentric form into an equivalent state-space form.

**Proposition 2.1** The rational function (2.1) can be written as

$$r(\lambda) = \begin{bmatrix} g(z_1) \omega_1 & \cdots & g(z_m) \omega_m \end{bmatrix} \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_{m-1} & \omega_m \cr \lambda - z_1 & z_2 - \lambda & \cdots & \lambda - z_{m-2} & \lambda - z_{m-1} \cr \lambda - z_2 & \cdots & \lambda - z_{m-2} & \lambda - z_{m-1} & \lambda - z_m \cr \cdots & \cdots & \cdots & \cdots & \cdots \cr \lambda - z_{m-1} & \cdots & \lambda - z_{m-2} & \lambda - z_{m-1} & \lambda - z_m \end{bmatrix}^{-1} \begin{bmatrix} 1 \cr 0 \cr \vdots \cr 0 \end{bmatrix},$$

(2.3)

where the entries that are not depicted are equal to zero.

**Proof.** Let $d(\lambda) = \sum_{j=1}^m \omega_j (\lambda - z_j)^{-1}$ and $n(\lambda) = \sum_{j=1}^m g(z_j) \omega_j (\lambda - z_j)^{-1}$ denote, respectively, the denominator and numerator of $r(\lambda)$ in (3.4). Then, it is easily verified that the vector

$$\frac{1}{d(\lambda)} \begin{bmatrix} (\lambda - z_1)^{-1} \cr \vdots \cr (\lambda - z_m)^{-1} \end{bmatrix}$$

is the first column of

$$\begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_{m-1} & \omega_m \cr \lambda - z_1 & z_2 - \lambda & \cdots & \lambda - z_{m-2} & \lambda - z_{m-1} \cr \lambda - z_2 & \cdots & \lambda - z_{m-2} & \lambda - z_{m-1} & \lambda - z_m \cr \cdots & \cdots & \cdots & \cdots & \cdots \cr \lambda - z_{m-1} & \cdots & \lambda - z_{m-2} & \lambda - z_{m-1} & \lambda - z_m \end{bmatrix}^{-1}.$$

Thus, we obtain the desired result as

$$\frac{1}{d(\lambda)} \begin{bmatrix} g(z_1) \omega_1 & \cdots & g(z_m) \omega_m \end{bmatrix} \begin{bmatrix} (\lambda - z_1)^{-1} \cr \vdots \cr (\lambda - z_m)^{-1} \end{bmatrix} = \frac{n(\lambda)}{d(\lambda)} = r(\lambda).$$

\[
2.2 \quad A \text{ set-valued AAA algorithm}
\]

There are applications where a large number of nonlinear functions $g_1, \ldots, g_s$ need to be approximated over the same region of the complex plane. One can of course use the AAA algorithm on each function
separately but, as we will see in the numerical examples in §4, it is sometimes beneficial to find support points and poles that work for all the functions at the same time. The result is then a linearization with a smaller total degree compared to the linearization obtained from the separate applications of AAA. In this section, we show how to extend the AAA approach to accomplish this.

Let \( g_1, \ldots, g_s \) be nonlinear functions of the same scale, which can be accomplished by simply scaling them as \( g_i(\lambda)/\max_j |g_i(z_j)| \). Our aim is to construct rational approximations to \( g_1, \ldots, g_s \) of the form

\[
g_i(\lambda) \approx \frac{\sum_{j=1}^{m} g_i(z_j) \omega_j}{\sum_{j=1}^{m} \lambda - z_j} =: n_{i,m}(\lambda) / d_{i,m}(\lambda)
\]

Note that all rational approximants share the same support points \( z_j \) and weights \( \omega_j \). In the spirit of the AAA algorithm, these support points and weights are computed iteratively. At the \( m \)-th step, the next support point \( z_m \) is chosen where the maximum of the residuals, i.e.,

\[
\max_i \left| g_i(\lambda) - \frac{n_{i,m-1}(\lambda)}{d_{i,m-1}(\lambda)} \right|
\]

attains its maximum. Then, denoting

\[
Z^{(m)} = \{ Z_1^{(m)} \ldots, Z_m^{(m)} \} := Z / \{ z_1, \ldots, z_m \} \quad \text{and} \quad G_i^{(m)} = \{ G_i 1^{(m)}, \ldots, G_i m^{(m)} \} := g_i(Z^{(m)}),
\]

the residual vector (2.2) can be written to incorporate different functions:

\[
\begin{bmatrix}
G_1^{(m)} - g_1(z_1) & \cdots & G_1^{(m)} - g_1(z_m) \\
G_2^{(m)} - g_2(z_1) & \cdots & G_2^{(m)} - g_2(z_m) \\
\vdots & \ddots & \vdots \\
G_s^{(m)} - g_s(z_1) & \cdots & G_s^{(m)} - g_s(z_m)
\end{bmatrix}
\begin{bmatrix}
\omega_1 \\
\omega_2 \\
\vdots \\
\omega_m
\end{bmatrix}
\]

The vector of weights \( \omega = [\omega_1 \cdots \omega_m]^T \) is computed as the vector minimizing the norm of (2.5) under the constraint \( \| \omega \|_2 = 1 \). We note that minimizing this norm is equivalent to minimizing the sum of squares of the norms of the residual vectors of the different functions.
The computational cost of the AAA algorithm and of our modified set-valued version might become an issue when a large number of nonlinear functions need to be approximated. To partly alleviate this extra cost, we can slightly reformulate how the AAA algorithm solves the least squares problems in (2.2) and (2.5) to compute the vector of weights. This is outlined in the next section.

2.3 Efficient solution of the least squares problem in AAA

The original AAA algorithm requires the repeated singular value decomposition of the tall but skinny Loewner matrix in (2.2) to find its right singular vector corresponding to the smallest singular value. This can become costly for large matrices, which is certainly the case if we have do this for a large number of functions as in (2.5). Fortunately, we can exploit that the Loewner matrices in (2.2) or in (2.5) differ only in a few rows and columns throughout each iteration of the AAA algorithm. In particular, by computing economy-size QR decompositions of said matrices, we only need to obtain the right singular vectors of much smaller matrices. In turn, these QR decompositions can be computed using updating strategies.

Let \( L_m \) be the Loewner matrix in (2.2) or (2.5), of size \( n \times m \). We recall that \( n \gg m \). The matrix \( L_m \) can be stored as

\[
L_m = QH,
\]

where \( Q \) is an \( n \times m \) matrix with orthonormal columns, and \( H \) is an \( m \times m \) matrix. Note that the right singular vectors of the matrix \( L_m \) can be computed as the right singular vectors of the small matrix \( H \).

The column of the matrix \( Q \) can be iteratively found by adding one column of the Loewner matrix \( L_m \) in each step and applying Gram-Schmidt orthogonalization. Note, however, that in each step of AAA, the number of rows of \( Q \) is reduced by one for (2.2) and by \( s \) for (2.5) because the support points \( z_i \) are removed from the set \( Z \) that defines the residual. With the removal of these rows, \( Q \) is no longer orthogonal. However, we can reorthogonalize \( Q \) cheaply as follows. Let \( Q_r \in \mathbb{C}^{r \times m} \) be the matrix whose rows are the rows that have been removed from \( Q \) in step \( m \) and let \( \tilde{Q} \) be the matrix obtained from \( Q \) after the removal of these rows. We then have, since \( Q \) is orthogonal,

\[
\tilde{Q}^* \tilde{Q} = I_m - Q_r^* Q_r.
\]

So, by taking the Cholesky decomposition,

\[
I_m - Q_r^* Q_r = S^* S,
\]

we have that matrix \( \tilde{Q} S^{-1} \) is orthogonal, and we update \( H = S H \). We can further avoid the (costly) explicit multiplication \( \tilde{Q} S^{-1} \) by storing \( S \) in matrix

\[
S_m = \begin{bmatrix} S_{m-1} & \cdot \cdot \cdot & \cdot \cdot \cdot & 1 \end{bmatrix} S^{-1},
\]

which is only used in matrix-vector and matrix-matrix multiplications with vector and matrices of size \( O(m) \), that is, of small size.

This procedure is implemented in our MATLAB version of the AAA algorithm in App. 5, which also shows how to rework the AAA algorithm to incorporate multiple functions, that is, the set-valued AAA algorithm. It should be compared with the algorithm presented in Nakatsukasa et al. (2016). The main cost of the algorithm is reduced to the Gram-Schmidt orthogonalization process of the long vectors of the Loewner matrix.
3. Rational approximations for NEPs using AAA

In this section we show how the scalar rational functions, computed by AAA, can be used efficiently to obtain a rational approximation of the NEP. In particular, we will present linearizations that build on the CORK (Van Beeumen et al., 2015) and the TS-CORK (Lietaert et al., 2017) frameworks and exploit possible low-rank terms. These frameworks allow that the eigenvalues of the linearizations can be computed efficiently by, for example, the rational Krylov method.

3.1 The CORK framework

The starting point of the compact rational Krylov (CORK) method in Van Beeumen et al. (2015) is a matrix-valued function of the form

$P(\lambda) = \sum_{i=0}^{k-1} (A_i - \lambda B_i) f_i(\lambda)$, \quad with \quad $A_i, B_i \in \mathbb{C}^{n \times n}$, \quad (3.1)

where $f_i : \mathbb{C} \to \mathbb{C}$ are polynomial or rational functions satisfying the linear relation

$(M - \lambda N) f(\lambda) = 0$ \quad with \quad $\text{rank}(M - \lambda N) = k - 1$ for all $\lambda \in \mathbb{C}$ \quad (3.2)

and $f(\lambda) = [f_0(\lambda) \quad \cdots \quad f_{k-1}(\lambda)]^T \neq 0$. Without much loss of generality, we further assume that $f_0(\lambda) \equiv 1$ has degree zero. This assumption is indeed not restrictive, as it covers most of the important cases in applications including monomials, Chebyshev polynomials, orthogonal polynomials, Newton polynomials, rational Newton functions. For a more general setting and explicit examples of $M - \lambda N$, we refer to Van Beeumen et al. (2015).

Given a matrix-valued function (3.1) satisfying (3.2), the matrix pencil

$L_P(\lambda) = \begin{bmatrix} A_0 - \lambda B_0 & \cdots & A_{k-1} - \lambda B_{k-1} \\ (M - \lambda N) \otimes I_n \end{bmatrix}$, \quad (3.3)

is called the CORK linearization of $P(\lambda)$. When $P(\lambda)$ is a matrix polynomial, the pencil (3.3) is a linearization in the usual sense (Gohberg et al., 1982), since it corresponds to a “block minimal bases pencil”, see M. Dopico et al. (2017); Robol et al. (2016). When $P(\lambda)$ is a rational matrix, it is not clear whether the pencil (3.3) is a linearization of $P(\lambda)$ in the sense of Amparan et al. (2016). In any case, for our purposes, we only need to use that $P(\lambda)$ and $L_P(\lambda)$ have the same eigenvalues and that the eigenvectors of $P(\lambda)$ can be easily recovered from those of $L_P(\lambda)$ (see Van Beeumen et al., 2015, Corollary 2.4).

The CORK linearization (3.3) is of size $kn \times kn$ which can become quite large. Fortunately, its Kronecker structure can be exploited when computing its eigenvalues by Krylov methods; see, e.g., (Van Beeumen et al., 2015, Algorithm 3) on how to efficiently use the rational Krylov method in this context.

3.2 Extending CORK with AAA

Let us now consider the NEP with $A(\lambda)$ as defined in (1.2), that is,

$A(\lambda) = P(\lambda) + \sum_{i=1}^{m} (C_i - \lambda D_i) g_i(\lambda)$.
Using AAA, or its set-valued generalization, we can approximate each function \( g_i(\lambda) \) on the region \( \Sigma \subset \mathbb{C} \) as
\[
g_i(\lambda) \approx r_i(\lambda) = \sum_{j=1}^{\ell_i} \frac{g_i(z^{(i)}_j) \omega^{(i)}_j}{\lambda - z^{(i)}_j} / \sum_{j=1}^{\ell_i} \omega^{(i)}_j,
\]
(3.4)
where \( \ell_i \) is the number of support points \( z^{(i)}_j \) and weights \( \omega^{(i)}_j \) for each \( i = 1, \ldots, s \). If some of the \( g_i \) are approximated together by the set-valued AAA algorithm, the \( z^{(i)}_j \) and \( \omega^{(i)}_j \) are the same for their corresponding indices \( i \). For now, we ignore this property. In any case, we can use the rational approximations \( r_i(\lambda) \) to obtain an approximation of the NEP on the same region \( \Sigma \):
\[
A(\lambda) \approx R(\lambda) = P(\lambda) + \sum_{i=1}^{s} (C_i - \lambda D_i) r_i(\lambda).
\]
(3.5)

We now show how to obtain a CORK-like linearization of \( R(\lambda) \). If we assume that \( P(\lambda) \) satisfies (3.1), then by making use of Prop. 2.1, we can also write the rational part in (3.5) in state-space form as
\[
R(\lambda) = \sum_{i=0}^{k-1} (A_i - \lambda B_i) f_i(\lambda) + \sum_{i=1}^{s} (C_i - \lambda D_i) a_i^T (E_i - \lambda F_i)^{-1} b_i
\]
(3.6)
for some vectors \( a_i, b_i \in \mathbb{C}^{\ell_i} \) and the \( \ell_i \times \ell_i \) matrices
\[
E_i = \begin{bmatrix}
\omega_1 & \omega_2 & \cdots & \omega_{\ell_i - 1} & \omega_{\ell_i} \\
-z_1 & z_2 & \cdots & \cdots & \cdots \\
& -z_2 & \cdots & \cdots & \cdots \\
& & \cdots & \cdots & \cdots \\
& & & \cdots & \cdots \\
& & & & \cdots \\
\end{bmatrix}
\quad \text{and} \quad
F_i = \begin{bmatrix}
0 & 0 & \cdots & 0 & 0 \\
1 & -1 & \cdots & & \cdots \\
& 1 & \cdots & -1 & \cdots \\
& & \cdots & 1 & \cdots \\
& & & & \cdots \\
& & & & & \cdots \\
\end{bmatrix}.
\]

Next, introduce for \( i = 1, \ldots, s \) the vector-valued function
\[
R_i: \mathbb{C} \rightarrow \mathbb{C}^{\ell_i}, \quad R_i(\lambda) = (E_i - \lambda F_i)^{-1} b_i.
\]
(3.7)
Assuming that \( P(\lambda) \) satisfies (3.1) with \( f_0(\lambda) = 1 \) and observing that \((E_i - \lambda F_i)R_i(\lambda) = b_i \) for all \( i = 1, \ldots, s \), we obtain the linear relation
\[
\begin{bmatrix}
M - \lambda \mathcal{N} \\
-\mathbf{b}_1 \quad 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-\mathbf{b}_s \quad 0 & \cdots & 0 \\
\end{bmatrix}
\begin{bmatrix}
f_0(\lambda) \\
f_1(\lambda) \\
\vdots \\
f_{\ell-1}(\lambda) \\
\end{bmatrix} = 0.
\]

Collecting the basis functions into the single vector
\[
f(\lambda) = \begin{bmatrix}
f_0(\lambda) \\
\vdots \\
f_{\ell-1}(\lambda)
\end{bmatrix}, \quad \Psi(\lambda) = \begin{bmatrix}
f(\lambda) \\
R_1(\lambda) \\
\vdots \\
R_s(\lambda)
\end{bmatrix},
\]
(3.8)
we arrive at the following result.

**Proposition 3.1** Let $\Psi(\lambda)$ be the vector-valued function (3.8) with $f_i(\lambda)$ scalar functions satisfying (3.2) such that $f_0(\lambda) = 1$ and $R_i(\lambda)$ satisfying (3.7). If $\lambda \in \mathbb{C}$ is such that $E_i - \lambda F_i$ is invertible for all $i = 1, \ldots, s$, then

$$
(M - \lambda N)\Psi(\lambda) = 0 \quad \text{with} \quad \hat{M} - \lambda \hat{N} = \begin{bmatrix} M - \lambda N & 0 \\ -b & E - \lambda F \end{bmatrix},
$$

(3.9)

where we used

$$
b = \begin{bmatrix} b_1^T & \cdots & b_s^T \end{bmatrix}^T \quad \text{and} \quad E - \lambda F = \text{diag}(E_1 - \lambda F_1, \ldots, E_s - \lambda F_s).
$$

(3.10)

Furthermore, the pencil $\hat{M} - \lambda \hat{N}$ has full row rank for any $\lambda \in \mathbb{C}$ such that $E_i - \lambda F_i$ is invertible for all $i = 1, \ldots, s$.

**Proof.** The identity (3.9) was shown above since $f_0(\lambda) = 1$. The second result is immediate since $M - \lambda N$ has full row rank by assumption and $E - \lambda F$ is only singular when one of the $E_i - \lambda F_i$ is singular.

In order to obtain a linearization of (3.6), we first write it using (3.7) equivalently as

$$
R(\lambda) = \sum_{i=0}^{k-1} (A_i - \lambda B_i)(f_i(\lambda) \cdot I_n) + \sum_{i=1}^{s} (C_i - \lambda D_i)(a_i^T R_i(\lambda) \cdot I_n)
$$

$$
= \sum_{i=0}^{k-1} (A_i - \lambda B_i)(f_i(\lambda) \otimes I_n) + \sum_{i=1}^{s} [a_i^T \otimes (C_i - \lambda D_i)] (R_i(\lambda) \otimes I_n).
$$

Observe that this is a trivial rewriting of scalar multiplications in terms of Kronecker products. However, using the vector $\Psi(\lambda)$ as defined in (3.8), it allows us to express the rational expression in (3.6) as

$$
R(\lambda) = \begin{bmatrix} A_0 - \lambda B_0 & \cdots & A_{k-1} - \lambda B_{k-1} \\ a_1^T \otimes (C_1 - \lambda D_1) & \cdots & a_s^T \otimes (C_s - \lambda D_s) \end{bmatrix} (\Psi(\lambda) \otimes I_n).
$$

Together with Prop. 3.1, this suggests the following CORK-like linearization of (3.6).

**Definition 3.2** (CORK linearization for AAA rational approximation) Let $R(\lambda)$ be the rational approximation (3.6) obtained by using the AAA algorithm or the set-valued AAA algorithm. We define the pencil $\mathcal{L}_R(\lambda)$ as follows

$$
\mathcal{L}_R(\lambda) = \begin{bmatrix} A_0 - \lambda B_0 & \cdots & A_{k-1} - \lambda B_{k-1} \\ a_1^T \otimes (C_1 - \lambda D_1) & \cdots & a_s^T \otimes (C_s - \lambda D_s) \end{bmatrix} \begin{bmatrix} M - \lambda N & 0 \\ -b & E - \lambda F \end{bmatrix},
$$

(3.11)

where the pencil $\hat{M} - \lambda \hat{N}$ has been defined in (3.9).

The size of $\mathcal{L}_R(\lambda)$ is $(k + \sum_{i=1}^{s} \ell_i)n$. Fortunately, one can again exploit the Kronecker structure and show that the CORK algorithm can be applied to (3.11), as long as the shifts in the shift-and-invert steps of the rational Krylov method are not poles of the rational interpolants (3.4). Furthermore, as a special case of Theorem 3.4 below with full-rank matrices, any $\lambda \in \mathbb{C}$ that is not a pole of any of the rational interpolants (3.4) is an eigenvalue of $R(\lambda)$ if and only if it is an eigenvalue of (3.11), and their associated
right eigenvectors are easily related. For the set-valued AAA approximation, we have that $E_i - \lambda F_i$ is the same for all $i$, as well as all $b_i$. As a result, linearization (3.11) becomes

$$\mathcal{L}_k(\lambda) = \begin{bmatrix} A_0 - \lambda B_0 & \cdots & A_{k-1} - \lambda B_{k-1} & \sum_{i=1}^{s} a_i^T \otimes (C_i - \lambda D_i) \\
M - \lambda N & 0 & 0 & 0 \\
-b_1 & 0 & 0 & E_1 - \lambda F_1 \end{bmatrix},$$

which is of size $(k + \ell_1)n$.

Conclusively, the CORK algorithm can be applied to (3.11) for computing eigenvalues of $R(\lambda)$ that are not poles of the rational interpolants (3.4) and their associated right eigenvectors. In practice, we have noticed that this assumption is not very restrictive, since the AAA algorithm tends to place the poles outside the region of interest.

### 3.3 Low-rank exploitation

In several applications, the matrix coefficients of the nonlinear valued function $G(\lambda)$ in (1.3) are usually of low rank. In this section, we show how the exploitation of these low ranks leads to a linearization of size smaller than that of $\mathcal{L}_k(\lambda)$. This linearization generalizes the one used in M. Dopico & González-Pizarro (2017); Su & Bai (2011), which is valid when $P(\lambda)$ in (3.1) is expressed using monomials, i.e., $f_i(\lambda) = \lambda^i$, to the more general setting used by CORK.

Suppose that the coefficients of the rational part in (3.6) admit the following structure

$$C_i - \lambda D_i = (\tilde{C}_i - \lambda \tilde{D}_i) \tilde{Z}_i^k$$

with $\tilde{C}_i, \tilde{D}_i, \tilde{Z}_i \in \mathbb{C}^{n \times k}$, and $\tilde{Z}_i^k \tilde{Z}_i = I_k$. (3.12)

Observe that this holds trivially for $\tilde{Z}_i = I_k$, but in many problems $k_i$ is potentially much smaller than $n$. In Su & Bai (2011), $\tilde{C}_i, \tilde{D}_i, \tilde{Z}_i$ are the result of a rank revealing decomposition of $C_i - \lambda D_i$.

Introducing the matrices $\tilde{Z}_i$ in the definition of $R(\lambda)$ in (3.6), we obtain

$$R(\lambda) = \sum_{i=0}^{k-1} (A_i - \lambda B_i) f_i(\lambda) + \sum_{i=1}^{s} [a_i^T \otimes (\tilde{C}_i - \lambda \tilde{D}_i)] (R_i(\lambda) \otimes I_k) \tilde{Z}_i^k,$$

where we recall that $R_i(\lambda) = (E_i - \lambda F_i)^{-1} b_i$. We can therefore write

$$R(\lambda) = \left[ \begin{array}{cc} A - \lambda B & C - \lambda D \end{array} \right] \cdot \Psi(\lambda)$$

using the matrices

$$A = \begin{bmatrix} A_0 & \cdots & A_{k-1} \end{bmatrix}, \quad B = \begin{bmatrix} B_0 & \cdots & B_{k-1} \end{bmatrix},$$

$$C = \begin{bmatrix} a_1^T \otimes \tilde{C}_1 & \cdots & a_s^T \otimes \tilde{C}_s \end{bmatrix}, \quad D = \begin{bmatrix} a_1^T \otimes \tilde{D}_1 & \cdots & a_s^T \otimes \tilde{D}_s \end{bmatrix},$$

$$f(\lambda) = \begin{bmatrix} f_0(\lambda) \\ \vdots \\ f_{k-1}(\lambda) \end{bmatrix}, \quad \Psi(\lambda) = \begin{bmatrix} f(\lambda) \otimes I_n \\ (R_i(\lambda) \otimes I_k) \tilde{Z}_i^k \\ \vdots \\ (R_i(\lambda) \otimes I_k) \tilde{Z}_i^k \end{bmatrix}.$$
Denoting by $O$ a matrix of all zeros (of suitable size), and using

$$M = M \otimes I_n, \quad N = N \otimes I_n,$$

where the pencil $M - \lambda N$ is the one in (3.2), we obtain from $(M - \lambda N)f(\lambda) = 0$ the identity

$$[M - \lambda N O] \cdot \Psi(\lambda) = O.$$

As before $(E_i - \lambda F_i)R_i(\lambda) = b_i$, whence

$$[(E_i - \lambda F_i) \otimes I_k] [R_i(\lambda) \otimes I_k] \tilde{Z}_i^* = [b_i \otimes I_k] \tilde{Z}_i^*.$$

Therefore by assuming again that $f_0(\lambda) \equiv 1$ and introducing

$$E = \text{diag}(E_1 \otimes I_k, \ldots, E_s \otimes I_k), \quad F = \text{diag}(F_1 \otimes I_k, \ldots, F_s \otimes I_k),$$

$$e_1^T = [1 \quad 0 \quad \cdots \quad 0] \in \mathbb{R}^k, \quad Z^* = \begin{bmatrix} -(b_1 \otimes I_k) \tilde{Z}_1^* \\ \vdots \\ -(b_s \otimes I_k) \tilde{Z}_s^* \end{bmatrix} (e_1^T \otimes I_n)$$

we obtain the identity

$$[Z^* E - \lambda F] \cdot \Psi(\lambda) = O.$$

Putting all the identities from above together, we obtain the following square matrix of size $d = nk + \sum_{i=1}^s \ell_i k_i$:

$$\mathcal{L}_R(\lambda) \Psi(\lambda) = \begin{bmatrix} R(\lambda) \\ O \end{bmatrix} \quad \text{with} \quad \mathcal{L}_R(\lambda) = \begin{bmatrix} A - \lambda B & C - \lambda D \\ M - \lambda N & O \\ Z^* & E - \lambda F \end{bmatrix}.$$

(3.13)

In Theorem 3.4 below we show that, as long as $\lambda$ is not a pole of any of the rational functions $r_i(\lambda)$ in (3.5), $\mathcal{L}_R(\lambda)$ is indeed a linearization for $R(\lambda)$ in the sense that we can use it to compute the eigenpairs of $R(\lambda)$. Observe that $d$ is never larger than $d = n(k + \sum_{i=1}^s \ell_i)$, the size of $L_R(\lambda)$. Hence, $\mathcal{L}_R(\lambda)$ is a trimmed linearization that effectively exploits the low-rank terms in the rational part of $R(\lambda)$. It is also possible to exploit low-rank terms in $P(\lambda)$ as is done in Van Beeumen et al. (2015). However, as this would complicate notation and the gain in size is typically less significant, we do not pursue this here.

Together with Theorem 3.4, we also have in Theorem 3.3 an explicit block-UL factorization of $\mathcal{L}_R(\lambda)$. The proof of both these results is fairly standard and is therefore devoted to the appendix—in particular, we refer to similar results in Su & Bai (2011) for rational terms $R_i(\lambda)$ with explicit state-space representation, in Van Beeumen et al. (2015) for $P(\lambda)$ in CORK form, and in M. Dopico & González-Pizarro (2017) for $P(\lambda)$ in companion form combined with explicit state space for $R_i(\lambda)$. However, a compact representation that is a combination of general $P(\lambda)$ in CORK form and rational terms stemming from AAA is new.

The theorems are stated for a certain permuted version of the columns of $\mathcal{L}_R(\lambda)$. By assumption, the $(k - 1) \times k$ pencil $M - \lambda N$ has rank $k - 1$. Hence, there exists a permutation $\Pi \in \mathbb{R}^{k \times k}$, possibly depending on $\lambda$, such that

$$(M - \lambda N)\Pi =: \begin{bmatrix} m_0 - \lambda n_0 & M_1 - \lambda N_1 \end{bmatrix} \quad \text{with} \quad M_1 - \lambda N_1 \text{ nonsingular.}$$
Denoting $\Pi = \Pi \otimes I_n$, we can also apply this permutation block-wise to the first $nk$ columns of $\widetilde{R}(\lambda)$. We then obtain

$$\begin{bmatrix} A - \lambda B & C - \lambda D \\ M - \lambda N & O \\ Z^* & E - \lambda F \end{bmatrix} \begin{bmatrix} \Pi \\ I \end{bmatrix} = \begin{bmatrix} A_0 - \lambda B_0 & A_1 - \lambda B_1 & C - \lambda D \\ M_0 - \lambda N_0 & M_1 - \lambda N_1 & O \\ Z_0^* & Z_1^* & E - \lambda F \end{bmatrix},$$

(3.14)

where $I$ denotes an identity matrix of suitable size, and with

$$M_0 = m_0 \otimes I_n, \quad n_0 = n_0 \otimes I_n, \quad M_1 = M_1 \otimes I_n, \quad N_1 = N_1 \otimes I_n.$$ 

The other block matrices are partitioned accordingly. This means $A_0 = A_j$ and $B_0 = B_j$ for some $j$ that corresponds to the column that $\Pi$ has permuted to the first position.

As mentioned above, one of the results is a block UL factorization. Amongst others, it is key for performing efficiently the shift-and-invert steps of the rational Krylov method when computing the eigenvalues of $\widetilde{R}(\lambda)$.

**Theorem 3.3** Let $\widetilde{R}(\lambda)$ be the pencil in (3.13) for the rational matrix $R(\lambda)$ in (3.6) with the low-rank structure (3.12). If $\mu \in \mathbb{C}$ is such that all $E_1 - \mu F_1, \ldots, E_s - \mu F_s$ are nonsingular, then using the block matrices as defined in (3.14), the following block-UL decomposition holds:

$$\widetilde{L}(\mu) \rho = \mathcal{U}(\mu) \mathcal{L}(\mu),$$

where (empty blocks are zero and $\rho = \sum_{i=1}^s \ell_i k_i$)

$$\rho = \begin{bmatrix} \Pi \\ I \end{bmatrix}$$

$$\mathcal{U}(\mu) = \begin{bmatrix} I_n & (A_1 - \mu B_1 - Z_1^* (C - \mu D)) [M_1 - \mu N_1]^{-1} (C - \mu D)(E - \mu F)^{-1} \\ I_{(k-1)n} & I_\rho \end{bmatrix},$$

$$\mathcal{L}(\mu) = \begin{bmatrix} \alpha(\mu)^{-1} R(\mu) \\ M_0 - \mu N_0 & M_1 - \mu N_1 \\ Z_0^* & Z_1^* & E - \mu F \end{bmatrix}, \quad \alpha(\mu) = \varepsilon_1^T \Pi^T f(\mu) \neq 0.$$

In addition,

$$\alpha(\mu)^{\rho} \det \widetilde{R}(\mu) = \det R(\mu) \left( \det (M_1 - \mu N_1) \right)^{k-1} \prod_{i=1}^s \left( \det (E_i - \mu F_i) \right)^{\ell_i}. \quad (3.15)$$

**Proof.** See appendix 5.

Next, we have the main result for the linearization: the relation of the eigenvalues (and their algebraic and geometric multiplicities) and eigenvectors of the rational matrix $R(\lambda)$ with those of the matrix trimmed pencil $\widetilde{R}(\lambda)$.

**Theorem 3.4** Let $\widetilde{R}(\lambda)$ be the pencil in (3.13) for the rational matrix $R(\lambda)$ in (3.6) with $f_0(\lambda) \equiv 1$ and the low-rank structure (3.12). Let $\lambda_0 \in \mathbb{C}$ be such that all $E_1 - \lambda_0 F_1, \ldots, E_s - \lambda_0 F_s$ are nonsingular. Denote $\rho = \sum_{i=1}^s \ell_i k_i.$
(a) If \( x \in \mathbb{C}^n \) is an eigenvector of \( R(\lambda) \) with eigenvalue \( \lambda_0 \), then \( \Psi(\lambda_0) x \in \mathbb{C}^{kn+\rho} \) is an eigenvector of \( \tilde{R}(\lambda) \) with eigenvalue \( \lambda_0 \).

(b) If \( z \in \mathbb{C}^{kn+\rho} \) is an eigenvector of \( \tilde{R}(\lambda) \) with eigenvalue \( \lambda_0 \), then \( z = \Psi(\lambda_0) x \) for some eigenvector \( x \in \mathbb{C}^n \) of \( R(\lambda) \) with eigenvalue \( \lambda_0 \).

(c) The algebraic and geometric multiplicities of \( \lambda_0 \) as an eigenvalue of \( \tilde{R}(\lambda) \) and as an eigenvalue of \( R(\lambda) \) are the same.

\[ \text{Proof.} \] See appendix 5. \[ \square \]

4. Numerical examples

This section illustrates the theory with a number of applications. The AAA algorithm is used to approximate different nonlinear matrix functions and the accuracy of the resulting rational approximations is compared to the accuracy obtained with potential theory (Leja–Bagby points). The approximations are compared in terms of accuracy and the number of poles (which is equal to the degree plus one) to achieve that accuracy. We used the rational Krylov method (more specifically, its CORK implementation in Van Beeumen et al. (2015)) to obtain eigenvalue and eigenvector estimates of the rational approximation.

We compare the following three methods.

NLEIGS This is the static variant from Güttel et al. (2014). The rational polynomial is expressed in a basis of rational Newton polynomials. The poles are selected in the branch cut of the nonlinear function and the nodes are Leja-Bagby points.

AAA-EIGS This is the rational Krylov method applied to linearization (3.3). The rational functions are determined by applying AAA to the \( m \) nonlinear functions from (1.3) separately.

SV-AAA-EIGS This is the rational Krylov method applied to linearization (3.3). The rational functions are determined using the set-valued AAA approach explain in §2.2.

In the next section, we review the rational Krylov method that is used in the numerical experiments for finding eigenvalues of the linearizations. In the numerical experiments, we do not use implicit restarting, i.e., the number of iterations corresponds to the dimension of the Krylov space.

4.1 The rational Krylov method

The rational Krylov method, sketched in Algorithm 1, is a generalization of the shift-and-invert Arnoldi method for solving large-scale generalized eigenvalue problems.

At step \( j \), Algorithm 1 computes a matrix \( V_{j+1} \) whose columns form an orthonormal basis for the rational Krylov subspace

\[
\mathcal{K}_j(A, B, v_1) = \text{span}\{u_1, u_2, \ldots, u_{j+1}\},
\]

where \( u_{i+1} = (A - \sigma_i B)^{-1} B v_i \), for \( i = 1, \ldots, j \). Furthermore, the matrix \( V_{j+1} \) satisfies the rational Krylov recurrence relation

\[
A V_{j+1} L_j = B V_{j+1} K_j,
\]

where
Algorithm 1 Rational Krylov method

1: Choose vector $v_1$, where $\|v\|_2 = 1$.
2: for $j = 1, 2, \ldots$ do
3: Choose shift $\sigma_j$.
4: Choose continuation vector $t_j$.
5: Compute $\hat{v} = (A - \sigma_jB)^{-1}Bw_j$, where $w_j = V_j t_j$.
6: Orthogonalize $\tilde{v} = v - V_j h_j$, where $h_j = V_j^* \hat{v}$.
7: Get new vector $v_{j+1} = \tilde{v} / h_{j+1}$, where $h_{j+1} = \|\tilde{v}\|_2$.
8: Set $V_{j+1} = [V_j \ v_{j+1}]$.
9: end for

where $H_j \in \mathbb{C}^{(j+1) \times j}$ is an upper Hessenberg matrix whose nonzero entries are the Gram-Schmidt coefficients computed by Algorithm 1, and

$$K_j = H_j \ \text{diag}(\sigma_1, \ldots, \sigma_j) + T_j \in \mathbb{C}^{(j+1) \times j},$$

where $T_j$ is an upper triangular matrix whose $i$th column is the continuation vector $t_i$ appended with some extra zeros. Note that we use $t_j = e_j$, where $e_j$ is the $j$-th vector of $I$, i.e. we always choose $w_j$ as the iteration vector of the previous step.

Approximations for the eigenvalues and right eigenvectors of the pencil $A - \lambda B$ are obtained by solving the small generalized eigenvalue problem

$$K_j s_i = \lambda_i H_j s_i,$$

where $H_j$ and $K_j$ are, respectively, the $j \times j$ upper part of $H_j$ and $K_j$. The pair $(\lambda_i, x_i = V_{j+1}^* h_j s_i)$ is referred to as a Ritz pair of the pencil $A - \lambda B$.

4.2 Gun problem

The radio-frequency gun cavity problem from the NLEVP collection Betcke et al. (2013) is described by the following matrix-valued function in $\lambda$

$$A(\lambda) = K - \lambda M + i\sqrt{(\lambda - \sigma_1^2)} W_1 + i\sqrt{(\lambda - \sigma_2^2)} W_2,$$

where $K, M, W_1, W_2 \in \mathbb{R}^{9956 \times 9956}$ are symmetric, positive semi-definite matrices, $\sigma_1 = 0$, and $\sigma_2 = 108.8774$.

We accurately approximate the nonlinear part in $A(\lambda)$ in a semi-circle $\Sigma$ in the complex plane, see Figure 1a. The function is approximated by a rational function in two different ways. First, for NLEIGS, a rational polynomial with 31 poles is used, with poles picked on the branch cut of $\sqrt{(\lambda - \sigma_j^2)}$ on the open interval $(-\infty, \sigma_2]$. Second, for AAA-EIGS, the AAA test set consists of 500 random points in the semi-disk combined with 500 equally distributed points on the boundary. The resulting poles for both functions $f_j = \sqrt{(\lambda - \sigma_j^2)}, j = 1, 2,$ are plotted in Figure 1a. Note that not all of the poles are shown, only the ones with a positive real part.

Figure 1b shows the approximation error as a function of the number of poles of the rational polynomial for NLEIGS, AAA-EIGS, and SV-AAA-EIGS. The approximation error is expressed as the relative
summed error on the nonlinear functions:

\[ E_f = \sqrt{\sum_i \left( \sum_j f_j(s_i) \right)^2}, \quad (4.1) \]

for \( j = 1, 2 \), and as the error on the matrix functions:

\[ E_m = \sqrt{\sum_i \frac{\| A(s_i) - R(s_i) \|_2^2}{\| A(s_i) \|_1^2}}, \]

where the points \( s_i, i = 1, \ldots, 1000 \) are points of a test set consisting of random points on the bounds and inside the semi-circle. AAA-EIGS leads to a small reduction in degree compared to NLEIGS. An important reduction, however, is achieved using the set-valued variant, SV-AAA-EIGS: 17 poles are sufficient for an accuracy of \( 10^{-13} \) for the AAA set, whereas NLEIGS requires 31 poles. Both errors \( E_f \) and \( E_m \) are very comparable.

We then determine eigenvalue and eigenvector estimates of the rational eigenvalue problems, around the central point \( \text{Re}(s) = 250^2 \), using the rational Krylov method. We used the same shifts as in Güttel et al. (2014), i.e., three shifts equally spaced on the real axis and two inside the semi-circle. Figure 1c shows residual norms of the five fastest converging Ritz values, as a function of the iteration count, for NLEIGS and SV-AAA-EIGS. The residual norms are defined as

\[ \rho_i = \frac{\| A(\lambda_i)x_i \|_2}{\| A(\lambda_i) \|_1 \| x_i \|_2}, \quad (4.2) \]

where \( \lambda_i \) is the \( i \)th eigenvalue estimate, or Ritz value, and \( x_i \) is an associated Ritz vector. In the figure, comparable convergence behavior is observed for both approaches, with slightly less accurate results for NLEIGS.

### 4.3 Bound states in semiconductor devices

Determining bound states of a semiconductor device requires the solution of the Schrödinger equation, which, after discretization, leads to a nonlinear eigenvalue problem with the matrix-valued function:

\[ A(\lambda) = H - \lambda I + \sum_{j=0}^{80} e^{i\sqrt{\lambda - \alpha_j}} S_j, \]

where \( H, S_j \in \mathbb{R}^{16281 \times 16281} \), see Vandenberghe et al. (2014); Van Beeumen (2015). Matrix \( H \) is symmetric and matrices \( S_j \) have low rank. This function has 81 branch points on the real axis at \( \lambda = \alpha_j, j = 0, \ldots, 80 \), between \(-0.19 \) and 22.3, as can be seen in Figure 2. There is a branch cut running from \([-\infty, \alpha_0]\) and one between each branch point.

For approximating the nonlinear functions with Leja–Bagby points, in Vandenberghe et al. (2014), a transformation is used that removes the branch cut between two predetermined, subsequent branch points, i.e., for \( \lambda \in [\alpha_{i-1}, \alpha_i] \). The interpolant based on these Leja–Bagby points is only valid for \( \lambda \)-values within this interval. For interval \([\alpha_0, \alpha_i]\), a rational approximation with 50 poles was used Vandenberghe et al. (2014). This corresponds to the green triangular marker in Figure 2.
Fig. 1: Results for the gun problem.
In contrast, using AAA and set-valued AAA, the 81 nonlinear functions are approximated on the real axis, over multiple branch points, without first transforming the problem. Figure 2 shows the resulting number of poles for approximating the nonlinear functions with an accuracy of $10^{-13}$ on a test set of 2000 equally spaced points, between $\alpha_0$ and the following seven branch points, i.e. for $\lambda \in [\alpha_0, \alpha_i], \ i = 1, \ldots, 7$. For example, the third bullet marker indicates the number of poles when the AAA test set runs from $\alpha_0$ to $\alpha_3$, so that it includes branchpoints $\alpha_0, \alpha_1, \alpha_2$ and $\alpha_3$ and the branch cuts in between. We show the results for AAA and set-valued AAA, the second resulting in a significant reduction of the number of poles for the 81 functions in the example.

The two eigenvalues between $\alpha_0$ and $\alpha_7$ are situated in the first interval, $[\alpha_0, \alpha_1]$, as can be seen in Figure 2. We used the rational Krylov method with five equally spaced shifts from $\alpha_0 + \varepsilon$ to $\alpha_1 - \varepsilon$, where $\varepsilon = 10^{-2}$. The convergence behavior, using a SV-AAA-EIGS with 141 poles for interval $[\alpha_0, \alpha_7]$, is shown in Figure 3. We used $\|A(\lambda)x\|_2$ as error measure, to be able to compare the results to those found in Van Beeumen (2015). The behavior is comparable to that observed for the static variant of NLEIGS.

![Graph showing number of poles for NLEIGS, AAA-EIGS and SV-AAA-EIGS for seven intervals, for the semiconductor problem.](image)

**Fig. 2:** Number of poles for NLEIGS, AAA-EIGS and SV-AAA-EIGS, for seven intervals, for the semiconductor problem.

### 4.4 Sandwich beam

A beam, consisting of two steel layers surrounding a damping layer, is modeled using the following matrix function

$$A(\lambda) = K - \lambda^2 M + \frac{G_0 + G_\infty (i\lambda \tau)^\alpha}{1 + (i\lambda \tau)^\alpha} C,$$
with $K, M, C \in \mathbb{R}^{168 \times 168}$ symmetric positive semi-definite matrices Van Beeumen et al. (2013). Here, $G_0 = 350.4\text{kPa}$ is the static shear modulus, $G_\infty = 3.062\text{MPa}$ is the asymptotic shear modulus, $\tau = 8.23\text{ns}$ is the relaxation time and $\alpha = 0.675$ a fractional parameter. Variable $\lambda$ is the angular frequency and we are interested in eigenvalues in the range $\lambda \in [200, 30000]$.

We use the AAA algorithm to approximate the function, where a sample set consists of $10^4$ equidistant points within the frequency range. The algorithm converges with 11 poles. The location of the poles is shown in Figure 4 and the approximation error, on a random test set of 1000 points $\lambda \in [200, 30000]$, is shown in Figure 5a. Note that $\text{Re} \lambda < -1$ and $\text{Im} \lambda > 0$ for all poles, so that we can visualize the poles on a logarithmic axes, with on the negative real axis $-\log_{10}|\text{Re} \lambda|$. Note also that we do not use set-valued AAA, as there is only one non-polynomial function.

We then use the rational Krylov method to obtain eigenvalue and eigenvector estimates of $A(\lambda)$. We use 10 shifts, $[2, 5, 10, 100, 200, 210, 220, 230, 240, 250] \cdot 100$, i.e., with more shifts located near the end of the interval. This results in the Ritz values also shown in Figure 4, on the same logarithmic axes, since $\text{Re} \lambda > 1$ for all Ritz values. The residual norms (4.2) can be seen in Figure 5b. Note that because of the large norm of $K$, most residual norm values lie around $10^{-8}$ for $k = 1$. After 50 iterations, the residual norms decrease by a factor $10^{-4}$ for most Ritz values.

**4.5 Car cavity problem**

The following model was generated by Dr. Axel van de Walle from KU Leuven, using a mesh from a Siemens tutorial, and using poro-elastic material properties from Allard & Noureddine (2009) and Chazot et al. (2013). The following matrix-valued function describes the nonlinear behavior of the sound pressure inside a car cavity with porous seats:

$$A(\lambda) = K_0 + h_K(\lambda)K_1 - \lambda^2(M_0 + h_M(\lambda)M_1),$$
where $K_0, K_1, M_0, M_1 \in \mathbb{R}^{15036 \times 15036}$ are symmetric, positive semidefinite matrices and $\lambda$ is the angular frequency $\lambda = 2\pi f$. The nonlinear functions are given as:

$$h_K(\lambda) = \frac{\phi}{\alpha(\lambda)}, \quad \alpha(\lambda) = \alpha_\infty + \frac{\sigma \phi}{i\lambda \rho_0} \left( 1 + \frac{4\alpha_\infty^2 \eta}{\sigma^2 \Lambda^2 \phi^2} \right),$$

and

$$h_M(\lambda) = \phi \left( \gamma - \frac{\gamma - 1}{\alpha'(\lambda)} \right), \quad \alpha'(\lambda) = 1 + \frac{8\eta}{i\lambda \rho_0 \Lambda^2 P_r} \sqrt{1 + \frac{\Lambda^2 P_r}{16\eta}},$$

with the parameters defined in Table 1. The nonlinear function $h_K$ is unbounded around $\lambda = 514i$ and has a branch point around $\lambda = 619i$, with a branch cut on the imaginary axis. The second nonlinear function $h_M$ is unbounded around $\lambda = 815i$ and has a branch point at $\lambda = 2089i$. These points and branch cuts are shown in Figure 6.

Since the physical model includes damping, we need to take into account that the eigenvalues have positive imaginary parts. We chose a test set with $5 \cdot 10^4$ real values, $\text{Re}(\lambda) \in [1, 300]$, and $5 \cdot 10^4$ random values in the rectangle with corner points 0 and $300 + 10^4i$, i.e. with very large imaginary part. We can see from Figure 6 that some AAA poles are picked very close to the singularities of the nonlinear functions. The degree of the set valued rational approximation is 42 for a tolerance of $10^{-13}$.

In order to compare with potential theory, we then chose a test set over a smaller region, namely the rectangle with corners 0 and $300 + 510i$. In this way, the set of points remains below the first singular

| $\alpha_\infty$ | 1.7 | $\sigma$ | $13500 \text{kgm}^{-3} \text{s}^{-1}$ | $\phi$ | 0.98 |
|---|---|---|---|---|---|
| $\eta$ | $1.839 \cdot 10^{-3}$ | $\Lambda$ | $80 \cdot 10^{-6} \text{m}$ | $\Lambda'$ | $160 \cdot 10^{-6} \text{m}$ |
| $\gamma$ | 1.4 | $\rho_0$ | 1.213 | $P_r$ | 0.7217 |
point of the function around 514i. The Leja–Bagby poles are picked on the imaginary axis, starting at \( \text{Im}(\lambda) = 514 \). This is shown in Figure 7. The degree necessary to reach a relative accuracy of \(10^{-12}\) on the border of the rectangle was 40. For AAA on this same, smaller rectangle, we found that the required approximation accuracy can be obtained with a polynomial of degree 11. We compared the results obtained by the rational Krylov method for both approximations, where we used 10 equally spaced shifts on the real axis \( \text{Re}(\lambda) \), from 1 to 300. Figure 8 shows some of the Ritz values together with the number of Krylov iterations required to reach a residual norm \(4.2\) below \(10^{-12}\). These Ritz values have low imaginary parts, confirming our choice of test set, and, as for the gun problem, convergence is comparable for both methods.

5. Conclusions
We have proposed a method for solving the nonlinear eigenvalue problem by first approximating the associated nonlinear matrix valued function using the AAA algorithm. This approximation is embedded in a state space representation, which leads to a strong linearization that can be solved using the Compact Rational Krylov method. We presented two versions: one that approximates each function separately and then the set-valued version that approximates all functions together. The latter version is very competitive with NLEIGS in terms of degree of the rational approximation, and in all our tests, AAA requires less poles than an approximation using potential theory, even for a problem with eighty different functions. The main advantage of the method is the fully automatic procedure in the determination of the rational approximations. Although we did not try solving a problem described by an
(a) Branch points and cuts and values for \( \text{Re}(\lambda) \in [-200, 300] \) and the poles nearest the two singularities.

(b) All of the AAA poles.

Fig. 6: Some test points and poles of AAA for the car cavity problem for large imaginary values.

(a) Leja–Bagby and AAA set and poles.

(b) All of the AAA poles.

Fig. 7: Leja–Bagby and AAA sets for the car cavity problem, up to 514i, and accompanying poles.
extremely large number of nonlinear functions, i.e., of the order of the size of the matrix, we expect that
the construction of the AAA approximation may become prohibitive, when the number of nonlinear
functions is extremely high, which is a disadvantage that NLEIGS does not share.

Acknowledgment
The authors thank Dr. Axel van de Walle for providing us with the car cavity model that he generated for
us using a model from a Virtual Lab tutorial (formerly LMS International, currently Siemens Industry
Software).

REFERENCES
ALLARD, J. & NOUREDDINE, A. (2009) Propagation of sound in porous media: modelling sound absorbing
materials, 2nd edn. John Wiley & Sons.
AMPARAN, A., DOPICO, F. M., MARCAIDA, S. & ZABALLA, I. (2016) Strong linearizations of rational matrices.
MIMS Preprint.
BETCKE, T., HIGHAM, N. J., MEHRMANN, V., SCHRÖDER, C. & TISSEUR, F. (2013) NLEVP: A collection of
nonlinear eigenvalue problems. ACM Trans. Math. Softw., 39, 7:1–7:28.
BEYN, W.-J. (2012) An integral method for solving nonlinear eigenvalue problems. Linear Algebra and its
Applications, 436, 3839–3863.
CHAZOT, J.-D., NENNIG, B. & PERREY-DEBAIN, E. (2013) Performances of the partition of unity finite element
method for the analysis of two-dimensional interior sound fields with absorbing materials. Journal of Sound
and Vibration, 332, 1918–1929.
EFFENBERGER, C. (2013) Robust successive computation of eigenpairs for nonlinear eigenvalue problems. SIAM
Journal on Matrix Analysis and Applications, 34, 1231–1256.
EFFENBERGER, C. & KRESSNER, D. (2012) Chebyshev interpolation for nonlinear eigenvalue problems. BIT, 52,
933–951.
FILIP, S.-I., NAKATSUKASA, Y., TREFETHEN, L. N. & BECKERMAN, B. (2017). Submitted.
GOHBERG, I., LANCASTER, P. & RODMAN, L. (1982) Matrix Polynomials. Classics in Applied Mathematics.
Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104).
Appendix A. Set-valued AAA implementation

function [r, pol, res, zer, z, f, w, errvec] = aaa2(F, Z, tol, mmax)
%AAA Computes a AAA rational approximation where F can have multiple
%outputs
%
% Input: F = matrix of data values with the number of sample points in the first
%dimension and the number of functions in the second dimension
% Z = vector of sample points
% tol = relative tolerance [1e-13]
% mmax = maximum degree [100]
%
% Output: r = AAA approximate function
% pol, res, zer = poles, residues and zeros
% z,f,w = interpolation points, function values and weights
% errvec = errors in each step
%
% Scale the functions
normF = max(abs(F),[],1);
F = bsxfun(@rdivide,F,normF);
% Left scaling matrix:
SF = spdiags(F, 0:-M:-M*(nF-1), M*nF, M);

% Initialize values
F = F(:);
R = mean(F);
errvec = zeros(mmax,1);
z = zeros(mmax,1);
f = zeros(mmax,nF);
ind = zeros(mmax,nF);
H = zeros(mmax,mmax-1);
S = zeros(mmax,mmax-1);
Q = zeros(M*nF,0);
C = zeros(M,0);

% AAA iteration:
for m = 1:mmax
    [errvec(m),loc] = max(abs(F-R));
    if ( errvec(m) <= tol )
        m = m-1;
        break
    end
    loc = mod(loc,M);
    ind(m,:) = loc + (M*(loc==0):M:(nF-1+(loc==0)*M));
    z(m) = Z(ind(m,1));
    f(m,:) = F(ind(m,:));
    C(:,end+1) = 1/(Z - z(m));
    C(ind(1:m,1),m) = 0;
    v = C(:,m)*f(m,:);
    v = SF*C(:,m)-v(:);
    q = Q(ind(m,:),1:m-1);
    q = q*S(1:m-1,1:m-1);
    Si = chol(eye(m-1,m-1)-q'*q);
    H(1:m-1,m-1) = Si*H(1:m-1,1:m-1);
    S(1:m-1,1:m-1) = S(1:m-1,1:m-1)/Si;
    S(m,m) = 1;
    Q(ind(1:m,:),:) = 0;
    nv = norm(v);
    H(1:m-1,m) = Q'*v;
    H(1:m-1,m) = S(1:m-1,1:m-1)*H(1:m-1,m);
    HH = S(1:m-1,1:m-1)*H(1:m-1,m);
    v = v - Q*H*HH;
    H(m,m) = norm(v);
    % Reorthogonalization is necessary for higher precision
    if it < 3 & & (H(m,m) < 1/sqrt(2)*nv)
        hnew = S(1:m-1,1:m-1)'*(Q'*v);
        v = v - Q*(S(1:m-1,1:m-1)*hnew);
        H(1:m-1,m) = H(1:m-1,m) + hnew;
        nv = H(m,m);
        H(m,m) = norm(v);
        it = it+1;
    end
end
AUTOMATIC RATIONAL APPROXIMATION FOR NONLINEAR EIGENVALUE PROBLEMS

v = v / H(m,m);

% Add v
Q(:,:,end+1) = v;

% Solve small least squares problem with H
[*,*V] = svd(H(1:m,1:m));
w = V(:,:,end);

% Get the rational approximation
N = C*bsxfun(@times,w,f(1:m,:));  % Numerator
D = C*bsxfun(@times,w,ones(m,nF));  % Denominator
R = N(:,:,1)/D(:,:,1);
R(ind(1:m,:)) = F(ind(1:m,:));

end

f = f(1:m,:);
w = w(1:m);
z = z(1:m);

% Scale function values back
f = bsxfun(@times,f,normF);

% Note: When M == 2, one weight is zero and r is constant.
% To obtain a good approximation, interpolate in both sample points.
if (M == 2)
z = Z;
f = F;
w = [1; -1];  % Only pole at infinity.
w = w/norm(w);  % Impose norm(w) = 1 for consistency.
errvec(2) = 0;
end

% Remove support points with zero weight:
I = find(w == 0);
z(I) = [];
w(I) = [];
f(I,:) = [];

% Construct function handle:
r = @(zz) eval(zz, z, f, w);

% Compute poles, residues and zeros:
[pol, res, zer] = prz(r, z, f, w);
end

Appendix B. Proofs of Theorems 3.3 and 3.4

The following simple lemma turns out be useful.

LEMMA A.1 Let $G_{22} \in \mathbb{C}^{m \times m}$ and $X_1 \in \mathbb{C}^{n \times n}$ be invertible matrices that satisfy

$$
\begin{bmatrix}
G_{11} & G_{12} \\
G_{21} & G_{22}
\end{bmatrix}
\begin{bmatrix}
X_1 \\
X_2
\end{bmatrix}
= 
\begin{bmatrix}
Y_1 \\
O
\end{bmatrix},
$$

(A.1)
then the following block-UL decomposition holds:

\[
\begin{bmatrix}
G_{11} & G_{12} \\
G_{21} & G_{22}
\end{bmatrix} = \begin{bmatrix}
I_h & G_{12}G_{22}^{-1} \\
O & I_m
\end{bmatrix} \begin{bmatrix}
Y_1X_1^{-1} & O \\
G_{21} & G_{22}
\end{bmatrix}.
\] (A.2)

**Proof.** After block elimination of the square matrix on the left-hand size of (A.2), we only need to show that its Schur complement \( S = G_{11} - G_{12}G_{22}^{-1}G_{21} \) equals \( Y_1X_1^{-1} \). But this follows directly from the identities \( G_{11} = (Y_1 - G_{12}X_2)X_1^{-1} \) and \( X_2 = -G_{22}^{-1}G_{21}X_1 \) that are implied by (A.1).

**Proof of Theorem 3.3.** Identity (3.15) follows from direct manipulation together with \( \det(A \otimes I_n) = (\det(A))^n \). To show the block-UL decomposition, recall from (3.13) and (3.14) that

\[
(\mathcal{L}_R(\lambda) \mathcal{P}) (\mathcal{P}^T \Psi(\lambda)) = \begin{bmatrix}
R(\lambda) \\
O
\end{bmatrix}
\text{ with } \mathcal{L}_R(\lambda) \mathcal{P} = \begin{bmatrix}
A_0 - \lambda B_0 & A_1 - \lambda B_1 & C - \lambda D \\
M_0 - \lambda N_0 & M_1 - \lambda N_1 & 0 \\
Z_0^* & Z_1^* & E - \lambda F
\end{bmatrix}.
\]

The vertical and horizontal lines indicate compatible block partitioning. The corresponding partitioning for \( \mathcal{P}^T \Psi(\lambda) \) satisfies

\[
\mathcal{P}^T \Psi(\lambda) = \begin{bmatrix}
p_1I_n \\
p_2I_n \\
\vdots \\
p_kI_n \\
(R_1(\lambda) \otimes I_k)\tilde{Z}_1^* \\
\vdots
\end{bmatrix}
\text{ with } p = \begin{bmatrix}
p_1 \\
p_2 \\
\vdots \\
p_k
\end{bmatrix} \equiv \Pi \psi(\lambda) \in \mathbb{C}^k.
\]

The required block-UL decomposition now follows from a direct calculation if we can apply Lemma A.1 to the partitioned matrix \( \mathcal{L}_R(\lambda) \mathcal{P} \). In order to be able to do this, we only have to establish that \( p_1 = e_1^T \Pi^T f(\lambda) \neq 0 \) since then \( X_1 = p_1I_n \) is invertible (and we already have \( Y_1 = R(\lambda) \)). To this end, we use the definition of \( \Pi \) to obtain

\[
(M - \lambda N)f(\lambda) = 0 \iff \begin{bmatrix}
m_0 - \lambda n_0 \\
M_1 - \lambda N_1
\end{bmatrix} p = 0.
\]

Since \( M_1 - \lambda N_1 \) is invertible and \( \dim \ker(M - \lambda N) = 1 \), the null-vector \( p \) has to be of the form

\[
p = \alpha \begin{bmatrix}
1 \\
-(m_0 - \lambda n_0)(M_1 - \lambda N_1)^{-1}
\end{bmatrix}, \quad \alpha \in \mathbb{C}.
\]

Since \( f(\lambda) \), and thus \( p \), cannot never be identically zero, we get as requested that \( \alpha = p_1 \neq 0 \).

**Proof of Theorem 3.4.** (a) Let \( (\lambda_0, x) \) be an eigenpair of \( R(\lambda) \), that is, \( R(\lambda_0)x = 0 \). Recalling that \( f(\lambda)^T = [f_0(\lambda) \cdots f_{k-1}(\lambda)] \) with \( f_0(\lambda) \equiv 1 \), we obtain

\[
y = \Psi(\lambda_0)x = \begin{bmatrix}
x \\
f_1(\lambda_0)x \\
\vdots \\
f_{k-1}(\lambda_0)x \\
(R_1(\lambda_0) \otimes I_k)\tilde{Z}_1^*x \\
\vdots \\
(R_s(\lambda_0) \otimes I_k)\tilde{Z}_s^*x
\end{bmatrix}
\]
and thus also $y \neq 0$ due to $x \neq 0$. Using (3.13), we see that $(\lambda_0, y)$ verifies the eigenpair equation

$$\tilde{\mathcal{L}}(\lambda_0)y = \begin{bmatrix} R(\lambda_0) \\ O \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$ 

(b) Let $(\lambda_0, y)$ be an eigenpair of $\tilde{\mathcal{L}}(\lambda)$. Thanks to (3.13), it suffices to show that $y = \Psi(\lambda_0)x$ for some nonzero $x \in \mathbb{C}^n$ since that implies $R(\lambda_0)x = 0$. To this end, consider the eigenpair equation $\tilde{\mathcal{L}}(\lambda_0)y = 0$ in partitioned form:

$$\begin{bmatrix} A - \lambda_0 B \\ M - \lambda_0 N \\ Z^* \end{bmatrix} \begin{bmatrix} C - \lambda_0 D \\ O \\ E - \lambda_0 F \end{bmatrix} \begin{bmatrix} y_{AB} \\ y_{CD} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$ 

The second block-row expresses that $y_{AB}$ is a null vector of $M - \lambda_0 N = (M - \lambda_0 N) \otimes I_n$. By definition of $f(\lambda)$, we have $(M - \lambda N)f(\lambda) = 0$ from which $(M - \lambda N)(f(\lambda) \otimes I_n) = 0$. Since $f_0(\lambda) \equiv 1$, the $kn \times n$ matrix $F(\lambda) := f(\lambda) \otimes I_n$ has rank $n$. In addition, by rank nullity, $\dim \ker (M - \lambda N) = n$ and so $F(\lambda)$ is a basis for $\ker (M - \lambda N)$. Hence, there exists $x \in \mathbb{C}^n$ such that (recall $f_0(\lambda) \equiv 1$)

$$y_{AB} = F(\lambda_0)x = \begin{bmatrix} x \\ f_1(\lambda_0)x \\ \vdots \\ f_{k-1}(\lambda_0)x \end{bmatrix}. \quad (A.3)$$

The third block-row reads

$$(E - \lambda_0 F)y_{CD} = -Z^*y_{AB}.$$ 

Together with the definitions of $Z, E, F$ and (A.3), we also obtain

$$\begin{bmatrix} (E_1 - \lambda_0 F_1) \otimes I_{k_1} \\ \vdots \\ (E_s - \lambda_0 F_s) \otimes I_{k_s} \end{bmatrix} y_{CD} = \begin{bmatrix} (b_1 \otimes I_{k_1}) \tilde{Z}_1^* \\ \vdots \\ (b_s \otimes I_{k_s}) \tilde{Z}_s^* \end{bmatrix} (e_1^T f(\lambda_0) \otimes I_n)x.$$ 

Hence, isolating for $y_{CD}$ and using $R_i(\lambda) = (E_i - \lambda F_i)^{-1}b_i$, we get

$$y_{CD} = \begin{bmatrix} ((E_1 - \lambda_0 F_1)^{-1} \otimes I_{k_1})(b_1 \otimes I_{k_1}) \tilde{Z}_1^* \\ \vdots \\ ((E_s - \lambda_0 F_s)^{-1} \otimes I_{k_s})(b_s \otimes I_{k_s}) \tilde{Z}_s^* \end{bmatrix} x = \begin{bmatrix} (R_1(\lambda_0) \otimes I_{k_1}) \tilde{Z}_1^* \\ \vdots \\ (R_s(\lambda_0) \otimes I_{k_s}) \tilde{Z}_s^* \end{bmatrix} x. \quad (A.4)$$

Now combining (A.3) and (A.4), we have indeed shown that there exists $x \in \mathbb{C}^n$ such that

$$y = \begin{bmatrix} y_{AB} \\ y_{CD} \end{bmatrix} = \Psi(\lambda_0)x.$$ 

Observe that $\Psi(\lambda)$ has full column rank thanks to $F(\lambda)$ being its upper block. Hence, from $y \neq 0$ it follows that $x \neq 0$ and we have proven (b).

(c) The statement about the algebraic multiplicity of $\lambda_0$ follows directly from (3.15) since the matrices $M_1 - \lambda_0 N_1$ and $E_i - \lambda_0 F_i$ are invertible by construction of $\Pi$ and by assumption, respectively,
and \( \alpha \neq 0 \). Next, we show the equality of the geometric multiplicity of \( \lambda_0 \), that is, \( \dim \ker R(\lambda_0) = \dim \ker \tilde{L}_R(\lambda_0) \). Let \( \{x_1, \ldots, x_t\} \) be a basis for \( \ker R(\lambda_0) \). Then by (a), \( \Psi(\lambda_0)x_i \in \ker \tilde{L}_R(\lambda_0) \) for \( i = 1, \ldots, t \). As argued above \( \Psi(\lambda_0) \) has full column rank, hence the vectors \( \Psi(\lambda_0)x_1, \ldots, \Psi(\lambda_0)x_t \) are linearly independent and so \( \dim \ker R(\lambda_0) \leq \dim \ker \tilde{L}_R(\lambda_0) \). Similarly, if \( \{y_1, \ldots, y_t\} \) is a basis for \( \ker \tilde{L}_R(\lambda_0) \), then by (b), \( y_i = \Psi(\lambda_0)x_i \) for some \( x_i \in \mathbb{C}^n \). Again due to \( \Psi(\lambda_0) \) having full column rank, the \( x_1, \ldots, x_t \) are linearly independent. Hence, \( \dim \ker R(\lambda_0) \geq \dim \ker \tilde{L}_R(\lambda_0) \), as we wanted to show. \( \square \)