Preprint

This is the submitted version of a paper published in *BIT Numerical Mathematics*.

Citation for the original published paper (version of record):

Mele, G., Jarlebring, E. (2018)
On restarting the tensor infinite Arnoldi method
*BIT Numerical Mathematics*, 58(1): 133-162
https://doi.org/10.1007/s10543-017-0671-z

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-218704
Restarting for the Tensor Infinite Arnoldi method

Giampaolo Mele · Elias Jarlebring

Received: date / Accepted: date

**Abstract** An efficient and robust restart strategy is important for any Krylov–based method for eigenvalue problems. The tensor infinite Arnoldi method (TIAR) is a Krylov–based method for solving nonlinear eigenvalue problems (NEPs). This method can be interpreted as an Arnoldi method applied to a linear and infinite dimensional eigenvalue problem where the Krylov basis consists of polynomials. We propose new restart techniques for TIAR and analyze efficiency and robustness. More precisely, we consider an extension of TIAR which corresponds to generating the Krylov space using not only polynomials, but also structured functions, which are sums of exponentials and polynomials, while maintaining a memory efficient tensor representation. We propose two restarting strategies, both derived from the specific structure of the infinite dimensional Arnoldi factorization. One restarting strategy, which we call semi–explicit TIAR restart, provides the possibility to carry out locking in a compact way. The other strategy, which we call implicit TIAR restart, is based on the Krylov–Schur restart method for linear eigenvalue problem and preserves its robustness. Both restarting strategies involve approximations of the tensor structured factorization in order to reduce complexity and required memory resources. We bound the error in the infinite dimensional Arnoldi factorization showing that the approximation does not substantially influence the robustness of the restart approach. We illustrate the effectiveness of the approaches by applying them to solve large scale NEPs that arise from a delay differential equation and a wave propagation problem. The advantages in comparison to other restart methods are also illustrated.

**Keywords** Nonlinear eigenvalue problem · Restart · Tensor Infinite Arnoldi · Krylov subspace method · Krylov-Schur method

Giampaolo Mele
Dept. Mathematics, KTH Royal Institute of Technology, SeRC swedish e-science research center, Lindstedtsvägen 25, Stockholm, Sweden E-mail: gmele@kth.se

Elias Jarlebring
Dept. Mathematics, KTH Royal Institute of Technology, SeRC swedish e-science research center, Lindstedtsvägen 25, Stockholm, Sweden E-mail: eliasj@kth.se
1 Introduction

We consider the nonlinear eigenvalue problem (NEP) defined as finding \((\lambda, v) \in \mathbb{C} \times \mathbb{C}^n \setminus \{0\}\) such that

\[ M(\lambda)v = 0 \quad (1) \]

where \(\lambda \in \Omega \subseteq \mathbb{C}\), \(\Omega\) is an open disk centered in the origin and \(M : \Omega \to \mathbb{C}^{n \times n}\) is analytic. The NEP has received a considerable attention in literature. See the review papers [25,37] and the problem collection [7].

There is a large number of methods available in a large amount of numerical linear algebra literature for (1). There are specialized methods for solving different classes of NEPs such as polynomial eigenvalue problems (PEPs) see [22,21,18] and [2, Chapter 9], in particular quadratic eigenvalue problems (QEPs) [32,24,23,3] and rational eigenvalue problems (REPs) [35,5,6,29]. There are also methods that exploit the structure of the operator \(M(\lambda)\) like Hermitian structure [31,30] or low rank of the matrix–coefficients [33]. Methods for solving a more general class of NEP are also present in literature. There exist methods based on modification of the Arnoldi method [36], which can be restarted for certain problems, Jacobi–Davidson methods [8], Newton–like methods [16,27,9]. Finally, there is a class of methods (to which the presented method belongs) based on Krylov methods and rational Krylov methods which can be interpreted as either dynamically expanding an approximation or applying a method on an infinite dimensional operator [11,34,14].

In principle, we do not assume any particular structure of the NEP except for the analyticity and the computability of certain quantities associated with \(M(\lambda)\) (further described later). This is similar to the infinite Arnoldi method (IAR) [14], which is in the same line of reasoning as our approach. IAR is equivalent to the Arnoldi method applied to a linear operator. More precisely, under the assumption that zero is not an eigenvalue, the problem (1) can be reformulated as \(\lambda B(\lambda)v = v\), where \(B(\lambda) = M(0)^{-1}(M(0) - M(\lambda))/\lambda\). This problem is equivalent to the linear and infinite dimensional eigenvalue problem \(\lambda \mathcal{B}\psi(\theta) = \psi(\theta)\), where \(\psi(\theta) : \mathbb{C} \to \mathbb{C}\) is an analytic function [14, Theorem 3]. The operator \(\mathcal{B}\) is linear, maps functions to functions, and is defined as

\[ \mathcal{B}\psi(\theta) := \int_0^\theta \psi(\hat{\theta})d\hat{\theta} + C(\psi), \]

where

\[ C(\psi) := \sum_{i=0}^{\infty} \frac{B^{(i)}(0)}{i!}\psi^{(i)}(0). \]

IAR has a particular structure such that it can be represented with a tensor and was the basis for the tensor infinite Arnoldi method (TIAR) in [13]. TIAR is equivalent to IAR but computationally more attractive because of the following features. In TIAR, the basis of the Krylov space, which consists of polynomials, is implicitly represented in a memory efficient way. This improves the performance in terms of memory and CPU-time. Another improvement of IAR was presented in [12]. This method consists in generating the Krylov space by using a particular type of structured functions, which are sums of polynomials and exponential functions. The main advantage of [12] is the possibility to perform a semi–explicit restart by imposing the structure. In this paper we extend the framework of TIAR to structured functions and study restart techniques.

A problematic aspect of any algorithm based on the Arnoldi method is that, when many iterations are performed, the computation time per iteration will eventually become large.
Moreover, finite arithmetic aspects may restrict the accuracy. Fortunately, an appropriate restart of the algorithm can resolve these issues in many situations. There exist two main classes of restarting strategies: explicit restart and implicit restart. Most of the explicit restart techniques correspond to selecting a starting vector that generates an Arnoldi factorization with the wanted Ritz values. The implicit restart consists in computing a new Arnoldi factorization with the wanted Ritz values. This process can be done deflating the unwanted Ritz values as in, e.g., IRA [20] or extracting a proper subspace of the Krylov space by using the Krylov–Schur restart approach [28]. The approaches are mathematically equivalent but, for reasons of numerical stability, implicit restart is often considered more robust. See [26] for further discussions about the restart of the Arnoldi method for the linear eigenvalue problems.

The paper is organized as follows: in Section 2 we extend TIAR to tensor structured functions. In Section 4 we propose a semi–explicit restart for TIAR. This new algorithm is equivalent to [12] but, the implicit representation of the Krylov basis gives improvements in terms of memory and CPU time. In section 5 we propose an implicit restart for TIAR based on an adaption of Krylov–Schur restart. The Krylov–Schur restart for the Arnoldi method in the linear case has constant CPU–time for outer iteration. In contrast to this, it turns out that a direct usage of the Krylov–Schur restart for TIAR does not give a substantial improvement due to the memory efficient representation of the Krylov basis. We show that the structure of the Arnoldi factorization allows us to perform approximations that reduce the complexity and the memory requirements. We prove that the coefficients matrix, representing the basis of the Krylov space, present a fast decay in the singular values. This allows us to effectively use a low rank approximation of such matrices. Moreover we prove that there is a fast decay in the coefficients of the polynomial part of the functions in the Krylov space. This can be exploited to introduce another approximation when the power series coefficients of $M(\lambda)$ decay to zero. We give explicit bounds on the errors due to those approximations.

There exist other Arnoldi–like methods combined with a companion linearization that use memory efficient representation of the Krylov basis matrix. See CORK [34], TOAR [17] and [38]. Similar to TIAR, the direct usage of the Krylov–Schur restart for these methods does not decrease the complexity unless SVD–based approximations are used (which is indeed suggested in the implementation of the methods). More precisely, the coefficients that represent the Krylov basis are replaced with their low rank approximations. In contrast to those approaches, our specific setting allows us to characterize the impact of the approximations.

Finally, in Section 7 we show the effectiveness of the restarting strategies with numerical simulations to large and sparse NEPs.

2 Tensor structured functions and TIAR factorizations

Similar to many restart strategies for linear eigenvalue problems, our approach is based on computation, representation and manipulation of an Arnoldi–type factorization. For our infinite-dimensional operator, the analogous Arnoldi–type factorization is defined as follows. The functions $\Psi_k$ are represented with a particular tensor structure which we further describe in Section 2.1 together with a natural sense to define orthogonality.

**Definition 1 (TIAR factorization)** Let $\Psi_{k+1}(\theta) \in \mathbb{C}^{n \times (k+1)}$ be a tensor structured function with orthogonal columns and let $H_k \in \mathbb{C}^{(k+1) \times k}$ be an Hessenberg matrix with positive ele-
ments in the sub–diagonal. The pair \((\Psi_{k+1}, \mathcal{B}_k)\) is a TIAR factorization of length \(k\) if
\[
\mathcal{B}^k \Psi_k(\theta) = \Psi_{k+1}(\theta) \mathcal{H}_k.
\]  

2.1 Representation and properties of the tensor structured functions

We consider a class of structured functions introduced in [12], which we represent in a different and more memory–efficient way.

**Definition 2** The vector–valued function \(\psi : \mathbb{C} \to \mathbb{C}^n\) is a tensor structured function if it exist \(Y, W \in \mathbb{C}^{n \times p}, \hat{a} \in \mathbb{C}^{d \times r}, \hat{b} \in \mathbb{C}^{d \times s}, \ell \in \mathbb{C}^p, S \in \mathbb{C}^{p \times p}, Z \in \mathbb{C}^{n \times l}\) where \([Z, W]\) is orthogonal and \(\text{span}(\ell) = \text{span}(W)\), such that
\[
\psi(\theta) = P_{d-1}(\theta) \left( \sum_{\ell=1}^r \hat{a}_{:,\ell} \otimes z_\ell + \sum_{\ell=1}^s \hat{b}_{:,\ell} \otimes w_\ell \right) + Y \exp_{d-1}(\theta S) \bar{c}
\]  
where
\[
P_{d}(\theta) := (1, \theta, \ldots, \theta^d) \otimes I_n
\]  
and \(\exp_{d-1}(\theta S) := \sum_{i=0}^d \theta^i S^i\) is consistent with [12].

The matrix–valued functions \(\Psi_k : \mathbb{C} \to \mathbb{C}^{n \times k}\) is a tensor structured function if it can be expressed as \(\Psi_k(\theta) = (\psi_1(\theta), \ldots, \psi_k(\theta))\), where each \(\psi_i\) is a tensor structured function. We denote the \(i\)-th column of \(\Psi_k\) by \(\psi_i\). The structure induced by (3) is now, in a compact form
\[
\Psi_k(\theta) = P_{d-1}(\theta) \left( \sum_{\ell=1}^r \hat{a}_{:,\ell} \otimes z_\ell + \sum_{\ell=1}^s \hat{b}_{:,\ell} \otimes w_\ell \right) + Y \exp_{d-1}(\theta S) \bar{c}
\]  
where \(a \in \mathbb{C}^d \times r, b \in \mathbb{C}^d \times s, c \in \mathbb{C}^p\). We say that \(\Psi_k(\theta)\) is orthogonal if the columns are orthogonal, i.e., \(<\psi_i, \psi_j> = \delta_{ij}\) for \(i, j = 1, \ldots, k\). We use the scalar product consistent with the other papers about the infinite Arnoldi method [12,14], i.e., if \(\psi(\theta) = \sum_{i=0}^n \theta^i y_i\) and \(\phi(\theta) = \sum_{i=0}^n \theta^i x_i\), then
\[
<\psi, \phi> = \sum_{i=0}^n x_i^H y_i.
\]  
The computation of this scalar product and norms for the tensor structured functions (3) can be done analogous to [12]. In particular, by definition of (4), we have the following identity which will be used several times in the derivations
\[
\|P_{d-1} W\| = \|W\|_F
\]  
for any \(W \in \mathbb{C}^{nd \times p}\).

We will carry out many operations on the tensor structured functions. Fortunately, many of the operations carry over directly to the coefficients, e.g., linear combinations.

**Observation 3 (Linearity with respect the coefficients)** Given the tensor structured function \(\Psi_k(\theta)\) represented by \((Z,W,Y,S)\) with coefficients \((a, b, C)\) and \(\overline{\Psi}_k(\theta)\) represented by the same matrices but with coefficients \((\hat{a}, \hat{b}, \hat{C})\). The function \(\overline{\Psi}_k(\theta)(M + \overline{\Psi}_k(\theta))N\) is also a tensor structured function represented by the same matrices and coefficients \((\hat{a}, \hat{b}, \hat{C})\) where for \(\ell = 1, \ldots, r\) we have defined
\[
\hat{a}_{:,\ell} := a_{:,\ell} M + \overline{\hat{a}}_{:,\ell} N, \quad \hat{b}_{:,\ell} := b_{:,\ell} M + \overline{\hat{b}}_{:,\ell} N, \quad \bar{C} := CM + \bar{C}N.
\]
We use the following notation: \( M : M^d(0) \) and \( M_d(\Psi, S) \) is defined as in [12]. In particular, any nonlinear function \( M(\lambda) \) can be represented as a sum of products of scalar nonlinearities

\[
M(\lambda) = \sum_{i=1}^{q} T_i f_i(\lambda), \quad T_i \in \mathbb{C}^{n \times n}, f_i : \Omega \to \mathbb{C}, \quad i = 1, \ldots, q,
\]

and we define \( M_d : \mathbb{C}^{n \times p} \times \mathbb{C}^{p \times p} \to \mathbb{C}^{n \times p} \) as

\[
M_d(\Psi, S) := \sum_{i=1}^{q} T_i Y f_i(S) - \sum_{i=1}^{d} \frac{MYS_i}{\ell!},
\]

which equivalently can be expressed as

\[
M_d(\Psi, S) = \sum_{i=d+1}^{\infty} \frac{MYS_i}{\ell!}.
\]

The function \( M_{-1}(\Psi, S) \) is the basis of block Newton methods for NEPs presented in [16].

The action of the operator \( B \) on functions represented as in (3) can now be expressed in a closed form using the above notation.

**Theorem 4 (Action of \( B \))** Suppose \( Y, W \in \mathbb{C}^{n \times p}, \ Z \in \mathbb{C}^{n \times r}, \ a \in \mathbb{C}^{d \times r}, \ b \in \mathbb{C}^{d \times p}, \ c \in \mathbb{C}^p \) and \( S \in \mathbb{C}^{p \times p} \). Suppose \( \lambda(S) \subset \Omega \), let \( \tilde{c} = S^{-1}c \) and

\[
\tilde{z} := -M_0^{-1} \left[ M_d(\Psi, S)\tilde{c} - \sum_{i=1}^{d} M_i \left( \sum_{i=1}^{r} \tilde{a}_{1,\ell} z_i + \sum_{i=1}^{p} \tilde{b}_{1,\ell} w_i \right) \right].
\]

Under the assumption that

\[
\tilde{z} \notin \text{span}(z_1, \ldots, z_r, w_1, \ldots, w_p),
\]

let \( z_{r+1} \) be the normalized orthogonal complement of \( \tilde{z} \) against \( z_1, \ldots, z_r, w_1, \ldots, w_p \) and \( \tilde{a}_{1,\ell} \) and \( \tilde{b}_{1,\ell} \) be the orthonormalization coefficients, i.e.,

\[
\tilde{z} = \sum_{i=1}^{r+1} \tilde{a}_{1,\ell} z_i + \sum_{i=1}^{p} \tilde{b}_{1,\ell} w_i.
\]

Then, the action of \( B \) on the tensor structured function defined by (3) is

\[
B\psi(\theta) = P_d(\theta) \left( \sum_{i=1}^{r+1} \tilde{a}_{1,\ell} \otimes z_i + \sum_{i=1}^{p} \tilde{b}_{1,\ell} \otimes w_i \right) + Y \exp_d(\theta S)\tilde{c}
\]

where

\[
\begin{align*}
\tilde{a}_{i,\ell+1} & := \begin{cases} 0, & i = 1, \ldots, d, \\ \tilde{a}_{i,\ell}/i, & i = 1, \ldots, d; \quad \ell = 1, \ldots, r, \end{cases} \\
\tilde{b}_{i,\ell+1} & := \tilde{b}_{i,\ell}/i, \quad i = 1, \ldots, d; \quad \ell = 1, \ldots, p.
\end{align*}
\]
To define approximations that avoid $r$ from being large. The hypothesis considered in this paper. Our focus is on large-scale NEPs and, in Section 5.1, we introduce

To expand a TIAR factorization

Orthogonalization

Remark

Proof

\[ x_i := \sum_{\ell=1}^r \tilde{a}_{i,\ell} z_{\ell} + \sum_{\ell=1}^u \tilde{b}_{i,\ell} w_{\ell}, \quad i = 0, \ldots, d - 1, \]

and $x := \text{vec}(x_0, \ldots, x_{d-1}) \in \mathbb{C}^{dn}$. $\psi(\theta)$ defined in (3) can be expressed as

\[ \psi(\theta) = P_{d-1}(\theta) x + Y \exp_{d-1}(\theta S) \bar{c}. \]

By invoking [12, theorem 4.2] and using (15), we can express the action of the operator as

\[ B\psi(\theta) = P_d(\theta) x_+ + Y \exp_d(\theta S) \bar{c} \]

where $x_+ := \text{vec}(x_{+,0}, \ldots, x_{+,d}) \in \mathbb{C}^{(d+1)r}$ with

\[ x_{+,j} := \sum_{i=1}^r \tilde{a}_{i,j} z_{\ell} + \sum_{\ell=1}^u \tilde{b}_{i,j} w_{\ell}, \quad i = 1, \ldots, d, \]

\[ x_{+,0} := -M_0^{-1} \left( M_0(Y, S) \bar{c} + \sum_{i=1}^d M_{d+i} \right). \]

Substituting (18) in (19) we obtain $x_{+,0} = \bar{z}$ given in (10). Using (14) and (12) we can express $x_+$ in terms of $\tilde{a}$ and $\tilde{b}$ and we conclude by substituting this expression for $x_+$ in (17).

Remark 5 The assumption (11) can only be satisfied if $r + p \leq n$. This is the case that we consider in this paper. Our focus is on large-scale NEPs and, in Section 5.1, we introduce approximations that avoid $r$ from being large. The hypothesis $\lambda(S) \subseteq \Omega$ is necessary in order to define $M_d(Y, S)$ that is used to compute $\bar{z}$ in equation (10).

2.2 Orthogonalization

In order to expand a TIAR factorization ($\Psi, H_{k-1}$), we need to orthogonalize the tensor structured function $B\psi_k$ (computed using the theorem 4) against the columns of $\Psi_k(\theta)$. The degree of $\Psi_k(\theta)$ is $d-1$ whereas the degree of $B\psi_k(\theta)$ is $d$. In order to perform the orthogonalization, we transform them to the same degree $d$. Starting from (5) we can rewrite $\Psi_k$ as

\[ \Psi_k(\theta) = P_{d-1}(\theta) \left( \sum_{\ell=1}^r a_{..., \ell} \otimes z_{\ell} + \sum_{\ell=1}^u b_{..., \ell} \otimes w_{\ell} \right) + \frac{Y S^d C}{d!} \theta^d + Y \exp_d(\theta S) C. \]

We define

\[ E := \frac{W^H Y S^d C}{d!}, \quad a_{d,j,\ell} := 0, \quad b_{d,j,\ell} := e_{\ell, j}, \quad \ell = 1, \ldots, r + 1, \]

for $j = 1, \ldots, k$. Since $\text{span}(W) = \text{span}(Y)$ and, since $W$ is orthogonal, we have that $Y = W W^H Y$. Hence, using this relation and (21), the function $\Psi_k$ in (20) can be expressed as

\[ \Psi_k(\theta) = P_d(\theta) \left( \sum_{\ell=1}^r a_{..., \ell} \otimes z_{\ell} + \sum_{\ell=1}^u b_{..., \ell} \otimes w_{\ell} \right) + Y \exp_d(\theta S) C \]
Theorem 6 (Orthogonalization) Let \((Z,W,Y,S) \in \mathbb{C}^{n \times r} \times \mathbb{C}^{m \times p} \times \mathbb{C}^{n \times p} \times \mathbb{C}^{p \times k}\) be the matrices and \((a,b,c), (\bar{a},\bar{b},\bar{c}) \in \mathbb{C}^{d \times k \times r} \times \mathbb{C}^{d \times k \times p} \times \mathbb{C}^{p \times k}\) the coefficients that represent \(\psi(\theta)\) given in (3) and \(\Psi_k(\theta)\) given in (5). Let
\[
h = \sum_{\ell=1}^{r}(a_{...})^H \alpha_{...} + \sum_{\ell=1}^{p}(b_{...})^H \beta_{...} + \sum_{i=d}^{\infty} c_i^H \left(S_i^H Y_i Y_i^H S_i\right)^{1/2} \bar{c},
\]
(22)
The orthogonal complement of \(\psi(\theta)\) against the columns of \(\Psi_k(\theta)\) is
\[
\psi^\perp(\theta) = P_{d-1}(\theta) \left(\sum_{i=1}^{r} a_{...}^\perp \otimes z_i + \sum_{\ell=1}^{p} b_{...}^\perp \otimes w_{\ell}\right) + Y \exp_{d-1}(\theta S) c^\perp
\]
where
\[
ce^\perp = \bar{c} - Ch,
\]
(23a)
\[
a_{...}^\perp = \bar{a}_{...} - a_{...} h, \quad \ell = 1, \ldots, r,
\]
(23b)
\[
b_{...}^\perp = \bar{b}_{...} - b_{...} h, \quad \ell = 1, \ldots, p,
\]
(23c)
The vector \(h\) contains the orthogonalization coefficients, i.e., \(h_j = \langle \psi_j, \psi \rangle\). Moreover, given
\[
\beta := \sqrt{\|b^\perp\|^2 + \|a^\perp\|^2 + \sum_{i=d}^{\infty} \left(c_i^\perp\right)^H \left(S_i^H Y_i Y_i^H S_i\right) c^\perp / (i!)^2},
\]
(24)
it holds \(\|\psi^\perp\| = \beta\).

Proof Let us define \(h_j := \langle \psi_j, \psi \rangle\) for \(j = 1, \ldots, k\). The orthogonal complement, computed with the Gram–Schmidt process, is \(\psi^\perp(\theta) = \psi(\theta) - \Psi_k(\theta) h\). Using the Observation 3 we obtain directly (23).

We express \(\psi(\theta)\) as (16) and, the columns of \(\Psi_k\) as
\[
\psi_j(\theta) = P_{d-1}(\theta) x_{(j)} + Y \exp_{d-1}(\theta S) c_j
\]
(25)
where \(x_{(j)} := \text{vec}(x_0^{(j)}, \ldots, x_{d-1}^{(j)}) \in \mathbb{C}^{dh}\), with
\[
x_{(j)} := \sum_{i=1}^{r} \bar{a}_{i+1,j} \otimes z_i + \sum_{\ell=1}^{p} \bar{b}_{i+1,j} \otimes w_{\ell}, \quad i = 0, \ldots, d-1.
\]
(26)

By applying [12, equation (4.32)] we obtain
\[
h_j = \sum_{i=0}^{d-1} (x_{(j)}^H x_i + c_j^H \sum_{i=d}^{\infty} \left(S_i^H Y_i Y_i^H S_i\right) c^\perp / (i!)^2) \bar{c}, \quad j = 1, \ldots, k.
\]
(27)
We now substitute (15) and (26) in (27) and, by using the orthogonormality of the vectors \(z_1, \ldots, z_r, w_1, \ldots, w_p\), we find that
\[
h_j = \sum_{\ell=1}^{r} (a_{...})^H \bar{a}_{...} + \sum_{\ell=1}^{p} (b_{...})^H \bar{b}_{...} + \sum_{i=d}^{\infty} c_i^H \left(S_i^H Y_i Y_i^H S_i\right) c^\perp / (i!)^2 \bar{c}, \quad j = 1, \ldots, k.
\]
Those are the elements of the right–hand side of (22). Using that \(\|\psi^\perp\|^2 = \langle \psi^\perp, \psi^\perp \rangle\), and repeating the same reasoning, we have
\[
\|\psi^\perp\|^2 = \sum_{\ell=1}^{r} (a_{...})^H a_{...} + \sum_{\ell=1}^{p} (b_{...})^H b_{...} + \sum_{i=d}^{\infty} c_i^H \left(S_i^H Y_i Y_i^H S_i\right) c^\perp / (i!)^2 \bar{c},
\]
which proves (24).
2.3 A TIAR expansion algorithm in finite dimension

One algorithmic component common in many restart procedures is the expansion of an Arnoldi–type factorizations. The standard way to expand Arnoldi–type factorizations (as, e.g., described in [28, Section 3]) involves the computation of the action of the operator/matrix and orthogonalization. We now show how we can carry out an expansion of the infinite dimensional TIAR factorization (2) by only using operations on matrices and vectors (of finite dimension).

In the previous subsections we presented the action of the operator $\mathcal{B}$ and orthogonalization for tensor structured functions (3). These results can be directly combined to expand the TIAR factorization. The resulting algorithm is summarized in Algorithm 1. The action of the operator $\mathcal{B}$, described in Theorem 4, is expressed in Steps 2–4. Step 5 corresponds to increasing the degree of the TIAR factorization as described in (20) and (21). The orthogonalization of the new function is carried out by using Theorem 6 and it is expressed in Steps 6–7. We can truncate the sum in (9), (24) and (22) analogously to [12]. Step 5 corresponds to increasing the degree of the TIAR factorization as described in (20) and (21). Due to the representation of $Ψ_k$ as tensor structured function, the expansion with one column corresponds to an expansion of all the coefficients representing $Ψ_k$. This expansion is visualized in Figure 1.

\textbf{Algorithm 1:} Expand TIAR factorization (tensor structured functions)

| input : A TIAR factorization $(Ψ_{k+1}, H_k)$ represented by $(Z, W, Y, S) \in \mathbb{C}^{n \times r} \times \mathbb{C}^{n \times p} \times \mathbb{C}^{n \times p} \times \mathbb{C}^{p \times p}$ and $(a, b, c) \in \mathbb{C}^{d \times k \times r} \times \mathbb{C}^{d \times k \times p} \times \mathbb{C}^{p \times k}$. |
| output: A TIAR factorization $(Ψ_{k+1}, H_{k+1})$ represented by $(Z, W, Y, S) \in \mathbb{C}^{n \times r} \times \mathbb{C}^{n \times p} \times \mathbb{C}^{n \times p} \times \mathbb{C}^{p \times p}$ and $(a, b, c) \in \mathbb{C}^{\delta \times m \times r} \times \mathbb{C}^{\delta \times m \times p} \times \mathbb{C}^{n \times m}$ where $\delta = r + m - k$ and $d = d + m - k$. |

1. Set $\tilde{r} = r$, $\tilde{d} = d$
2. for $k = k + 1, 2, \ldots, m$ do
3. \hspace{1em} Compute $\bar{a}$ using (10), where $\bar{a} = a_{k-1}$, $\bar{b} = b_{k-1}$ and $\bar{c} = c_{k}$
4. \hspace{1em} Compute $\psi_{k-1}$ and increase $\tilde{r} = \tilde{r} + 1$
5. \hspace{1em} Set $\tilde{a}$, $\tilde{b}$ and $\tilde{c}$ as in (12) and (14)
6. \hspace{1em} Compute $E$ and expand the tensors $\bar{a}$ and $\bar{b}$ as (21) and increase $\tilde{d} = \tilde{d} + 1$
7. \hspace{1em} Compute $h$ using (22), where $\bar{a} = \bar{a}$, $\bar{b} = \bar{b}$ and $\bar{c} = \bar{c}$
8. \hspace{1em} Compute $a^{k-1}, b^{k-1}, c^{k-1}$ using (23) and $\beta$ using (24) and extend $H_k = \begin{pmatrix} H_{k-1} & h \\ 0 & \beta \end{pmatrix} \in \mathbb{C}^{(\tilde{d}+1) \times \tilde{d}}$
9. \hspace{1em} Expand $c_{k+1} := c^{k-1} / \beta$ and $a_{k+1} := a^{k-1} / \beta$ and $b_{k+1} := b^{k-1} / \beta$. end

3 Restarting for TIAR in an abstract setting

3.1 The Krylov–Schur decomposition for TIAR factorizations

We briefly recall the reasoning for the Krylov–Schur type restarting [28] in an abstract and infinite dimensional setting. We later show that the operations can be carried out with oper-
Fig. 1 Graphical illustration of the expansion of the tensor structured function that represents the TIAR factorization in Algorithm 1.

\[ \begin{align*}
C &= \tilde{c} = c_\perp \\
\alpha &= \tilde{\alpha} = a_\perp \\
b &= \tilde{b} = E = b_\perp
\end{align*} \]

\[ \begin{align*}
Z &= z_{r+1} + 1 = H = h = \beta
\end{align*} \]

\[ \begin{align*}
\Phi_{m+1} &= \Phi_m + \left( \begin{array}{ccc}
R_{1,1} & R_{1,2} & R_{1,3} \\
R_{2,2} & R_{2,3} & R_{3,3} \\
a_1^H & a_2^H & a_3^H
\end{array} \right)
\end{align*} \] (28)
where \( \Psi_{m+1} = [\Psi_m^m, \psi_{m+1}] \). The matrix \( P \) is selected in a way that the matrix \( R^{(p,1)} \in \mathbb{C}^{p \times p} \) contains the converged Ritz values, the matrix \( R^{(p,2)} \in \mathbb{C}^{(p-p) \times (p-p)} \) contains the wanted Ritz values and the matrix \( R^{(m-p)} \in \mathbb{C}^{(m-p) \times (m-p)} \) contains the Ritz values that we want to purge.

From (28) we find that

\[
\mathcal{P}\Psi = \Psi_{p+1} \begin{pmatrix} R_{1,1} & R_{1,2} \\ a_1^H & a_2^H \end{pmatrix}
\]

(29)

where \( \Psi_{p+1} = [\Psi_m^m, \psi_{p+1}] = [\bar{\Psi}_p, \psi_{m+1}] \).

Using a product of Householder reflectors, we compute a matrix \( Q \) such that

\[
\mathcal{P}\Psi = \Psi_{p+1} \begin{pmatrix} R_{1,1} & F \\ a_1^H & H \end{pmatrix}
\]

(30)

where \( \Psi_{p+1} = \Psi_{p+1}[Q] \in \mathbb{C}^{(m+1) \times (m+1)} = [\bar{\Psi}_p Q, \psi_{m+1}] \).

Since we want to lock the Ritz values in the matrix \( R_{1,1} \), we replace in (30) the vector \( a_1 \) with zeros, which introduces an error \( \mathcal{O}(||a_1||) \). With this approximation, (30) is the wanted TIAR factorization of length \( p \).

**Observation 7** In the TIAR factorization (30), \( (\Psi_p, R_{1,1}) \) is an invariant pair, i.e., \( \mathcal{P}\Psi_p = \Psi_p, R_{1,1} \). Moreover \( (\Psi_p(0), R_{1,1}^{-1}) \) is invariant of the original NEP in the sense of [16, Definition 1], see [12, Theorem 2.2].

### 3.2 Two structured restarting approaches

The standard restart approach for TIAR using Krylov-Schur type restarting, as described in the previous section, involves expansions and manipulations of the TIAR factorization. Due linearity of tensor structured functions described in Observation 3, the manipulations for \( \Psi_m \) leading to \( \Psi_p \) can be directly carried out on the coefficients representing \( \Psi_m \). Unfortunately, due to the implicit representation of \( \Psi_m \), the memory requirements are not substantially reduced since the basis matrix \( Z_0 \in \mathbb{C}^{m \times r} \) is not modified in the manipulations. The size of the basis matrix \( Z \) is the same before and after the restart.

We propose two ways of further exploiting the structure of the functions in order to avoid a dramatic increase in the required memory resources.

- **Semi–explicit restart (Section 4):** An invariant pair can be completely represented by exponentials and therefore does not contribute to the memory requirement for \( Z \). The fact that invariant pairs are exponentials was exploited in the restart in [12]. We show how the ideas in [12] can be carried over to tensor structured functions. More precisely, the adaption of [12] involves restarting the iteration with a locked pair, i.e., only the first \( p_1 \) columns of (30), and a function \( f \) constructed in a particular way. The approach is outlined in Algorithm 2 with details specified in Section 4.

- **Implicit restart (Section 5):** By only representing polynomials, we show that the TIAR factorization has a particular structure such that it can be accurately approximated. This allows us to carry out a full implicit restart, and subsequently approximate the TIAR factorization reducing the size of the matrix \( Z \). The adaption is given in Algorithm 3 with details about the approximation specified in section 4. The approximation of the TIAR factorization in Step 6 of Algorithm 3 is specified Algorithm 4.
Algorithm 2: Semi–explicit restarting for TIAR in operator setting

input: A normalized tensor structured function represented by
\[(Z, W, Y, S) \in \mathbb{C}^{m \times r} \times \mathbb{C}^{n \times p} \times \mathbb{C}^{p \times \ell} \times \mathbb{C}^{p \times m}\]
output: \( p \) eigenvalues of \( \Phi \)
1 Set \( \Psi^{(j)} = [\psi], H^{(j)} \) empty matrix of size \( 1 \times 0 \) and \( j = 1 \)
while \( p_l \leq p \) do
2 Expand the the TIAR factorization \( \{\Psi^{(j)}, H^{(j)}\} \) to length \( m \) using algorithm 1
3 Compute the \( p_l \) converged Ritz pairs and \( P, R_l, a \), given in (28)
4 Compute the matrices \( Q, F, H \) and \( \beta \), given in (30)
5 Lock the invariant pair \( \bar{\Psi} = \Psi^{(j)} P \bar{a} \) and \( R_{l,1} \)
6 Select \( f \) and compute \( \bar{f} \) the orthogonal complement with respect \( \Phi \)
7 Set \( \Psi^{(j+1)} = [\Psi, f] \) and \( H^{(j+1)} = \begin{pmatrix} R_{l,1} & \ell \\ 0 & 0 \end{pmatrix} \) and \( j = j + 1 \)
end
8 Return the eigenvalues of \( R_{l,1} \)

Algorithm 3: Implicit restart for TIAR in operator setting

input: A normalized tensor structured function represented by
\[(Z, 0, 0, 0) \in \mathbb{C}^{m \times r} \times \mathbb{C}^{n \times p} \times \mathbb{C}^{p \times \ell} \times \mathbb{C}^{p \times m}\]
output: \( p \) eigenvalues of \( \Phi \)
1 Set \( \Psi^{(j)} = [\psi], H^{(j)} \) empty matrix of size \( 1 \times 0 \) and \( j = 1 \)
while \( p_l \leq p \) do
2 Expand the the TIAR factorization \( \{\Psi^{(j)}, H^{(j)}\} \) to length \( m \) using algorithm 1
3 Compute the \( p_l \) converged Ritz pairs and \( P, R_l, a \), given in (28)
4 Compute the matrices \( Q, F, H \) and \( \beta \), given in (30)
5 Set \( \Psi^{(j+1)} = [\psi, 0] \) and \( H^{(j+1)} = \begin{pmatrix} R_{l,1} & \ell \\ 0 & 0 \end{pmatrix} \)
6 Approximation of TIAR factorization, algorithm 4
end
7 Return the eigenvalues of \( R_{l,1} \)

4 Tensor structure exploitation for the semi–explicit restart

The IAR restart approach in [12] is based representing functions as sums of exponentials and polynomials. An attractive feature of that approach is that the invariant pairs can be exactly represented, and locking can be efficiently incorporated. Due to the explicit storage of polynomial coefficients in [12], the approach still requires considerable memory. We here show that, by representing the functions implicitly as tensor structured functions (3), we can maintain the advantages of [12] but improve performance (both in memory and CPU-time). This construction is equivalent to [12], but more efficient.

The expansion of the TIAR factorization with tensor structured functions (as described in Algorithm 1) combined with the locking procedure (as described in Section 3.1) results in Algorithm 2. Steps 3-7 follow the procedure described in [12] adapted for tensor structured functions. In Step 6 the function used as a new starting function can be extracted from the tensor structured representation as follows, completely equivalent with [12].

\[ f(\theta) = \bar{\Psi} \exp(\theta S) e_{p_l+1}, \quad S := \left( R_{l,1} F \right)^{-1}, \quad \bar{\Psi} := \Psi^{(0)} P I_{k,p} Q. \]
We can use Observation 3 to compute \( \tilde{Y} \) from the tensor structured function representation. With \( M := P_{l_k,p} Q \), we obtain
\[
\tilde{Y} := \Psi_n(0) M
\]
\[
= P_d(0) \left( \sum_{i=1}^r a_{1:;i} M \otimes z_i + \sum_{i=1}^p b_{1:;i} M \otimes w_i \right) + Y \exp_d(0) C
\]
\[
= \sum_{i=1}^r a_{1:;i} M \otimes z_i + \sum_{i=1}^p b_{1:;i} M \otimes w_i.
\]

5 Tensor structure exploitation for the implicit polynomial restart

In contrast to the procedure in Section 4, where the main idea was to do locking with exponentials and restart with a factorization of length \( p_d \), we now propose a fully implicit procedure involving a factorization of length \( p \). In this setting we use \( Y = 0 \), i.e., only representing polynomials with the tensor structured functions. This allows us to derive theory for the structure of the coefficient matrix, which shows how to approximate of the TIAR factorization. The algorithm is summarized in Algorithm 3.

The approximation in Step 6 is done in order to avoid the growth in memory requirements for the representation. The approximation technique is derived in the following subsections and summarized in Algorithm 4.

Our approximation approach is based on approximation with a truncated singular value decomposition and a degree reduction. A compression with a truncated singular value decomposition was also made for the compact representations in CORK [34] and TOAR [17]. Our specific setting allows to prove bounds on the error introduced by the approximations (Section 5.1–5.2). We also show the effectiveness by proving a bound on the decay of the singular values (Section 5.3).

We first note the following decay in the magnitude of the elements of the tensor \( a \), which are the coefficients representing \( \Psi_i \).

**Theorem 8** Let \( Z \in \mathbb{C}^{n \times p} \) be a matrix and \( a \in \mathbb{C}^{(k+1) \times (k+1) \times r} \) a set of coefficients that represent the tensor structured function \( \Phi_{k+1} \) and \( H_k \in \mathbb{C}^{(k+1) \times k} \) such that \( (\Phi_{k+1}, H_k) \) is a TIAR factorization. Assume that \( \psi_i(\theta) \) is a constant function, i.e., \( a_{i,1,\ell} = 0 \) if \( i > 1 \). Then
\[
||a_{i:;;i}|| \leq \frac{C}{(i-1)!} \text{ for } i = 1, \ldots, k + 1,
\]
where \( C = k([c_k, c_k+1, \ldots, c_k^{r+1}]) \) is defined in [14, equation (29)] and \( v = \sum_{\ell=1}^r a_{1,1,\ell} z_\ell \).

**Proof** Let \( \Phi_{k+1}(\theta) = (\psi_1(\theta), \psi_2(\theta), \ldots, \psi_k(\theta)) \). Applying theorem 4 with \( Y = 0 \), we obtain
\[
\Phi_{k+1}(\theta) = P_d(\theta) \left( \sum_{\ell=1}^r \tilde{a}_{1:;;\ell} \otimes z_\ell \right)
\]
where
\[
\tilde{a}_{1:;;\ell} := \begin{pmatrix}
a_{1,1,\ell} & a_{1,2,\ell} & a_{1,3,\ell} & \cdots & a_{1,k+1,\ell} \\
a_{2,1,\ell} & a_{2,2,\ell} & a_{2,3,\ell} & \cdots & a_{2,k+1,\ell} \\
a_{3,1,\ell} & a_{3,2,\ell} & a_{3,3,\ell} & \cdots & a_{3,k+1,\ell} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
a_{p,1,\ell} & a_{p,2,\ell} & a_{p,3,\ell} & \cdots & a_{p,k+1,\ell}
\end{pmatrix}
\]
Since $\Psi_{k+1}(\theta_1, \ldots, \theta_k)$ forms a TIAR factorization, it holds span $(\Phi_{k+1}) = \text{span} (\Psi_{k+1})$. Therefore it exists an invertible matrix $R \in \mathbb{C}^{(k+1) \times (k+1)}$ such that $\Phi_{k+1}R = \Psi_{k+1}$. Using the Observation 3 we have that $a_{i,\ell} = \tilde{a}_{i,\ell}R$ and by submultiplicativity of the Euclidean norm we have that for $i = 1, \ldots, k+1$

$$\|a_{i,\ell}\| = \|\tilde{a}_{i,\ell}R\| \leq \|\tilde{a}_{i,\ell}\|\|R\|. \quad (32)$$

Using the structure of $\tilde{a}_{i,\ell}$ we have

$$\|\tilde{a}_{i,\ell}\|^2 \leq \frac{1}{(i-1)!} \sum_{j=1}^{k+1} \tilde{a}_{i,j,\ell}^2 \leq \frac{1}{(i-1)!} \sum_{j=1}^{k+1} \tilde{a}_{i,j,\ell}^2 = \frac{\|\tilde{a}_{i,\ell}\|^2}{(i-1)!}. \quad (33)$$

Combining (32) and (33) we obtain

$$\|a_{i,\ell}\| \leq \frac{\|\tilde{a}_{i,\ell}\|\|R\|}{(i-1)!} = \frac{\|a_{i,\ell}R^{-1}\|}{(i-1)!} \leq \frac{\|a_{i,\ell}\|\|R\|}{(i-1)!} \kappa(R).$$

Setting $C := \kappa(R)$ we obtain (31). It remains to show that $C = \kappa([v; C_{k+1}v, \ldots, C_{k+1}^{k+1}v])$. Due to the equivalence of TIAR and IAR and the companion matrix interpretation of IAR [14, theorem 6], we have that TIAR is equivalent to use the Arnoldi method on the matrix $C_{k+1}$ and starting vector $v = \sum_{j=1}^{r} a_{1,j}z_{j}$. More precisely, the relation $\Phi_{k+1}R = \Psi_{k+1}$ can be written in terms of vectors as $VR = W$ where the first column of $V$ and $W$ is $v = \sum_{j=1}^{r} a_{1,j} \otimes z_{j}$ and $W = [v; C_{k+1}v, \ldots, C_{k+1}^{k+1}v]$. By using the orthogonality of $V$ we conclude that that $C = \kappa([v; C_{k+1}v, \ldots, C_{k+1}^{k+1}v])$.

**Observation 9** In the numerical simulations, we observed a very fast decay of the norm of the matrices $\|a_{i,\ell}\|$ with respect $i$. Unfortunately, the condition number of the Krylov matrix $[v; C_{k+1}v, \ldots, C_{k+1}^{k+1}v]$ grows at least exponentially with respect $k$. See [4] and the references therein for a characterization of similar properties for the linear case. The bound provided by Theorem 8 is pessimistic and not sharp; we use it only for theoretical purposes.

**Corollary 10** If $\Psi(\theta)$ given in (5) satisfies $\|a_{i,\ell}\| \leq C/(i-1)!$ for $i = 1, \ldots, d$, then for any matrix $M$, $\tilde{\Psi}(\theta) = \Psi(\theta)M$ satisfies $\|\tilde{a}_{i,\ell}\| \leq \tilde{C}/(i-1)!$ for $i = 1, \ldots, d$ where $\tilde{C} \leq C\kappa(M)$.

Algorithm 3 corresponds to a constant starting function in Step 1, i.e., $\psi(\theta)$ is such that $a_{1,\ell} = 0$ if $i > 1$. As consequence of Theorem 8, after expansion of the TIAR factorization in Step 2, we have that the norm of $a_{1,\ell}$ satisfies (31). By using the Corollary 10 we obtain that this relation is preserved also after the Step 5. In conclusion, in the Algorithm 3 the coefficients of Krylov basis $\Psi^{(j)}$ always fulfill (31). This allows us to introduce an approximation of the TIAR factorization.

5.1 Approximation by SVD compression

Given a TIAR factorization with basis function $\Psi$, we now show (in the following theorem) how we can approximate the basis function with less memory, by using a thinner $Z$-matrix. The theorem also shows how this approximation influences the influence the approximation $\Psi$. Moreover, we show that the approximation has a small impact also on the residual of the TIAR factorization.
Theorem 11 Let \( a \in \mathbb{C}^{(d+1) \times k \times r} \), \( Z \in \mathbb{C}^{n \times r} \) be the coefficients that represent the tensor structured function (3) and suppose that \((\Psi_{k+1}, H_k)\) is a TIAR factorization. Suppose \(|z| \leq R \subseteq \Omega \) with \( R > 1 \). Let \( A := [A_1, \ldots, A_d] \in \mathbb{C}^{r \times dm} \) be the unfolding of the tensor \( a \) in the sense that \( A_i = (a_{i:,})^T \). Given the singular value decomposition of \( A \)

\[
A = [U_1, U] \text{diag}(\Sigma_1, \Sigma)[V_1^H, \ldots, V_d^H] \\
\Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_r) \\
\Sigma = \text{diag}(\sigma_{r+1}, \ldots, \sigma_d),
\]

(34)

let

\[
\tilde{Z} := ZU_1, \quad \tilde{A}_i := \Sigma V_i^H \quad i = 1, \ldots, d + 1.
\]

(35)

and \( \Psi_{k+1} \) the tensor structured function defined by the coefficients \( \tilde{a} \in \mathbb{C}^{(d+1) \times (k+1) \times r} \) and \( \tilde{Z} \in \mathbb{C}^{n \times r} \), with \( \tilde{a}_{i:,} := \tilde{A}_i^T \). Then,

\[
\|\Psi_{k+1} - \Psi_{k+1}\| \leq \sqrt{(d + 1)(k + 1)} \|M\| \|\sigma_{k+1}\| \quad (36a)
\]

\[
\|H_k - \tilde{H}_k\| \leq \sqrt{(C_d + C_r)} \|\sigma_{k+1}\|
\]

(36b)

with

\[
C_d := \chi + \log(d + 1) + (d + 1)\|H_k\|
\]

\[
C_r := \|M_r\| \left( \chi + \log(s + 1) \right) \max_{1 \leq i \leq s} \|M_i\| \max_{|\lambda| \leq R} |\lambda|
\]

where \( \chi \approx 0.57721 \) is the Euler–Mascheroni constant and

\[
s := \min \left\{ s \in \mathbb{N} : \frac{C(d - s)}{R^s} \leq \sigma_f \right\}
\]

where \( C \) is defined in Corollary 10.

Proof The proof of (36a) is based on construction a difference function \( \Psi_{k+1} = \Psi_{k+1} - \Psi_{k+1} \) as follows. We define

\[
\tilde{Z} := ZU_1, \quad \tilde{A}_i := \Sigma V_i^H,
\]

\[
\tilde{X}_i := ZA_{i+1}, \quad \tilde{X}_i := \tilde{Z}\tilde{A}_i, \quad \tilde{X}_i := \tilde{Z}\tilde{A}_i,
\]

\[
X := [X_0^H, \ldots, X_d^H]^H, \quad \tilde{X} := [\tilde{X}_0^H, \ldots, \tilde{X}_d^H]^H, \quad \tilde{X} := [\tilde{X}_0^H, \ldots, \tilde{X}_d^H]^H.
\]

then we can express \( \Psi_{k+1} = P_d(\theta)X \) where \( \Psi_{k+1}(\theta) = P_d(\theta)\tilde{X} \) and \( \Psi_{k+1}(\theta) = P_d(\theta)\tilde{X} \). By using (6) and \( \|\tilde{X}_i\|_{F}^2 = \|Z\tilde{A}_{i+1}\|_{F}^2 \leq (k + 1)\|Z\tilde{A}_{i+1}\|_{2}^2 = (k + 1)\|\tilde{A}_{i+1}\|_{2}^2 = (k + 1)\|\Sigma V_i\|_{2}^2 \leq (k + 1)\|\Sigma\|_{2}^2 = (k + 1)\|\sigma_{k+1}\|_2^2 \) we obtain

\[
\|\tilde{\Psi}_{k+1}\|_{F}^2 = \sum_{i=0}^{d} \|\tilde{X}_i\|_{F}^2 \leq (d + 1)(k + 1)\|\sigma_{k+1}\|^2
\]

which proves (36a).

In order to show (36b) we first use that \( \|P_{k+1} - \tilde{P}_{k+1}\| = \|\tilde{P}_{k+1} - \tilde{P}_{k+1}\| \) since \((\Psi_{k+1}, H_k)\) is a TIAR factorization and subsequently use the decay of \( A_i \) and analyticity of \( M \) as follows. For notational convenience we define

\[
Y_i := \tilde{X}_{i,k+1}^i, \quad \text{for } i = 0, \ldots, d - 1
\]

(37)
and \( Y := [Y_0^H \cdots Y_d^H]^H \) such that we can express \( \hat{\Psi}_k(\theta) = P_{d-1}(\theta)Y \). Using [12, theorem 4.2] for each column of \( \hat{\Psi}_k(\theta) \), we get
\[
\hat{\Psi}_k(\theta) = P_d(\theta)Y,
\]
with
\[
Y_{+,i+1} := \frac{Y_i}{i+1} \quad \text{for} \quad i = 0, \ldots, d-1 \quad \text{and} \quad Y_{+,0} := -M_0^{-1}\sum_{i=1}^d M_i Y_{+,i}.
\]

By definition and (6) we have
\[
\| \hat{\Psi}_k - \hat{\Psi}_{k+1} H_s \| = \| P_d(\theta)Y_k - P_d(\theta)\hat{X} H_s \| = \| Y_k - \hat{X} H_s \|_F.
\]

Moreover, by using the two-norm bound of the Frobenius norm, (37) and that \( \| \hat{X} \| \leq \sigma_{r+1} \),
\[
\| Y_k - \hat{X} H_s \|_F \leq \sqrt{k} \sum_{i=0}^d \| Y_{+,i} - \hat{X}_i H_s \|_F \leq \sqrt{k} \sum_{i=0}^d (\| Y_{+,i} \| + \| \hat{X}_i \| \| H_s \|)
\]
\[
= \sqrt{k} \left( \| Y_{+,0} \| + \sum_{i=1}^d \| Y_{+,i} \| + \sum_{i=0}^d \| \hat{X}_i \| \| H_s \| \right)
\]
\[
\leq \sqrt{k} \left( \| Y_{+,0} \| + \sum_{i=1}^d \frac{\sigma_{r+1}}{i} + \sum_{i=0}^d \sigma_{r+1} \| H_s \| \right)
\]
\[
\leq \sqrt{k} \| Y_{+,0} \| + \sigma_{r+1} (\gamma + \log(d+1) + (d+1)\| H_s \|).
\]

In the last inequality we use the Euler-Mascheroni inequality where \( \gamma \) is defined in [1, Formula 6.1.3]. It remains to bound \( \| Y_{+,0} \|. \) By using the definition of \( Y_{+,0} \) and again applying the Euler-Mascheroni inequality we have that
\[
\| Y_{+,0} \| \leq \| M_0^{-1} \| \sum_{i=1}^d \| M_i \| \frac{\| \hat{X}_{+,i} \|}{i} \leq \| M_0^{-1} \| \left( \sum_{i=1}^d \| M_i \| \frac{\| \hat{X}_{+,i} \|}{i} + \sum_{i=1}^d \| M_i \| \frac{\| \hat{X}_{+,i} \|}{i} \right)
\]
\[
\leq \| M_0^{-1} \| \left( \sigma_{r+1} (\gamma + \log(s+1)) \max_{\lambda \leq s} \| M_i \| + \sum_{i=1}^d \| M_i \| \frac{\| \hat{X}_{+,i} \|}{i} \right),
\]
(39)

As consequence of the Cauchy integral formula
\[
\| M_i \| \frac{\| \hat{X}_{+,i} \|}{i} \leq \| M_i \| \frac{\| A_i \|}{i} \leq C \frac{\| M_i \|}{i!} \leq C \frac{\max | M(\lambda) |}{R^s}.
\]
(40)

By substituting (40) in (39) we obtain
\[
\| Y_{+,0} \| \leq \sigma_{r+1} \| M_0^{-1} \| (\gamma + \log(s+1)) \max_{\lambda \leq s} \| M_i \| + \| M(\lambda) \| c^{d-s} \frac{R^s}{R^s}
\]
\[
\leq \sigma_{r+1} \| M_0^{-1} \| (\gamma + \log(s+1)) \max_{\lambda \leq s} \| M_i \| + \| M(\lambda) \| c^{d-s} \frac{R^s}{R^s}.
\]
(41)

We reach the conclusion (36b) from the combination of (41) in (38).
5.2 Approximation by reducing the degree

Another approximation which reduces the storage requirements can be done by truncating the polynomial in $\Psi_k$. The following theorem illustrated the approximation properties of this approach.

**Theorem 12** Let $a \in \mathbb{C}^{(d+1) \times (k+1) \times r}$, be the representation of the tensor structured function $\Psi_{k+1}$ with $Y = 0$. For $\bar{d} \leq d$ let

$$\bar{\Psi}_{k+1}(\theta) := P_d(\theta) \left( \sum_{i=1}^{d} \bar{a}_{i,j}\ell \otimes z^i \right)$$

(42)

where $\bar{a}_{i,j}\ell$ for $i = 1, \ldots, \bar{d}$, $j = 1, \ldots, k + 1$ and $\ell = 1, \ldots, r$. Then

$$\|\Psi_{k+1} - \bar{\Psi}_{k+1}\| \leq C \sqrt{k+1} \frac{(d-\bar{d})}{d!}$$

(43)

$$\|\bar{\Psi}_k - \Psi_{k+1} H_k\| \leq C \sqrt{k+1} \left( \max_{d+1 \leq d} \|M_i\| \right) \|M_0^{-1}\| \frac{d-\bar{d}}{(d+1)!}$$

(44)

**Proof** We define $X_i := Z A_{i+1}$ for $i = 0, \ldots, d$ and $X := [X_0^T, \ldots, X_d^T]$ and $\bar{X} := [\bar{X}_0^T, \ldots, \bar{X}_d^T]$ such that $\Psi_{k+1}(\theta) = P_d(\theta)X$ and $\bar{\Psi}_{k+1}(\theta) = P_d(\theta)\bar{X}$. We have

$$\|\Psi_{k+1}(\theta) - \bar{\Psi}_{k+1}(\theta)\|^2 = \sum_{i=d+1}^d \|X_i\|_F^2 = \sum_{i=d+1}^d \|A_i\|_F^2 \leq (k+1) \sum_{i=d+1}^d \|A_i\|^2.$$  

(45)

By using Corollary 10 we obtain (43).

By definition $\Psi_k(\theta) = \Psi_{k+1}(\theta)I_{k+1,k}$ and $\bar{\Psi}_k(\theta) = \bar{\Psi}_{k+1}(\theta)I_{k+1,k}$, using the observation 3, if we define $Y_i := X_{d,k+1}$ for $i = 0, \ldots, d-1$ and $Y := [Y_0^H \cdots Y_{d-1}^H]^H$ and $\bar{Y} := [\bar{Y}_0^H \cdots \bar{Y}_{d-1}^H]^H$ we can express $\Psi_k(\theta) = P_{d-1}(\theta)Y$ and $\bar{\Psi}_k(\theta) = P_{d-1}(\theta)\bar{Y}$.

Using [12, theorem 4.2] for each column of $\Psi_k(\theta)$ and $\bar{\Psi}_k(\theta)$, we get $\bar{\Theta}_k(\theta) = P_{d}(\theta)Y_+$ and $\bar{\Theta}_k(\theta) = P_{d}(\theta)\bar{Y}_+$ with

$$Y_{+,i+1} := \frac{Y_i}{i+1} \quad \text{for} \quad i = 0, \ldots, d-1 \quad \text{and} \quad Y_{+,0} := -M_0^{-1} \sum_{i=1}^d M_i Y_{+,i}$$

(46)

$$\bar{Y}_{+,i+1} := Y_{+,i+1} \quad \text{for} \quad i = 0, \ldots, d-1 \quad \text{and} \quad \bar{Y}_{+,0} := -M_0^{-1} \sum_{i=1}^d M_i Y_{+,i}$$

(47)

In our notation, the fact that $(\Psi_{k+1}, H_k)$ is a TIR factorization, can be expressed as $P_d(\theta)Y_+ = P_d(\theta)X H_k$, which implies that the monomial coefficients are equal, i.e.,

$$Y_{+,i} = X_i H_k \quad \text{for} \quad i = 0, \ldots, d.$$  

(45)

Hence, from (6) we have

$$\|\bar{\Theta}_k - \bar{\Psi}_{k+1} H_k\|^2 = \|P_d(\theta)Y_+ - P_d(\theta)X H_k\|^2$$

$$= \|Y_+ - X H_k\|^2_F$$

$$= \|\bar{Y}_{+,0} - X_0 H_k\|^2_F + \sum_{i=1}^d \|Y_{+,i} - X_i H_k\|^2_F$$

$$= \|\bar{Y}_{+,0} - X_0 H_k\|^2_F$$
In the last step we applied (45). Moreover, by again using (45), we have

\[ Y_{+0} - X_0 H_k = -M_0^{-1} \sum_{i=1}^d M_i Y_{i0} - X_0 H_k \]

\[ = -M_0^{-1} \sum_{i=1}^d M_i \bar{Y}_{i0} - M_0^{-1} \sum_{i=d+1}^d M_i Y_{i0} - X_0 H_k \]

\[ = \bar{Y}_{+0} - X_0 H_k - M_0^{-1} \sum_{i=d+1}^d M_i \bar{X}_{i0} Y_{i0} H_k. \]

Therefore

\[ \| \bar{Y}_{+0} - X_0 H_k \| \leq \| M_0^{-1} \| \sum_{i=d+1}^d \| M_i \| \| A_i \|. \]

We obtain (44) by using the Corollary 10.

**Remark 13** The approximation given in Theorem 12 can only be effective under the condition that \( \max_{1 \leq i \leq d} \| M_i \| / (d+1)! \) is small. In particular this condition is satisfied if the Taylor coefficients \( \| M_i \| / i! \) present a fast decay. This condition correspond to have the coefficients of the power series expansion of \( M(\lambda) \) that are decaying to zero.

**Algorithm 4:** Approximation of TIAR factorization

| input | A TIAR factorization \( (\Psi_{i+1}, H_i) \) expressed by \( Y, W \in \mathbb{C}^{n \times p}, a \in \mathbb{C}^{d \times k \times r}, b \in \mathbb{C}^{d \times k \times p} \) and \( C \in \mathbb{C}^{r \times k} \) |
| output | A TIAR factorization \( (\Psi_{i+1}, H_i) \) expressed by \( Y, W \in \mathbb{C}^{n \times p}, a \in \mathbb{C}^{d \times k \times r}, b \in \mathbb{C}^{d \times k \times p} \) and \( C \in \mathbb{C}^{r \times k} \) |
| 1 Compute the SVD decomposition given in (34) partitioned such that \( \delta_i \leq \varepsilon \) |
| 2 Set \( r = i, Z = \tilde{Z}, a_{i1,:} = \tilde{A}_i^\top \) for \( i = 1, \ldots, d \) given in (35) |
| 3 Compute \( d \) such that |
| \( \left( \max_{1 \leq i \leq d} \| M_i \| \right) \| M_0^{-1} \| \frac{d-d}{(d+1)!} < \varepsilon \) |
| 4 Reduce the size of the tensor \( a_{i1,:} = a_{i1,:} \) and set \( d = \tilde{d} \) |

5.3 The fast decay of singular values

Finally, as a further justification for our approximation procedure, we now show how fast the singular values decay. The fast decay in the singular values illustrated below justifies the effectiveness of the truncation in Section 5.1.

**Lemma 14** Let \( Z \in \mathbb{C}^{n \times r}, a \in \mathbb{C}^{d \times (k+1) \times i} \) represent the tensor structured function \( \Psi_{i+1} \) as in (5) with \( Y = W = 0 \) and let \( H_k \in \mathbb{C}^{(k+1) \times k} \) be a Hessenberg matrix such that \( (\Psi_{i+1}, H_k) \) is TIAR factorization. Then, the tensor \( a \) is generated by \( d \) vectors, in the sense that each vector \( a_{i,j,:} \) for \( i = 1, \ldots, d \) and \( j = 1, \ldots, k \) can be expressed as linear combination of the vectors \( a_{11,:} \) and \( a_{1,k,:} \) for \( i = 1, \ldots, k-d \) and \( j = 1, \ldots, k \).
The proof is based on induction over the length $k$ of the TIAR factorization. The result is trivial if $k = 1$. Suppose the result holds for some $k$. Let $Z \in \mathbb{C}^{n \times (r-1)}$, $a \in \mathbb{C}^{(d-1) \times k \times r}$ represent the tensor structured function $\Psi_k$ and let $H_{k-1} \in \mathbb{C}^{k \times (k-1)}$ an Hessenberg matrix such that $(\Psi_k, H_{k-1})$ is TIAR factorization. If we expand the TIAR factorization $(\Psi_k, H_{k-1})$ by using the Algorithm 1, more precisely by using (14b) and (23b), we obtain

$$\beta_{a_{i+1,k+1},i} = \frac{a_{i,k}}{i} - \sum_{j=1}^{k} h_j a_{i,j}$$

$i = 1, \ldots, d - 1$.

We reach the condition of the theorem by induction.

**Theorem 15** Under the same hypothesis of Lemma 14, let $A$ be the unfolding of the tensor $a$ in a sense that $A = [A_1, \ldots, A_d]$ such that $A_i := (a_{i,:})^T$. We have the following decay in the singular values

$$\sigma_i \leq C \frac{d - R - k + 2}{(R - k + 1)!}$$

$i = R + 1, \ldots, d$,

where $k \leq R \leq d$ and $C$ is the constant provided by Corollary 10.

**Proof** We define the matrix $\tilde{A} := [A_1, \ldots, A_{R-k+1}, 0, \ldots, 0] \in \mathbb{C}^{r \times dk}$. Notice that the columns of the matrices $A$ and $\tilde{A}$ correspond to the vectors $a_{i,j}$. In particular, using the Lemma 14, we have that rank($A_1$) = $k$ whereas rank($A_j$) = 1 if $j \leq d - k + 1$ otherwise rank($A_j$) = 0. Then we have that rank($A$) = $d$ and rank($\tilde{A}$) = $R$. Using Weyl’s theorem [10, Corollary 8.6.2] and Corollary 10 we have for $i \geq R + 1$

$$\sigma_i \leq \|A - \tilde{A}\| \leq \sum_{i=R-k+2}^{d} \|A_i\| \leq \sum_{i=R-k+2}^{d} \frac{C}{(i-1)!} \leq C \frac{d - R - k + 2}{(R - k + 1)!}$$

6 Complexity analysis

We presented above two different restarting strategies: the structured semi–explicit restart and the implicit restart. They have different performance for different problems and we have not been able to conclusively determine if one is better than the other. The best choice of the restarting strategy appears to depend on many problem properties. It may be convenient to test both methods on the same problem. We now discuss the general performances, in terms of complexity and stability. The complexity discussion is based on the assumption that the complexity of the action of $M_0^{-1}$ is neglectable in comparison to the other parts of the algorithm.

**Complexity of expanding the TIAR factorization**

Independently of which restarting strategy is used, the main computational effort of the algorithms 2 and 3 is the expansion of a TIAR factorization described in algorithm 1. The essential computational effort of the algorithm 1 is the computation of $\tilde{z}$, given in equation (10). This operation has complexity $O(drn)$ for each iteration. In both restarting strategies $r$ and $d$ are, in general, not large due to the way they are automatically selected in the algorithm 4.
Complexity of the restarting strategies

After an implicit restart we obtain a TIAR factorization of length $p$, whereas after a semi-explicit restart, we obtain a TIAR factorization of length $p_\ell$. This means that the semi-explicit restart requires a re-computation phase, i.e., after the restart we need to perform extra $p - p_\ell$ steps in order to have a TIAR factorization of length $p$. If $p - p_\ell$ is large, i.e., if not many Ritz values converged in comparison to the restarting parameter $p$, then the re-computation phase is the essential computational effort of the algorithm. Notice that this is hard to predict since we do not know how fast the Ritz values will converge in advance.

Stability of the restarting strategies

We will illustrate in section 7 that the restarting approaches have different stability properties. The semi-explicit restart tends to be efficient if only a few eigenvalues are wanted, i.e., if $p$ is small. This is due to the fact that we impose the structure in the starting function. On the other hand, the implicit restart requires a thick restart in order to be stable in several situations, see corresponding discussions for the linear case in [19, chapter 8]. Then $p$ has to be large enough in a sense that at each restart the $p$ wanted Ritz values have the corresponding residual not small. This leads to additional computational and memory resources.

If we use the semi-explicit restart, then the computation of $\tilde{z}$, in equation (10), involves the term $M_d(Y, S)$. This quantity can be computed in different ways. In the simulations we must choose between (8) or (9). The choice influences the stability of the algorithm. In particular if one eigenvalue of $S$ is close to $\partial\Omega$ and $M(\lambda)$ is not analytic in $\partial\Omega$, the series (9) converges slowly and in practice overflow can occur. In such situations, (8) is preferable. Notice that it is not always it is possible to use (8) since many problems cannot be formulated as (7) with small $q$.

Memory requirements of the restarting strategies

From a memory point of view, the essential part of the semi-explicit restart is the storage of the matrices $Z$ and $Y$, that is $O(nm + np)$. In the implicit restart the essential part is the storage of the matrix $Z$ and requires $O(nr_{\text{max}})$ where $r_{\text{max}}$ denotes the maximum value that the variable $r$ takes in the algorithm. The size of $r_{\text{max}}$ is not predictable since it depends on the svd–approximation introduced in algorithm 4. Since in each iteration of the algorithm 1 the variable $r$ is increased, it holds $r_{\text{max}} \geq m - p$. Therefore, in the optimal case where $r_{\text{max}}$ takes the lower value, the two methods are comparable in terms of memory requirements. Notice that, the semi-explicit restart requires less memory and has the advantage that the required memory is problem independent.

7 Numerical experiments

7.1 Delay eigenvalue problem

In order to illustrate properties of the proposed restart methods and advantages in comparison to other approaches, we carried out numerical simulations 1 for solving the delay eigenvalue problem (DEP). More precisely, we consider the DEP associated with the delay

---

1 All simulations were carried out with Intel octa core i7-3770 CPU 3.40GHz and 16 GB RAM
differential equation defined in [15, sect 4.2] with \( \tau = 1 \). By using a standard second order finite difference discretization, the DEP is formulated as

\[
M(\lambda) = -\lambda^2 I + \lambda A_1 + A_0 + e^{-\lambda} A_2.
\]

We show how the proposed methods perform in terms of \( m \), the maximum length of the TIAR factorization, and \( p \), the number of wanted Ritz values.

Table 1a and Table 1b show the advantages of our semi–explicit restart approach in comparison to the equivalent method described in [12]. Our new approach is faster in terms of CPU–time and can solve larger problems due to the memory efficient representation of the Krylov basis.

Table 2a and Table 2b show the effectiveness of approximations introduced in Section 5.1 and 5.2 in comparison to the corresponding restart procedure without approximations. In particular, in Algorithm 4 we consider a drop tolerance \( \varepsilon = 10^{-14} \). Since the DEP is defined by entire functions, the power series coefficients decay to zero and, according to Remark 13, the approximation by reducing the degree is expected to be effective. By approximating the TIAR factorization, the implicit restart requires less resources in terms of memory and CPU–time and can solve larger problems.

We now illustrate the differences between the semi–explicit and the implicit restart. More precisely, we show how the parameters \( m \) and \( p \) influence the convergence of the Ritz values with respect the number of iterations. The convergence of the semi–explicit restart appears to be slower in the semi–explicit restart when \( p \) is not sufficiently large. See Figure 2. The convergence speed of both restarting strategies is comparable for a larger \( m \) and \( p \). See Figure 3.

In practice, the performance of the two restarting strategies corresponds to a trade-off between CPU–time and memory. In particular, due to the fact that we impose the structure, the semi–explicit restart does not have a growth in the polynomial part at each restart and therefore requires less memory. On the other hand, for this problem, the semi–explicit restart appears to be slower in term of CPU–time. See Figure 2 and 3.
Fig. 3 Implicit and semi-explicit restart for DEP of size $n = 40401$ with $m = 40$, $p = 10$ and restart=4

Table 1 DEP, Semi-explicit restart
| Size  | m  | p  | restart | compression | no compression |
|-------|----|----|---------|-------------|---------------|
| 10201 | 20 | 5  | 7       | 31.41 s  | 65.38 MB      |
| 40401 | 30 | 14 | 11      | 258.92 MB |               |
| 160801| 10 | 7  | 3       | 1.01 GB   |               |
| 641601| 30 | 10 | 7       | 4.02 GB   |               |
| 1002001| 50 | 10 | 4       | 749.18 MB |               |

Table 2 DEP, Implicit restart
| Size  | m  | p  | restart | compression | no compression |
|-------|----|----|---------|-------------|---------------|
| 10201 | 20 | 5  | 7       | 31.41 s  | 65.38 MB      |
| 40401 | 30 | 14 | 11      | 258.92 MB |               |
| 160801| 10 | 7  | 3       | 1.01 GB   |               |
| 641601| 30 | 10 | 7       | 4.02 GB   |               |
| 1002001| 50 | 10 | 4       | 749.18 MB |               |

7.2 Waveguide eigenvalue problem

In order to illustrate how the performance depends on the problem properties, we now consider a NEP defined by functions with branch point and branch cut singularities. More precisely, we consider the waveguide eigenvalue problem (WEP) described in [13, Section 5.1] after the Cayley transformation. In this problem, $\Omega$ is the unit disc and there are branch point singularities in $\partial \Omega$. Thus, due to the slow convergence of the power series, in the semi-explicit restart we have to use (9) in order to compute $M_d(Y, S)$. This also implies that the approximation by reducing the degree is not expected to be effective since the power series coefficients of $M(\lambda)$ are not decaying to zero.

In analogy to the previous subsection, we carried out numerical simulations in order to compare the semi-explicit and the implicit restart.
With Figure 4 and 5, we illustrate the performance of the two restarting approaches with respect the choice of the parameters $m$ and $p$. When $p$ is sufficiently large, the residual in the semi–explicit restart appears to stagnate after the first restart whereas it decreases in a regular way in the implicit restart. See Figure 4. On the other hand, when $p$ is small, the behavior of the residual appear to be specular. See Figure 5. This is due to the fact that semi–explicit restart imposes the structure on $p$ vectors which is not beneficial when they do not contain eigenvector approximations.

It is known that this specific problem has two eigenvalues. Therefore, in order to reduce the CPU–time and the memory resources, the the number of wanted Ritz values $p$ should be selected small. As consequence of the above discussion, we conclude that the semi–explicit restart is the best restarting strategy for this problem.
Table 3 WEP, Implicit restart

| a) m = 40, p = 20, restart=4 | Implicit | Semi-explicit |
|-------------------------------|----------|--------------|
| Size            | CPU      | Memory       | CPU      | Memory       |
| 2703            | 5.33s    | 2.4 MB       | 13.02s   | 2.4 MB       |
| 40803           | 9.78s    | 9.4 MB       | 22.54s   | 9.4 MB       |
| 161603          | 29.15s   | 36.7 MB      | 1m14s    | 36.7 MB      |
| 643203          | 1m50s    | 143.0 MB     | 4m01s    | 143.0 MB     |
| 1006009         | 7m27s    | 579.1 MB     | 11m54s   | 579.1 MB     |

| b) m = 20, p = 4, restart=6 | Implicit | Semi-explicit |
|-------------------------------|----------|--------------|
| Size            | CPU      | Memory       | CPU      | Memory       |
| 2703            | 5.09s    | 1.5 MB       | 5.17s    | 0.9 MB       |
| 40803           | 8.83s    | 5.9 MB       | 10.74s   | 3.7 MB       |
| 161603          | 25.93s   | 23.0 MB      | 24.11s   | 14.3 MB      |
| 643203          | 6m31s    | 91.2 MB      | 1m20s    | 56.7 MB      |
| 1006009         | 10m25s   | 566.8 MB     | 8m57s    | 352.4 MB     |

8 Concluding remarks and outlook

In this work we have derived an extension of the TIAR algorithm and two restarting strategies. Both restarting strategies are based on approximating the TIAR factorization. In other works on the IAR–method it has been proven that the basis matrix contains a structure that allows exploitations, e.g. for NEPs with low rank structure in the coefficients [33]. An investigation about the combination of the approximations of the TIAR factorization with such structures of the NEP seems possible but deserve further attention.

Although the framework of TIAR and restarted TIAR is general, a specialization of the methods to the NEP is required in order to efficiently solve the problem. More precisely, an efficient computation procedure for computing (10) is required. This is a nontrivial task for many application and requires problem specific research.

Acknowledgements We gratefully acknowledge the support of the Swedish Research Council under Grant No. 621-2013-4640.

References

1. Abramowitz, M., Stegun, I.: Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol. 55. Courier Corporation (1964)
2. Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., van der Vorst, H.: Templates for the solution of algebraic eigenvalue problems: a practical guide, vol. 11. Siam (2000)
3. Bai, Z., Su, Y.: SOAR: A second-order Arnoldi method for the solution of the quadratic eigenvalue problem. SIAM J. Matrix Anal. Appl. 26(3), 640–659 (2005)
4. Beckermann, B.: The condition number of real vandermonde, Krylov and positive definite Hankel matrices. Numer. Math. 85(4), 553–577 (2000)
5. Betcke, M.M., Voss, H.; Restarting projection methods for rational eigenproblems arising in fluid-solid vibrations. Math. Model. Anal. 13(2), 171–182 (2008)
6. Betcke, M.M., Voss, H.: Restarting iterative projection methods for Hermitian nonlinear eigenvalue problems with minmax property. Numer. Math., 1–34 (2016)
7. Betcke, T., Higham, N.J., Mehrmann, V., Schröder, C., Tisseur, F.: NLEVP: A collection of nonlinear eigenvalue problems. Tech. rep., Manchester Institute for Mathematical Sciences (2011)
8. Betcke, T., Voss, H.: A Jacobi–Davidson-type projection method for nonlinear eigenvalue problems. Future Gener. Comp. Sy. 20(3), 363–372 (2004)
9. Effenberger, C.: Robust solution methods for nonlinear eigenvalue problems. Ph.D. thesis, École polytechnique fédérale de Lausanne (2013)
10. Golub, G.H., Van Loan, C., Charles, F.: Matrix computations, vol. 3. JHU Press (2012)
11. Güttel, S., Van Beeumen, R., Meerbergen, K., Michiels, W.: NLEIGS: A class of fully rational Krylov methods for nonlinear eigenvalue problems. SIAM J. Sci. Comput. 36(6), A2842–A2864 (2014)
12. Jarlebring, E., Meerbergen, K., Michiels, W.: Computing a partial Schur factorization of nonlinear eigenvalue problems using the infinite Arnoldi method. SIAM J. Matrix Anal. Appl. 35(2), 411–436 (2014)
13. Jarlebring, E., Mele, G., Runborg, O.: The waveguide eigenvalue problem and the tensor infinite Arnoldi method. Tech. rep. (2015). ArXiv:1503.02096
14. Jarlebring, E., Michiels, W., Meerbergen, K.: A linear eigenvalue algorithm for the nonlinear eigenvalue problem. Numer. Math. 122(1), 169–195 (2012)
15. Jarlebring, E., Poloni, F.: Iterative methods for the delay Lyapunov equation with T-Sylvester preconditioning. Tech. rep. (2015). ArXiv:1507.02100
16. Kressner, D.: A block Newton method for nonlinear eigenvalue problems. Numer. Math. 114(2), 355–372 (2009)
17. Kressner, D., Roman, J.E.: Memory-efficient Arnoldi algorithms for linearizations of matrix polynomials in Chebyshev basis. Numer. Linear Algebra Appl. 21(4), 569–588 (2014)
18. Lancaster, P., Psarrakos, P.: On the pseudospectra of matrix polynomials. SIAM J. Matrix Anal. Appl. 27(1), 115–129 (2005)
19. Lehoucq, R.B.: Analysis and implementation of an implicitly restarted Arnoldi iteration. Ph.D. thesis, Rice University (1995)
20. Lehoucq, R.B., Sorensen, D.C.: Deflation techniques for an implicitly restarted Arnoldi iteration. SIAM J. Matrix Anal. Appl. 17(4), 789–821 (1996)
21. Mackey, D.S., Mackey, N., Mehl, C., Mehrmann, V.: Structured polynomial eigenvalue problems: Good vibrations from good linearizations. SIAM J. Matrix Anal. Appl. 28(4), 1029–1051 (2006)
22. Mackey, D.S., Mackey, N., Tisseur, F.: Polynomial eigenvalue problems: Theory, computation, and structure. In: Numerical Algebra, Matrix Theory, Differential-Algebraic Equations and Control Theory, pp. 319–348. Springer (2015)
23. Meerbergen, K.: Locking and restarting quadratic eigenvalue solvers. SIAM J. Sci. Comput. 22(5), 1814–1839 (2001)
24. Meerbergen, K.: The quadratic Arnoldi method for the solution of the quadratic eigenvalue problem. SIAM J. Matrix Anal. Appl. 30(4), 1463–1482 (2008)
25. Mehrmann, V., Voss, H.: Nonlinear eigenvalue problems: A challenge for modern eigenvalue methods. GAMM-Mitt. 27(2), 121–152 (2004)
26. Morgan, R.: On restarting the Arnoldi method for large nonsymmetric eigenvalue problems. Math. Comp. 65(215), 1213–1230 (1996)
27. Neumaier, A.: Residual inverse iteration for the nonlinear eigenvalue problem. SIAM J. Numer. Anal. 22(5), 914–923 (1985)
28. Stewart, G.W.: A Krylov–Schur algorithm for large eigenproblems. SIAM J. Matrix Anal. Appl. 23(3), 601–614 (2002)
29. Su, Y., Bai, Z.: Solving rational eigenvalue problems via linearization. SIAM J. Matrix Anal. Appl. 32(1), 201–216 (2011)
30. Szylf, D., Vecharynski, E., Xue, F.: Preconditioned eigensolvers for large-scale nonlinear Hermitian eigenproblems with variational characterizations. I. Interior eigenvalues. SIAM J. Sci. Comput. 37(6), A2969–A2997 (2015)
31. Szylf, D., Xue, F.: Preconditioned eigensolvers for large-scale nonlinear Hermitian eigenproblems with variational characterizations. II. Extreme eigenvalues. Math. Comp. , (2016)
32. Tisseur, F., Meerbergen, K.: The quadratic eigenvalue problem. SIAM Rev. 2, 235–286 (2001)
33. Van Beeumen, R., Jarlebring, E., Michiels, W.: A rank-exploiting infinite Arnoldi algorithm for nonlinear eigenvalue problems. Numer. Linear Algebra Appl. , (2016)
34. Van Beeumen, R., Meerbergen, K., Michiels, W.: Compact rational Krylov methods for nonlinear eigenvalue problems. SIAM J. Matrix Anal. Appl. 36(2), 820–838 (2015)
35. Voss, H.: A maximin principle for nonlinear eigenvalue problems with application to a rational spectral problem in fluid-solid vibration. Appl. Math. 48(6), 607–622 (2003)
36. Voss, H.: An Arnoldi method for nonlinear eigenvalue problems. BIT 44(2), 387–401 (2004)
37. Voss, H.: Nonlinear eigenvalue problems. In: L. Hogben (ed.) Handbook of Linear Algebra, Second Edition, no. 164 in Discrete Mathematics and Its Applications. Chapman and Hall/CRC (2013)
38. Zhang, Y., Su, Y.: A memory-efficient model order reduction for time-delay systems. BIT 53(4), 1047–1073 (2013)