Networks and Heterogeneous Media

https://www.aimspress.com/journal/nhm

NHM, 20(1): 15–34.
DOI: 10.3934/nhm.2025002
Received: 21 October 2024
Revised: 20 December 2024
Accepted: 27 December 2024
Published: 06 January 2025

Research article

Theory of Krylov subspace methods based on the Arnoldi process with inexact inner products

Meng Su[1], Chun Wen[1,*], Zhao-Li Shen[2,*] and Stefano Serra-Capizzano[3]

[1] School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China

[2] College of Science, Sichuan Agricultural University, Ya'an, Sichuan 625000, China

[3] Department of Science and High Technology, University of Insubria, Como Campus 22100 , Italy

* **Correspondence:** Email: wchun17@163.com, szlxiaoyao@163.com.

Abstract: Several Krylov subspace methods are based on the Arnoldi process, such as the full orthogonalization method (FOM), GMRES, and in general all the Arnoldi-type methods. In fact, the Arnoldi process is an algorithm for building an orthogonal basis of the Krylov subspace. Once the inner products are performed inexactly, which cannot be avoided due to round-off errors, the orthogonality of Arnoldi vectors is lost. In this paper, we presented a new analysis framework to show how the inexact inner products influence the Krylov subspace methods that are based on the Arnoldi process. A new metric was developed to quantify the inexactness of the Arnoldi process with inexact inner products. In addition, the proposed metric can be used to approximately estimate the loss of orthogonality in the practical use of the Arnoldi process. The discrepancy in residual gaps between Krylov subspace methods employing inexact inner products and their corresponding exact counterparts was discussed. Numerical experiments on several examples were reported to illustrate our theoretical findings and final observations were presented.

Keywords: Arnoldi process; Krylov subspace methods; inexact inner products; residual gaps

1. Introduction

Consider the solution of large and sparse linear systems of the form

$$Ax = b, \tag{1.1}$$

where $A \in \mathbb{R}^{n\times n}$ is a nonsingular matrix, $b \in \mathbb{R}^n$ is a given right-hand vector, and $x \in \mathbb{R}^n$ is the unknown vector that needs to be computed. Many numerical methods have been discussed to solve the linear system (1.1) in [1–4]. For example, the basic iterative methods (Jacobi, Gauss-Seidel, Successive Over-Relaxation (SOR)), the Krylov subspace methods (FOM, GMRES, Conjugate Gradient (CG), Bi-Conjugate Gradient (BCG)), the incomplete LU (ILU) factorization preconditioning techniques, etc. In addition to solving linear systems like Eq (1.1), Krylov subspace methods, such as the Arnoldi-type algorithm, are widely used for approximating eigenvalues and eigenvectors of large matrices. These methods provide an efficient way to compute a subset of eigenpairs for large and sparse systems.

In recent years, the application of inexact Krylov subspace methods based on the Arnoldi process has obtained extensive attention (e.g., [5–10]). Specifically speaking, if only the matrix-vector multiplication is performed inexactly in the Arnoldi process and the other arithmetic operations are carried out with exact computations, then we obtain an inexact Arnoldi relation

$$(A + \mathcal{E}_m)V_m = V_{m+1}H_m, \quad V_{m+1}^T V_{m+1} = I_{m+1}, \tag{1.2}$$

where $I_{m+1} \in \mathbb{R}^{(m+1)\times(m+1)}$ is an identity matrix, $\mathcal{E}_m \in \mathbb{R}^{n\times n}$ is the perturbation matrix evolving with m, $H_m \in \mathbb{R}^{(m+1)\times m}$ is an upper-Hessenberg matrix, and $V_m \in \mathbb{R}^{n\times m}$ is a basis matrix whose column vectors are the basis vectors of the Krylov subspace $\mathcal{K}_m(A + \mathcal{E}_m, v_1)$ with $v_1 = b/\|b\|$. For more details, refer to [8, 11, 12]. It is clear that the Krylov subspace is generated by $A + \mathcal{E}_m$ and not by the coefficient matrix A in the considered linear system (1.1). However, the matrix V_m has the same property as in the exact computations, that is, its column vectors are mutually orthogonal. One natural question is how to control the accuracy of the inexact matrix-vector multiplication within Krylov subspace methods in such a way that the efficiency is preserved, without affecting the convergence or the final achieved accuracy in a substantial manner. There is a vast literature related to this question, (e.g., [5–9, 13] and the references therein). Relaxation strategies proposed in [5, 13, 14] have shown to be surprisingly effective for a range of different Krylov subspace methods, and their empirical results have been confirmed by Van Den Eshof and Sleijpen in [7]. They argue that the success of a relaxation strategy depends on the underlying way that the Krylov subspace is constructed and not on the optimality properties for the residuals. Simoncini and Szyld [8] have provided a general framework to understand the inexact Krylov subspace methods for the solution of the linear system (1.1), as well as for certain eigenvalue calculations.

Rounding errors in inner products represent a concern in Krylov subspace methods like Arnoldi-type and GMRES. Due to finite floating-point precision, these errors cause discrepancies between exact and computed inner products, leading to a loss of orthogonality in the Krylov basis and affecting the algorithm's stability and convergence. Accumulated errors can degrade performance. In the Arnoldi process, for instance, the classical Gram-Schmidt (CGS) method suffers from loss of orthogonality due to inner product rounding errors, which led to the development of reorthogonalized Gram-Schmidt algorithms to mitigate these effects. If only the inner products are performed inexactly in the Arnoldi process, and the other arithmetic operations are carried out exactly, then it has an inexact Arnoldi relation as showed in [10]

$$AV_m = V_{m+1}H_m, \quad V_{m+1}^T V_{m+1} = I_{m+1} + F_{m+1}, \tag{1.3}$$

where $F_{m+1} \in \mathbb{R}^{(m+1)\times(m+1)}$ is an error matrix and $V_m \in \mathbb{R}^{n\times m}$ is a basis matrix with its column vectors being the basis vectors of the Krylov subspace $\mathcal{K}_m(A, v_1)$, $v_1 = b/\|b\|$. It is easy to find that the Krylov subspace is generated by A in Eq (1.3), not the matrix $A + \mathcal{E}_m$ as given in Eq (1.2). The basis matrix-sequence $\{V_m\}$ in Eq (1.2) holds the orthogonality, but the basis matrix-sequence $\{V_m\}$ in Eq (1.3) loses the orthogonality in general. Similarly, one question is how the accuracy of the inexact inner products computed in Krylov subspace methods influences the convergence rate or the final accuracy achieved by the iterates. Gratton et al. [10] have investigated this question and derived implementable conditions by bounding the loss of orthogonality in GMRES with inexact inner products, where they have adapted techniques used in the rounding-error analysis of the modified Gram-Schmidt (MGS) algorithm [15–17] and of the MGS-GMRES algorithm [18–20].

In our view a further analysis is of interest concerning the properties of the Arnoldi relation, when using inexact inner products in the Arnoldi process. Here a new analysis framework is provided for understanding the extent to which inexact inner products affect the convergence and final accuracy of several Arnoldi-based Krylov subspace methods. Our work focuses on the following items:

- We introduce a transition matrix between the basis vectors of the Krylov subspace $\mathcal{K}_m(A, v_1)$ with exact inner products and the basis vectors of the Krylov subspace $\mathcal{K}_m(A, v_1)$ with inexact inner products, and we analyze its properties.
- We discuss the perturbation of the upper-Hessenberg matrix caused by inexact inner products operations.
- We investigate the effect of an approximately computed inner products on the convergence and accuracy of several Krylov subspace methods, by providing bounds regarding their residual gaps, including FOM, GMRES, and Arnoldi-type methods.

The remaining sections of the paper are structured as follows. In Section 2, we begin with a standard Arnoldi process, and establish a connection between it and the Arnoldi process with inexact inner products by introducing a transition matrix. In Section 3, we discuss the perturbation of the upper-Hessenberg matrix and then analyze the residual gaps between exact and inexact inner products versions of the FOM, GMRES, and Arnoldi-type methods. In Section 4, we present several numerical experiments to give practical evidence of our theoretical results. Finally, conclusions and final observations are reported in Section 5.

2. Analysis for the Arnoldi process with inexact inner products

It is well known that many Krylov subspace methods are based on the Arnoldi process, such as FOM, GMRES, and Arnoldi-type methods. The standard Arnoldi process can be described as in Algorithm 1 [1,4].

If all the arithmetic operations in the standard Arnoldi process are computed exactly, then we have the following Arnoldi relation

$$AV_m^{(e)} = V_{m+1}^{(e)} H_m^{(e)}, \quad V_{m+1}^{(e)T} V_{m+1}^{(e)} = I_{m+1}, \tag{2.1}$$

where $V_m^{(e)} = [v_1^{(e)}, v_2^{(e)}, \cdots, v_m^{(e)}] \in \mathbb{R}^{n\times m}$ is a basis matrix whose column vectors are the basis vectors of the Krylov subspace $\mathcal{K}_m(A, v)$ and $H_m^{(e)} \in \mathbb{R}^{(m+1)\times m}$ is an upper-Hessenberg matrix. For the Arnoldi process with inexact inner products, as given in [10], we could suppose the inexact inner products computed as

$$h_{ij} = v_i^{(e)T} w + \eta_{ij}, \quad h_{j+1,j} = \sqrt{w^T w + \eta_{j+1,j}}, \tag{2.2}$$

where η_{ij} and $\eta_{j+1,j}$ can be regarded as random errors. Comparing Eqs (1.3) and (2.1), it is easy to see that the Arnoldi vectors in $V_m^{(e)}$ obtained by exact inner products are orthogonal, while the Arnoldi vectors in V_m obtained by inexact inner products have lost their orthogonality. However, it is worth noting that both the basis matrix $V_m^{(e)}$ and V_m are generated from the same Krylov subspace $\mathcal{K}_m(A, v)$, that is, both the Arnoldi vectors $v_1^{(e)}, v_2^{(e)}, \cdots, v_m^{(e)}$ and $v_1, v_2, \cdots, v_m$ represent a basis for the Krylov subspace $\mathcal{K}_m(A, v)$. Since we observe only a change of basis, there necessarily exists a unique transition matrix $T_m \in \mathbb{R}^{m\times m}$ such that $V_m = V_m^{(e)} T_m$. Based on this fact, we deduce the following interesting propositions.

Algorithm 1 Arnoldi process with modified Gram-Schmidt

Require: $A \in \mathbb{R}^{n\times n}, v = b \in \mathbb{R}^n, m \in \mathbb{N}$

$\beta^{(e)} = \|v\|$
$v_1^{(e)} = v/\beta^{(e)}$
for $j = 1$ to m **do**
    $w = Av_j^{(e)}$
    for $i = 1$ to j **do**
        $h_{ij}^{(e)} = v_i^{(e)T} w$
        $w = w - h_{ij}^{(e)} v_i^{(e)}$
    end for
    $h_{j+1,j}^{(e)} = \|w\|$
    if $h_{j+1,j}^{(e)} = 0$ **then**
        break
    end if
    $v_{j+1}^{(e)} = w/h_{j+1,j}^{(e)}$
end for

Proposition 2.1. Assume that the Arnoldi process and its inexact inner products version do not break. Under the above notations, we have $H_m^{(e)} = T_{m+1} H_m T_m^{-1}$ and $T_{m+1}\begin{bmatrix} I_m \\ O \end{bmatrix} = \begin{bmatrix} T_m \\ O \end{bmatrix}$.

Proof. Using the relation $V_m = V_m^{(e)} T_m$ and $V_{m+1} = V_{m+1}^{(e)} T_{m+1}$, the equation $AV_m = V_{m+1} H_m$ in Eq (1.3) can be rewritten as

$$AV_m^{(e)} T_m = V_{m+1}^{(e)} T_{m+1} H_m. \tag{2.3}$$

Since the transition matrix is invertible. Multiplying both sides of Eq (2.3) by T_m^{-1} to the right and by $V_{m+1}^{(e)T}$ to the left, we have

$$V_{m+1}^{(e)T} A V_m^{(e)} = V_{m+1}^{(e)T} V_{m+1}^{(e)} T_{m+1} H_m T_m^{-1}. \tag{2.4}$$

Substituting Eq (2.1) into Eq (2.4), it is easy to obtain

$$H_m^{(e)} = T_{m+1} H_m T_m^{-1}. \tag{2.5}$$

On the other hand, because of the fact that V_m and $V_m^{(e)}$ are the top m columns of V_{m+1} and $V_{m+1}^{(e)}$, respectively, we deduce the expressions

$$V_m = V_{m+1}\begin{bmatrix} I_m \\ O \end{bmatrix}, \quad V_m^{(e)} = V_{m+1}^{(e)}\begin{bmatrix} I_m \\ O \end{bmatrix}. \tag{2.6}$$

Using the relations $V_m = V_m^{(e)} T_m$ and $V_{m+1} = V_{m+1}^{(e)} T_{m+1}$ again, we find

$$V_{m+1}^{(e)} T_{m+1}\begin{bmatrix} I_m \\ O \end{bmatrix} = V_{m+1}^{(e)}\begin{bmatrix} T_m \\ O \end{bmatrix}. \tag{2.7}$$

Since $V_{m+1}^{(e)}$ is a column full rank matrix, we obtain the result

$$T_{m+1}\begin{bmatrix} I_m \\ O \end{bmatrix} = \begin{bmatrix} T_m \\ O \end{bmatrix}. \tag{2.8}$$

From Eqs (2.5) and (2.8), we complete the proof of the current proposition.

Proposition 2.2. Assume that the Arnoldi process and its inexact inner products version do not break. Under the above notations, T_{m+1} is an upper triangular matrix whose main diagonal entries are all positive.

Proof. We would prove this proposition by mathematical induction. In the case of $m = 0$, the result is obvious. Suppose this proposition holds when $m = k - 1 (k \geq 2)$, that is, T_k is an upper triangular matrix whose main diagonal entries are all positive. Then, when $m = k$, we have

$$T_{k+1} = V_{k+1}^{(e)T} V_{k+1} = \begin{bmatrix} V_k^{(e)T} \\ v_{k+1}^{(e)T} \end{bmatrix} \begin{bmatrix} V_k & v_{k+1} \end{bmatrix} = \begin{bmatrix} T_k & V_k^{(e)T} v_{k+1} \\ v_{k+1}^{(e)T} V_k & v_{k+1}^{(e)T} v_{k+1} \end{bmatrix}. \tag{2.9}$$

From Eq (1.3), we observe $AV_k = V_{k+1}H_k$ and thus

$$v_{k+1} = \frac{1}{h_{k+1,k}}(Av_k - h_{k,k}v_k - \cdots - h_{1,k}v_1). \tag{2.10}$$

Since $v_1^{(e)}, v_2^{(e)}, \cdots, v_{k+1}^{(e)}$ are orthogonal basis vectors of the Krylov subspace $\mathcal{K}_{k+1}(A, v)$. It is clear that the vector $v_{k+1}^{(e)}$ is orthogonal to the Krylov subspace $\mathcal{K}_k(A, v) = span(v_1^{(e)}, v_2^{(e)}, \cdots, v_k^{(e)}) = span(v_1, v_2, \cdots, v_k)$. Consequently we deduce

$$v_{k+1}^{(e)T} V_k = O, \quad v_{k+1}^{(e)T} v_{k+1} = \frac{1}{h_{k+1,k}} v_{k+1}^{(e)T} A v_k.$$

In addition, from the relation $v_k = V_k e_k$, where $e_k = [0, \cdots, 0, 1]^T \in \mathbb{R}^{k\times 1}$, we obtain

$$v_{k+1}^{(e)T} v_{k+1} = \frac{1}{h_{k+1,k}} v_{k+1}^{(e)T} A V_k^{(e)} T_k e_k. \tag{2.11}$$

Furthermore, using Eq (2.1), the equality $AV_k^{(e)} = V_{k+1}^{(e)} H_k^{(e)}$ holds and we have $v_{k+1}^{(e)T} A V_k^{(e)} = h_{k+1,k}^{(e)} e_k^T$. Then, Eq (2.11) becomes

$$v_{k+1}^{(e)T} v_{k+1} = \frac{h_{k+1,k}^{(e)}}{h_{k+1,k}} e_k^T T_k e_k.$$

By the induction hypothesis, T_k is an upper triangular matrix with its main diagonal entries are all positive, we have $e_k^T T_k e_k > 0$. Combining with $h_{k+1,k}^{(e)} > 0$ and $h_{k+1,k} > 0$, we infer $v_{k+1}^{(e)T} v_{k+1} > 0$. Hence, the matrix T_{k+1} is also an upper triangular matrix with its main diagonal entries are all positive, which means that the claimed thesis holds when $m = k$. By mathematical induction, the proof is completed.

The relation $V_{m+1} = V_{m+1}^{(e)} T_{m+1}$ can be seen as a QR factorization of the matrix V_{m+1}. Generally speaking, the QR factorization of a matrix is of course non-unique. However, once the R factor is an upper triangular matrix with all positive diagonal entries, the QR factorization of any column full rank matrix would be unique [21]. According to Proposition 2.2, as the R factor, the matrix T_{m+1} is an upper triangular matrix whose main diagonal entries are all positive. Therefore, the QR factorization of the matrix V_{m+1} is unique. Also, the basis matrix $V_{m+1}^{(e)}$ could be obtained by QR factorization of V_{m+1} that does not involve inner products operation, such as Givens rotation QR factorization [22] or Cholesky QR factorization [23].

Moreover, according to Proposition 2.2, we can further explain why the basis vectors $v_1, v_2, \cdots, v_{m+1}$ generated from the Krylov subspace $\mathcal{K}_{m+1}(A, v)$ with inexact inner products are impossible to be orthogonal, i.e., the basis matrix V_{m+1} loses its orthogonality as given in Eq (1.3). Because if V_{m+1} is an orthogonal basis matrix, then from the uniqueness of its QR factorization, it is easy to have $V_{m+1} = V_{m+1}^{(e)}$ and $T_{m+1} = I_{m+1}$. If this case $V_{m+1} = V_{m+1}^{(e)}$ holds true, then there is no inexact computations in the Arnoldi process. Hence, the Arnoldi process with inexact inner products inevitably leads to the basis matrix V_{m+1} loses orthogonality. And the discrepancy $V_{m+1} - V_{m+1}^{(e)}$ can be interpreted as the perturbation caused by inexact inner products applied to $V_{m+1}^{(e)}$ obtained through the standard Arnoldi process. It is natural to use $\|V_{m+1} - V_{m+1}^{(e)}\|$ to quantify the perturbation induced by inexact inner products on the Arnoldi process. Additionally, because of the unitary invariance of the spectral matrix norm $\|\cdot\|$, that is, the induced l_2-norm, we have

$$\|V_{m+1} - V_{m+1}^{(e)}\| = \|V_{m+1}^{(e)}(T_{m+1} - I_{m+1})\| = \|T_{m+1} - I_{m+1}\|.$$

It follows that $\|T_{m+1} - I_{m+1}\|$ can be regarded as a new metric to quantify the impact of inexact inner products on the Arnoldi process.

On the other hand, it is worth mentioning that $\|T_{m+1} - I_{m+1}\|$ not only can be used to estimate the inexactness of inexact inner products, but also can be applied to exhibit the loss of orthogonality caused by the inexact inner products, that is, $\|T_{m+1} - I_{m+1}\|$ has a relationship with the loss of orthogonality $\|V_{m+1}^T V_{m+1} - I_{m+1}\|$ presented in [10]. Again, using the unitary invariance of the spectral norm $\|\cdot\|$, it is easy to see that $\|V_{m+1}^T V_{m+1} - I_{m+1}\| = \|T_{m+1}^T T_{m+1} - I_{m+1}\|$. By the triangle inequality and compatibility of norm, we have

$$\begin{aligned} \left| \|T_{m+1} - I_{m+1}\| - \|T_{m+1}^T T_{m+1} - I_{m+1}\| \right| &\le \|T_{m+1} - T_{m+1}^T T_{m+1}\| \\ &\le \|I_{m+1} - T_{m+1}^T\| \|T_{m+1}\| \\ &\le \|I_{m+1} - T_{m+1}\| \left(\|I_{m+1} - T_{m+1}\| + 1\right). \end{aligned} \tag{2.12}$$

From Eq (2.12), if $\|T_{m+1} - I_{m+1}\|$ is small enough, then $\|T_{m+1} - I_{m+1}\|$ and $\|V_{m+1}^T V_{m+1} - I_{m+1}\|$ are close enough. In other words, to some extent, $\|T_{m+1} - I_{m+1}\|$ can replace $\|V_{m+1}^T V_{m+1} - I_{m+1}\|$ to represent the loss of orthogonality.

3. Analysis for several inexact Krylov subspace methods

According to Proposition 2.1, $H_m^{(e)} = T_{m+1} H_m T_m^{-1}$, it is clear that there exists a multiplicative perturbation with respect to the upper-Hessenberg matrix $H_m^{(e)}$. Since many Krylov subspace methods use the information stored in the upper-Hessenberg matrix to solve the linear system in Eq (1.1), it is significant to study the perturbation of the upper-Hessenberg matrix $H_m^{(e)}$ and then investigate the residual gaps between exact and inexact inner products versions of the FOM, GMRES, and Arnoldi-type methods.

3.1. Perturbation of the upper-Hessenberg matrix

In the present section, we analyze the multiplicative perturbation of the upper-Hessenberg matrix $H_m^{(e)}$ based on the following two lemmas.

Lemma 1. ([24], Corollary 3.1.3) Let $A \in \mathbb{C}^{m\times n}$ be given, and let A_r denote a submatrix of A obtained by deleting a total of r rows and(or) columns from A. Then,

$$\sigma_k(A) \geq \sigma_k(A_r) \geq \sigma_{k+r}(A), \quad k = 1, \cdots, \min\{m, n\} \tag{3.1}$$

where for $X \in \mathbb{C}^{p\times q}$ we set $\sigma_j(X) = 0$ if $j \geq \min\{p, q\}$.

Lemma 2. ([25], Theorem 6.6) Let $A, B \in \mathbb{C}^{m\times n}$, their singular values are

$$\sigma_1 \geq \cdots \geq \sigma_n \geq 0, \quad \tau_1 \geq \cdots \geq \tau_n \geq 0 \tag{3.2}$$

respectively, then $|\sigma_i - \tau_i| \leq \|B - A\|, i = 1, 2, \cdots, n$.

Proposition 3.1. Under the above notations, if $\|T_{m+1} - I_{m+1}\| \leq \epsilon$ ($\epsilon < 1$), then

$$\|H_m - H_m^{(e)}\| \leq \frac{2\epsilon}{1-\epsilon}\|H_m^{(e)}\|. \tag{3.3}$$

Proof. According to Proposition 2.1, we have $H_m = T_{m+1}^{-1} H_m^{(e)} T_m$. Then, the following chain of relationships holds

$$\begin{aligned} \|H_m - H_m^{(e)}\| &= \|T_{m+1}^{-1} H_m^{(e)} T_m - H_m^{(e)}\| \\ &= \|T_{m+1}^{-1} H_m^{(e)} T_m - T_{m+1}^{-1} H_m^{(e)} + T_{m+1}^{-1} H_m^{(e)} - H_m^{(e)}\| \\ &\leq \|T_{m+1}^{-1}\| \|H_m^{(e)}\| \left(\|T_m - I_m\| + \|T_{m+1} - I_{m+1}\|\right). \end{aligned} \tag{3.4}$$

Since $T_m - I_m$ is a submatrix of $T_{m+1} - I_{m+1}$, obtained by deleting the last column and the last row, from Lemma 1, we deduce

$$\|T_m - I_m\| = \sigma_1(T_m - I_m) \leq \sigma_1(T_{m+1} - I_{m+1}) = \|T_{m+1} - I_{m+1}\| \leq \epsilon. \tag{3.5}$$

On the other hand, according to Lemma 2, it has $|\sigma_{min}(T_{m+1}) - 1| \leq \|T_{m+1} - I_{m+1}\| \leq \epsilon$. Then, we have

$$\frac{1}{1+\epsilon} \leq \frac{1}{\sigma_{min}(T_{m+1})} = \|T_{m+1}^{-1}\| \leq \frac{1}{1-\epsilon}. \tag{3.6}$$

Combining the inequalities in Eqs (3.4)–(3.6), we proved the result in Eq (3.3).

Let $f(x) = \frac{2x}{1-x}$, $0 \leq x < 1$. It is easy to see that the function $f(x)$ is monotonically increasing on the interval $[0, 1)$ and $f(0) = 0$. Therefore, if the upper bound ϵ of $\|T_{m+1} - I_{m+1}\|$ decreases, the perturbation of $H_m^{(e)}$ diminishes as well. In particular, when $\epsilon = 0$, there is no perturbation with $H_m^{(e)}$. Next, we employ the result concerning the perturbation of the upper-Hessenberg matrix for analyzing several inexact Krylov subspace methods.

3.2. GMRES with inexact inner products

Based on the standard Arnoldi process, GMRES [26] is a classical method for solving the linear system (1.1). First of all, let us briefly review the GMRES when all the arithmetic operations are performed exactly as given in Algorithm 2. The core idea of GMRES is searching a solution x_m such that

$$\min_{x_m\in\mathcal{K}_m(A,b)} \|b - Ax_m\| = \min_{y\in\mathbb{R}^m} \|V^{(e)}_{m+1}(\beta^{(e)}e_1 - H^{(e)}_m y)\| = \min_{y\in\mathbb{R}^m} \|\beta^{(e)}e_1 - H^{(e)}_m y\|, \tag{3.7}$$

in which $\beta^{(e)} = \|b\|$ and $e_1 = [1, 0, \cdots, 0]^T \in \mathbb{R}^{m+1}$. Finally, the algorithm is reduced to solve the least square problem associated with the relation $H^{(e)}_m y = \beta^{(e)}e_1$.

Algorithm 2 GMRES for solving $Ax = b$

Require: $A \in \mathbb{R}^{n\times n}, b \in \mathbb{R}^n, m \in \mathbb{N}$

1: Obtain $V^{(e)}_m$ and $H^{(e)}_m$ by Algorithm 1.
2: Compute the solution $y^{(e)}$ of the least squares problem $H^{(e)}_m y^{(e)} = \beta^{(e)}e_1$, where $\beta^{(e)} = \|b\|$.
3: Compute $x^{(e)}_m = V^{(e)}_m y^{(e)}$.

In the inexact inner products GMRES, we could not acquire the same result as shown in Eq (3.7) since V_{m+1} is not a orthogonal basis matrix. Therefore, it is flawed that we solve the least square problem $H_m y = \beta e_1$ like in the exact GMRES. Now we analyze the inexact inner products GMRES by perturbation analysis of the least square problem. A lemma regarding the least square problem and its perturbed least square problem is given as follows.

Lemma 3. ([27], Theorem 9.7) Let $Ax = b$ and $\bar{A}x = \bar{b}$ represent least square problem and perturbed least square problem respectively, where A, $\bar{A} \in \mathbb{C}^{m\times n}$ and b, $\bar{b} \in \mathbb{C}^m$. Additionally, x_{LS} and $\bar{x}_{LS}$ are the least square solutions for the above least square problems, respectively, $\Delta A = \bar{A} - A$, and $\Delta b = \bar{b} - b$. Suppose $\|\Delta A\| = \epsilon_A\|A\|$ and $\|\Delta b\| = \epsilon_b\|b\|$, with $Rank(A) = p \le \min\{m, n\}$. If $\kappa_A\epsilon_A < 1$ and $Rank(\bar{A}) = Rank(A)$, then with $\Delta x = \bar{x}_{LS} - x_{LS}$,

$$\|\Delta x\| = \bar{\kappa}_A\left(\epsilon_b\frac{\|b\|}{\|A\|} + \epsilon_A\|x_{LS}\| + \kappa_A\epsilon_A\frac{\|r\|}{\|A\|}\right) + \kappa_A\epsilon_A\|x_{LS}\|, \tag{3.8}$$

where κ_A represents the spectral condition number of A and $\bar{\kappa}_A = \frac{\kappa_A}{1-\kappa_A}$.

Consider the least square problem $H^{(e)}_m y = \beta^{(e)}e_1$ generated from GMRES when all arithmetic operations are performed exactly and the perturbed least square problem $H_m y = \beta e_1$ obtained from GMRES with inexact inner products. Let $y^{(e)}_{LS}$ and y_{LS} represent the least square solutions of the above least square problems respectively. Assume $r^{(e)} = \beta^{(e)}e_1 - H^{(e)}_m y^{(e)}_{LS}$, $\Delta H^{(e)}_m = H^{(e)}_m - H_m$, $\Delta\beta^{(e)}e_1 = \beta^{(e)}e_1 - \beta e_1$, $\Delta y^{(e)}_{LS} = y^{(e)}_{LS} - y_{LS}$, $\kappa_{H^{(e)}_m}$ is the spectral condition number of $H^{(e)}_m$ and $\bar{\kappa}_{H^{(e)}_m} = \kappa_{H^{(e)}_m}/(1 - \kappa_{H^{(e)}_m})$. Then, in the following proposition we provide a bound for $\|y^{(e)}_{LS} - y_{LS}\|$.

Proposition 3.2. Under the above notations, if $\|T_{m+1} - I_{m+1}\| \le \epsilon$ ($\epsilon < 1$), then

$$\|\Delta y^{(e)}_{LS}\| \le \bar{\kappa}_{H^{(e)}_m}\left(\frac{\epsilon}{1-\epsilon}\frac{\|b\|}{\|H^{(e)}_m\|} + \frac{2\epsilon}{1-\epsilon}\|y^{(e)}_{LS}\| + \frac{2\epsilon}{1-\epsilon}\frac{\kappa_{H^{(e)}_m}\|r^{(e)}\|}{\|H^{(e)}_m\|}\right) + \frac{2\epsilon}{1-\epsilon}\kappa_{H^{(e)}_m}\|y^{(e)}_{LS}\|. \tag{3.9}$$

Proof. For completing the proof of the inequality in Eq (3.9), we analyze the perturbation of the right side term $\beta^{(e)}e_1$ in the least square problem $H^{(e)}_m y = \beta^{(e)}e_1$. Because of $b = V_{m+1}\beta e_1$, multiplying the two sides by $V^{(e)T}_{m+1}$, we find $\beta^{(e)}e_1 = T_{m+1}\beta e_1$ by using the relation $V_{m+1} = V^{(e)}_{m+1}T_{m+1}$. Thus, we can bound the perturbation of $\beta^{(e)}e_1$ as

$$\|\beta e_1 - \beta^{(e)}e_1\| \le \|T^{-1}_{m+1} - I_{m+1}\|\|\beta^{(e)}e_1\| \le \|T^{-1}_{m+1}\|\|T_{m+1} - I_{m+1}\|\|b\|. \tag{3.10}$$

If $\|T_{m+1} - I_{m+1}\| \leq \epsilon$, then the relation $\|\beta e_1 - \beta^{(e)} e_1\| \leq \frac{\epsilon}{1-\epsilon}\|b\|$ holds by using the result in Eq (3.6).

On the other hand, as shown in Proposition 3.1, we have $\|H_m^{(e)} - H_m\| \leq \frac{2\epsilon}{1-\epsilon}\|H_m^{(e)}\|$. Hence, by Lemma 3, we obtain

$$\begin{aligned}\|\Delta y_{LS}^{(e)}\| &= \bar{\kappa}_{H_m^{(e)}} \left(\frac{\|\Delta\beta^{(e)} e_1\|}{\|H_m^{(e)}\|} + \frac{\|\Delta H_m^{(e)}\|}{\|H_m^{(e)}\|}\|y_{LS}^{(e)}\| + \kappa_{H_m^{(e)}} \frac{\|\Delta H_m^{(e)}\|\|r^{(e)}\|}{\|H_m^{(e)}\|^2} \right) + \kappa_{H_m^{(e)}} \frac{\|\Delta H_m^{(e)}\|}{\|H_m^{(e)}\|}\|y_{LS}^{(e)}\| \\ &\leq \|\bar{\kappa}_{H_m^{(e)}} \left(\frac{\epsilon}{1-\epsilon} \frac{\|b\|}{\|H_m^{(e)}\|} + \frac{2\epsilon}{1-\epsilon}\|y_{LS}^{(e)}\| + \frac{2\epsilon}{1-\epsilon} \frac{\kappa_{H_m^{(e)}}\|r^{(e)}\|}{\|H_m^{(e)}\|} \right) + \frac{2\epsilon}{1-\epsilon}\kappa_{H_m^{(e)}}\|y_{LS}^{(e)}\|.\end{aligned}$$

According to Proposition 3.2, the right-hand side term of the inequality (3.9) can be regarded as a function of the parameter ϵ. Let this function be

$$\phi(\epsilon) = \bar{\kappa}_{H_m^{(e)}} \left(\frac{\epsilon}{1-\epsilon} \frac{\|b\|}{\|H_m^{(e)}\|} + \frac{2\epsilon}{1-\epsilon}\|y_{LS}^{(e)}\| + \frac{2\epsilon}{1-\epsilon} \frac{\kappa_{H_m^{(e)}}\|r^{(e)}\|}{\|H_m^{(e)}\|} \right) + \frac{2\epsilon}{1-\epsilon}\kappa_{H_m^{(e)}}\|y_{LS}^{(e)}\|.$$

It is easy to find that $\phi(\epsilon)$ is a strict monotone increasing function on the interval $[0, 1)$ and $\phi(0) = 0$. In other words, the smaller ϵ is, the smaller the error in computing $y_{LS}^{(e)}$ is.

Let $x_m^{(e)}$ and x_m denote the solution of the exact GMRES and the GMRES with inexact inner products, respectively, and $res^{(e)} = b - Ax_m^{(e)}$ and $res = b - Ax_m$ represent their residual vectors, respectively. Now we can bound the residual gap $\|res^{(e)} - res\|$ as follows.

Proposition 3.3. Assume that GMRES with inexact inner products is used for solving the linear system $Ax = b$. Under the above notations, if $\|T_{m+1} - I_{m+1}\| \leq \epsilon$ $(\epsilon < 1)$, then

$$\|res^{(e)} - res\| \leq \|H_m^{(e)}\| \left[(1+\epsilon)\phi(\epsilon) + \epsilon\|y_{LS}^{(e)}\|\right]. \tag{3.11}$$

Proof. Using the relations $x_m = V_m y_{LS}$, $x_m^{(e)} = V_m^{(e)} y_{LS}^{(e)}$, $V_m = V_m^{(e)} T_m$, and $AV_m^{(e)} = V_{m+1}^{(e)} H_m^{(e)}$, we find

$$\begin{aligned}\|res^{(e)} - res\| &= \|(b - Ax_m^{(e)}) - (b - Ax_m)\| \\ &= \|Ax_m^{(e)} - Ax_m\| \\ &= \|AV_m^{(e)} y_{LS}^{(e)} - AV_m y_{LS}\| \\ &= \|AV_m^{(e)} y_{LS}^{(e)} - AV_m^{(e)} T_m y_{LS}\| \\ &= \|V_{m+1}^{(e)} H_m^{(e)} y_{LS}^{(e)} - V_{m+1}^{(e)} H_m^{(e)} T_m y_{LS}\| \\ &= \|H_m^{(e)} (y_{LS}^{(e)} - T_m y_{LS})\| \\ &\leq \|H_m^{(e)}\| \left(\|T_m\|\|y_{LS}^{(e)} - y_{LS}\| + \|T_m - I_m\|\|y_{LS}^{(e)}\| \right).\end{aligned}$$

According to Proposition 3.2 and the assumptions, the inequality in Eq (3.11) follows.

Furthermore, $(1+\epsilon)\phi(\epsilon)$ is a non-negative monotonically increasing function when $\epsilon \in [0, 1)$, because it can be seen as the product of two nonnegative monotonically increasing function defined on $[0, 1)$. Hence, from Proposition 3.3, we see that the residual vector *res* of the inexact GMRES approaches the residual vector $res^{(e)}$ of the exact GMRES when the upper bound ϵ tends to 0.

3.3. FOM with inexact inner products

For the FOM with all arithmetic operations performed exactly, we need to seek a solution $x_m \in \mathcal{K}_m(A,b)$ such that $b - Ax_m \perp \mathcal{K}_m(A,b)$. The key point is to compute the linear system $\tilde{H}_m^{(e)} y = \beta^{(e)}\tilde{e}_1$, where $\tilde{H}_m^{(e)} \in \mathbb{R}^{m\times m}$ is a nonsingular submatrix formed by the top m rows of the upper-Hessenberg matrix $H_m^{(e)}$, with $\beta^{(e)} = \|b\|$ and $\tilde{e}_1 = [1,0,\cdots,0]^T \in \mathbb{R}^m$. Usually, we obtain the exact solution $y^{(e)} = (\tilde{H}_m^{(e)})^{-1}\beta^{(e)}\tilde{e}_1$, which results in $\tilde{r}^{(e)} = \beta^{(e)}\tilde{e}_1 - \tilde{H}_m^{(e)} y^{(e)}$ being the zero vector.

Similarly to GMRES, when the Arnoldi process with inexact inner products is applied into the FOM iteration, a perturbed linear system $\tilde{H}_m y = \beta\tilde{e}_1$ is generated with $\tilde{H}_m$ being a nonsingular submatrix constituted by the top m rows of the upper-Hessenberg matrix H_m. By Lemma 1, the perturbation of $\tilde{H}_m^{(e)}$ is such that the subsequent relationships stand

$$\|\tilde{H}_m^{(e)} - \tilde{H}_m\| = \sigma_{max}(\tilde{H}_m^{(e)} - \tilde{H}_m) \le \sigma_{max}(H_m^{(e)} - H_m) = \|H_m^{(e)} - H_m\| \le \frac{2\epsilon}{1-\epsilon}\|H_m^{(e)}\|.$$

On the other hand, as analyzed in the proof of Proposition 3.2, we also have the perturbation of the right-hand term $\beta^{(e)}\tilde{e}_1$ equipped with the inequality $\|\beta\tilde{e}_1 - \beta^{(e)}\tilde{e}_1\| \le \frac{\epsilon}{1-\epsilon}\|b\|$.

Let $\tilde{x}_m^{(e)}$ and $\tilde{x}_m$ denote the solution of the exact FOM and the FOM with inexact inner products, respectively, and $\widetilde{res}^{(e)} = b - A\tilde{x}_m^{(e)}$ and $\widetilde{res} = b - A\tilde{x}_m$ represent their residual vectors, respectively. Then, as discussed above, we can bound the residual gap $\|\widetilde{res}^{(e)} - \widetilde{res}\|$ as performed in the following proposition.

Proposition 3.4. Assume that FOM with inexact inner products is used for solving the linear system $Ax = b$. Under the above notations, if $\|T_{m+1} - I_{m+1}\| \le \epsilon$ $(\epsilon < 1)$, then

$$\|\widetilde{res}^{(e)} - \widetilde{res}\| \le \|H_m^{(e)}\|\left[(1+\epsilon)\tilde{\phi}(\epsilon) + \epsilon\|y^{(e)}\|\right], \tag{3.12}$$

where $\tilde{\phi}(\epsilon) = \bar{\kappa}_{H_m^{(e)}}\left(\frac{\epsilon}{1-\epsilon}\frac{\|b\|}{\|H_m^{(e)}\|} + \frac{2\epsilon}{1-\epsilon}\|y^{(e)}\|\right) + \frac{2\epsilon}{1-\epsilon}\kappa_{H_m^{(e)}}\|y^{(e)}\|$.

Proof. The proof is similar to that of Proposition 3.3.

Similarly, it is easy to find that $(1+\epsilon)\tilde{\phi}(\epsilon)$ is a monotonically increasing function on the interval [0,1). According to Proposition 3.4, when the parameter ϵ tends to 0, the residual vector $\widetilde{res}$ of the inexact FOM approaches the residual vector $\widetilde{res}^{(e)}$ of the exact FOM.

3.4. Arnoldi-type method with inexact inner products

The Arnoldi-type method is an efficient approximate method for computing the eigenvectors of a matrix with the Arnoldi process as main core. For the eigenvector problem $Ax = \lambda x$ $(x \ne 0)$, it ultimately boils down to seeking a solution $x_m \in \mathcal{K}_m(A,v)$ that satisfies the optimality property

$$\|(A-\lambda I)x_m\| = \min_{u\in\mathcal{K}_m(A,v),\|u\|=1}\|(A-\lambda I)u\|. \tag{3.13}$$

Algorithm 3 shows the exact Arnoldi-type method for solving the eigenvector problem $(A-\lambda I)x = 0$. In Step 1 of Algorithm 3, we obtain the basis matrix $V_m^{(e)}$ and the upper-Hessenberg matrix $H_m^{(e)}$ by the standard Arnoldi process as given in Algorithm 1. In Step 2 of Algorithm 3, a singular value decomposition of the matrix $H_m^{(e)} - [\lambda I_m; O]$ is performed. In Step 3 of Algorithm 3, we get the right singular vector $s_m^{(e)}$ corresponding to the minimum singular value $\sigma_{min}(H_m^{(e)} - [\lambda I_m; O])$. Then, in Step 4 of Algorithm 3, an approximate vector $x_m^{(e)} = V_m^{(e)} s_m^{(e)}$ is obtained. Let $\sigma_m^{(e)} = \sigma_{min}(H_m^{(e)} - [\lambda I_m; O])$, and

$res_{AT}^{(e)} = (A - \lambda I)x_m^{(e)}$ be the residual vector of the exact Arnoldi-type method. As discussed in [28, 29], the value of $\sigma_m^{(e)}$ can be used as a stopping criterion, i.e., $\sigma_m^{(e)} = \|res_{AT}^{(e)}\|$.

Algorithm 3 Arnoldi-type method for solving $(A - \lambda I)x = 0$

Require: $A \in \mathbb{R}^{n\times n}$, initial nonzero vector $v \in \mathbb{R}^n$, $m \in \mathbb{N}$
1: Obtain $V_m^{(e)}$ and $H_m^{(e)}$ by Algorithm 1.
2: Compute $H_m^{(e)} - [\lambda I_m; O] = U^{(e)}\Sigma^{(e)}S^{(e)T}$.
3: Set $s_m^{(e)}$ is the right singular vector corresponding to $\sigma_{min}(H_m^{(e)} - [\lambda I_m; O])$.
4: Obtain $x_m^{(e)} = V_m^{(e)}s_m^{(e)}$.

When the Arnoldi process with inexact inner products is applied into the Arnoldi-type method, we obtain the basis matrix V_m and the upper-Hessenberg matrix H_m. Thus a perturbation of the matrix $H_m^{(e)} - [\lambda I_m; O]$ is generated, which is denoted as $H_m - [\lambda I_m; O]$. Similarly, a singular value decomposition for the matrix $H_m - [\lambda I_m; O]$ is computed, and the right singular vector s_m is obtained corresponding to the minimum singular value $\sigma_{min}(H_m - [\lambda I_m; O])$. Finally, an approximate vector $x_m = V_m s_m$ is computed. Let $\sigma_m = \sigma_{min}(H_m - [\lambda I_m; O])$, and let $res_{AT} = (A - \lambda I)x_m$ be the residual vector of the inexact Arnoldi-type method. Now we give an upper bound of $\|res_{AT}\|$ as follows.

Proposition 3.5. Assume that the Arnoldi-type method with inexact inner products is used for solving the eigenvector problem $(A - \lambda I)x = 0$. Under the above notations, if $\|T_{m+1} - I_{m+1}\| \le \epsilon$ $(\epsilon < 1)$, then

$$\|res_{AT}\| \le (1 + \epsilon)\left(\frac{2\epsilon}{1-\epsilon}\|H_m^{(e)}\| + \|res_{AT}^{(e)}\|\right). \tag{3.14}$$

Proof. Using the relations $AV_m = V_{m+1}H_m$ and $V_{m+1} = V_{m+1}^{(e)}T_{m+1}$, we find

$$\begin{aligned}
\|res_{AT}\| &= \|(A - \lambda I)V_m s_m\| \\
&= \|V_{m+1}(H_m - [\lambda I_m; O])s_m\| \\
&= \|T_{m+1}(H_m - [\lambda I_m; O])s_m\| \\
&\le (\|T_{m+1} - I_{m+1}\| + 1)\,\|(H_m - [\lambda I_m; O])s_m\| \\
&= (\|T_{m+1} - I_{m+1}\| + 1)\,\sigma_m.
\end{aligned} \tag{3.15}$$

By Lemma 2 and Proposition 3.1, we infer

$$|\sigma_m - \sigma_m^{(e)}| \le \|(H_m - [\lambda I_m; O]) - (H_m^{(e)} - [\lambda I_m; O])\| = \|H_m - H_m^{(e)}\| \le \frac{2\epsilon}{1-\epsilon}\|H_m^{(e)}\|, \tag{3.16}$$

and as a consequence the following inequality holds

$$\sigma_m \le \sigma_m^{(e)} + \frac{2\epsilon}{1-\epsilon}\|H_m^{(e)}\|. \tag{3.17}$$

Substituting Eq (3.17) into Eq (3.15), the proof of the relation in Eq (3.14) is complete.

From the right-hand term of the inequality (3.14), it is easy to see that if ϵ is close to zero, then the quantity $\|res_{AT}\|$ is sufficiently close to $\|res_{AT}^{(e)}\|$.

4. Numerical experiments

In this section, we illustrate our theoretical results with a few numerical examples. All the numerical experiments are run in MATLAB R2018b on a 64-bit Windows 10 computer equipped with a core i7-8550u processor and with 16GB of RAM memory.

To compare the Arnoldi process-based Krylov subspace methods with their inexact inner products counterparts, we introduce Algorithm 4, Arnoldi process with inexact inner products. This algorithm replaces the inner products operation in Algorithm 1 with Eq (2.2). The standard Arnoldi process as given in Algorithm 1 and the Arnoldi process with inexact inner products as shown in Algorithm 4 is applied to GMRES, FOM, and Arnoldi-type methods, respectively.

Algorithm 4 Arnoldi process with inexact inner products

Require: $A \in \mathbb{R}^{n\times n}, v = b \in \mathbb{R}^n, m \in \mathbb{N}$

1: $\beta^{(e)} = \|v\|$
2: $v_1^{(e)} = v/\beta^{(e)}$
3: **for** $j = 1$ to m **do**
4: $\quad w = Av_j^{(e)}$
5: $\quad$ **for** $i = 1$ to j **do**
6: $\quad\quad h_{ij}^{(e)} = v_i^{(e)T} w + \eta_{ij}$
7: $\quad\quad w = w - h_{ij}^{(e)} v_i^{(e)}$
8: $\quad$ **end for**
9: $\quad h_{j+1,j}^{(e)} = \sqrt{w^T w + \eta_{j+1,j}}$
10: $\quad$ **if** $h_{j+1,j}^{(e)} = 0$ **then**
11: $\quad\quad$ **break**
12: $\quad$ **end if**
13: $\quad v_{j+1}^{(e)} = w/h_{j+1,j}^{(e)}$
14: **end for**

4.1. How inexact inner products influence the upper-Hessenberg matrix

Our first numerical example aims at illustrating Proposition 3.1, which provides bounds for the perturbation of the upper-Hessenberg matrix $H_m^{(e)}$. Here, we consider the Grcar matrix of parameter 5 and size 100, i.e., the Toeplitz matrix $A \in \mathbb{R}^{100\times 100}$ reported below

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & \cdots & 0 \\ -1 & 1 & 1 & 1 & 1 & 1 & 1 & \ddots & 0 \\ 0 & -1 & 1 & 1 & 1 & 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & -1 & 1 \end{pmatrix}_{100\times 100} . \tag{4.1}$$

More precisely, $A = T_n(f_5)$, $n = 100$, $f_5 = \sum_{j=0}^{5} \exp(ij\theta) - \exp(-i\theta)$, $i^2 = -1$, $\theta \in [-\pi, \pi]$, and according to the theory, the matrix-sequence $\{T_n(f_5)\}_n$ is canonically distributed as $|f_5|$ in the singular value sense, while the eigenvalue canonical distribution does not hold (see [30, 31] and references therein): hence the singular values of $T_n(f_5)$ behave as a uniform sampling of $|f_5(\theta)|$ of cardinality 100, while the moduli of the eigenvalues are much smaller than the singular values and the latter discrepancy represents a measure of its high nonnormality [32].

The right-hand side is set as $b = A[\sin(1), \cdots, \sin(100)]^T$. Let $\epsilon = \|I_{m+1} - T_{m+1}\|$ represent a measure of inexactness of the inexact inner products. Assume errors η_{ij} and $\eta_{j+1,j}$ in Algorithm 4 are uniformly distributed between $-\eta$ and η. The value of η that changes with k is set as $10^{-6} \times 2^{1-k}, k = 1, \cdots, 30$. Taking the η as X-axis, we conduct numerical experiments with Krylov subspace dimension m taking values of 10, 40, and 70, and the numerical results are plotted in Figure 1.

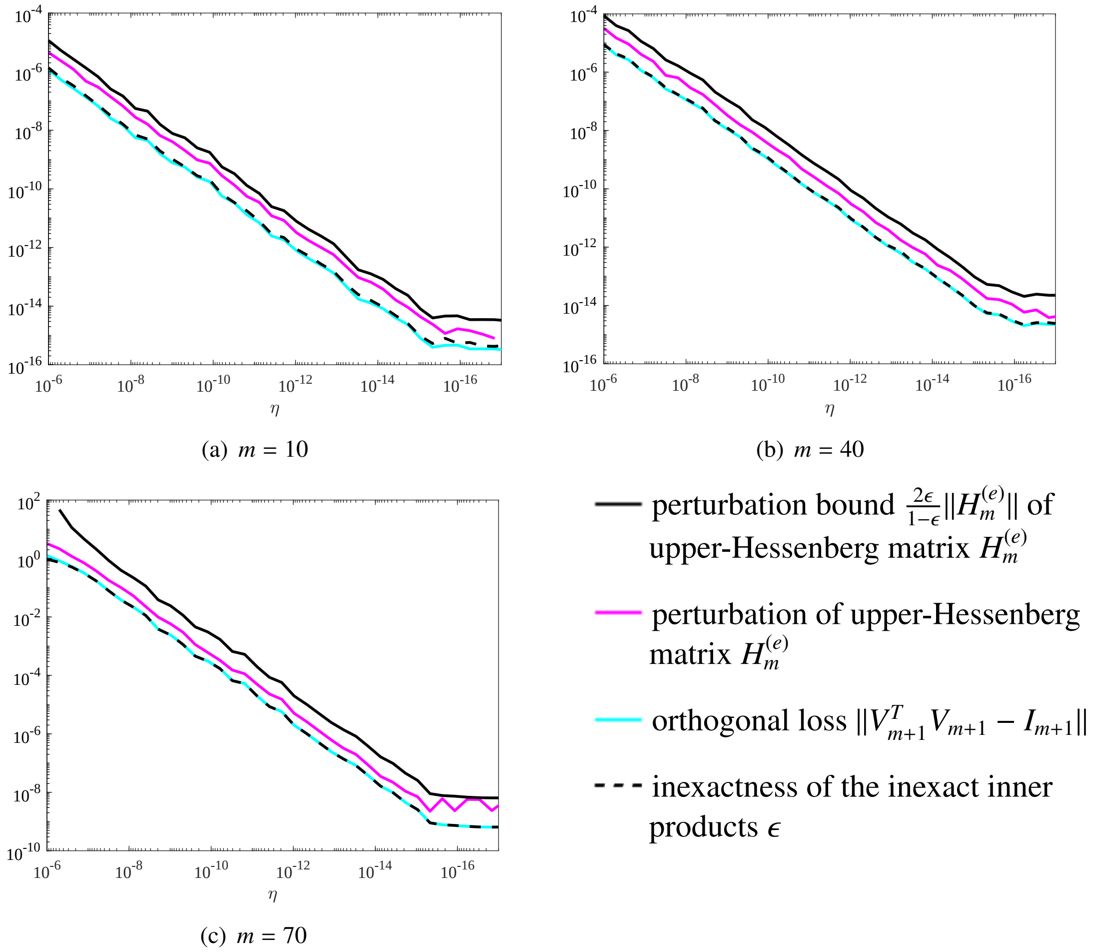


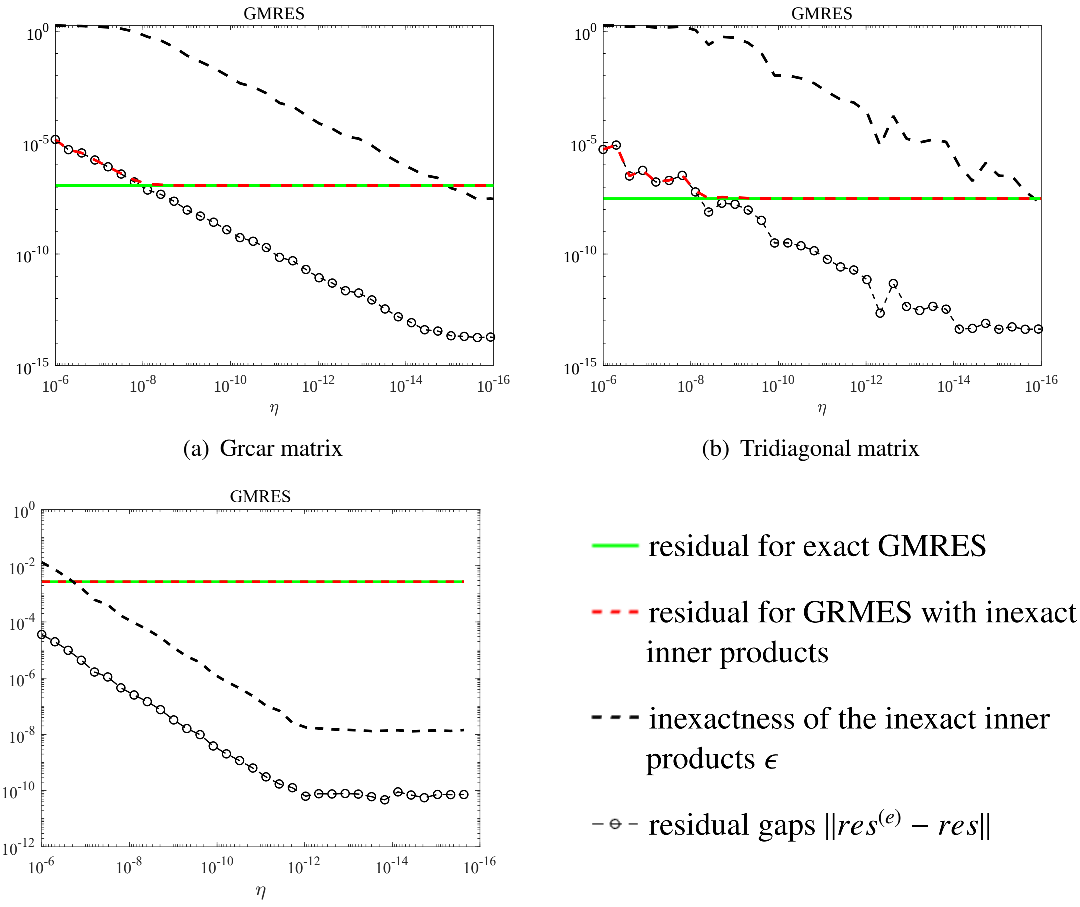
Figure 1. Numerical behaviors for the Grcar matrix by the Arnoldi process with inexact inner products.

Numerical results are shown in Figure 1. From Figure 1, we can see that $\frac{2\epsilon}{1-\epsilon}\|H_m^{(e)}\|$ is a good upper bound for the perturbation of the upper-Hessenberg matrix $H_m^{(e)}$ caused by the inexact inner products. In addition, we find that the curve of orthogonal loss described by $\|V_{m+1}^T V_{m+1} - I_{m+1}\|$ almost coincides with the curve of inexactness of the inexact inner products given by $\epsilon = \|I_{m+1} - T_{m+1}\|$ to a certain degree extent, which testifies the sharpness of our analysis as shown in Eq (2.12). Particularly, the curve of inexactness of the inexact inner products is lower than the curve of orthogonal loss slightly. This numerical behaviour indicates that inexactness of the inexact inner products described by $\epsilon = \|I_{m+1} - T_{m+1}\|$ may be a better measure to quantify the orthogonal loss.

4.2. How inexact inner products influence GMRES and FOM

Here we illustrate the theoretical analysis about GMRES and FOM with inexact inner products. For the numerical examples, we not only continue to use the matrix given in the Section 4.1, but also we introduce two new examples to further illustrate our theoretical results. The first example is the following tridiagonal matrix:

$$A = \begin{pmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & -1 \\ & & & & -1 & 2 \end{pmatrix}_{1000\times 1000} . \tag{4.2}$$



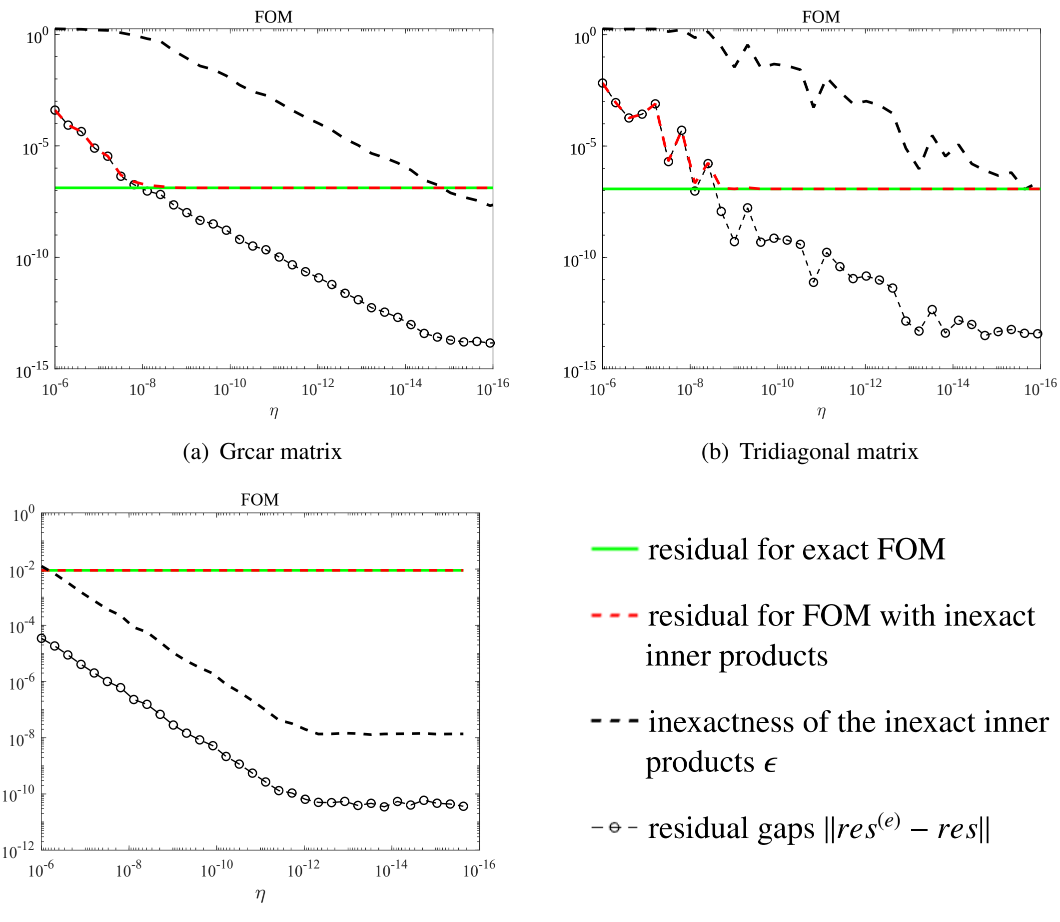
(a) Grcar matrix

(b) Tridiagonal matrix

(c) 494_bus

Figure 2. GMRES with inexact inner products for the Grcar matrix, the tridiagonal matrix and the 494_bus matrix.

The second example is the 494_bus matrix, $A \in \mathbb{R}^{494\times494}$, which can be available from the SuiteSparse matrix collection [33]. The condition number of the 494_bus matrix is $\kappa_2(A) \approx 10^6$. For the tridiagonal matrix, we set $b = A[\sin(1), \cdots, \sin(1000)]^T$, and for 494_bus matrix, we set $b = A[\sin(1), \cdots, \sin(494)]^T$. As shown in the Section 4.1, no matter what value Krylov subspace dimension m is, the decreasing trend of the perturbation norm $\|H_m^{(e)} - H_m\|$ as η decreases does not change. Hence, we can set the value of Krylov subspace dimesion m according to the specific problem, although other values are also possible. For the Grcar matrix, we set $m = 80$, for the tridiagonal matrix, $m = 400$, and for the 494_bus matrix, $m = 200$. The setting of η is the same as it is in Section 4.1.



(a) Grcar matrix

(b) Tridiagonal matrix

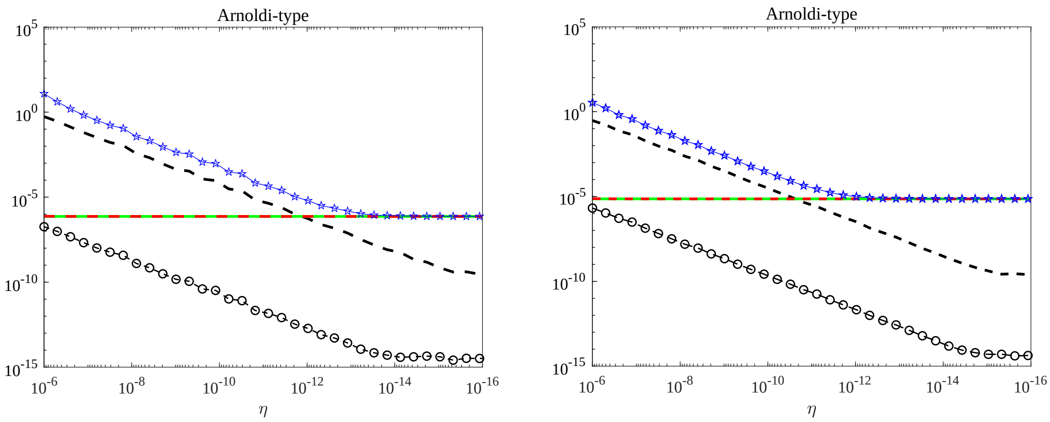
(c) 494_bus

Figure 3. FOM with inexact inner products for for the Grcar matrix, the tridiagonal matrix and the 494_bus matrix.

From Figures 2 and 3, as η decreases, we observe a concurrent decrease in the residual gap. This trend implies that we can manage errors in inner products computations within a specific range without substantially affecting the convergence in GRMES and FOM. Then, during solving the tridiagonal matrix problem by GMRES and FOM, we note that the change of residual gaps are almost stagnant when $\eta \leq 10^{-14}$. For the considered problems, this observation suggests that enhancing the precision of inner products computations largely may not lead to significant improvements in the convergence rate.

4.3. *How inexact inner products influence Arnoldi-type method*

In this section, we continue to use the Grcar matrix and the tridiagonal matrix as given previously to illustrate our theoretical analysis for the Arnoldi-type method. Regarding parameters, the same setting of η is used as previously; in this part, we set $m = 75$ for the Grcar matrix and $m = 400$ for the tridiagonal matrix, respectively. In addition, because the classical PageRank problem [34] can be regarded as a typical problem for computing the principal eigenvector with its principal eigenvalue is 1, we consider using the exact and inexact inner products versions of the Arnoldi-type method to calculate the PageRank vector. Taking the Web-stanford matrix as an example, which contains 281,903 pages and 2,312,497 links and can be available from http://www.ciseuf.edu/research/sparse/matrices/groups.html. For the Web-stanford matrix, we assume the damping factor $\alpha = 0.99$, the Krylov subspace dimension $m = 100$, and other settings are the same as the Grcar matrix and the tridiagonal matrix. Numerical results are shown in Figures 4 and 5. Regarding large damping factors as $\alpha = 0.99$, the standard power method becomes too slow and acceleration techniques are necessary, such as either (preconditioned) Krylov methods [35] or extrapolation techniques [36]. For a general structural and spectral analysis of Web matrices as a function of the damping factor α (see [37]).



(a) Grcar matrix (b) Tridiagonal matrix

residual for exact Arnoldi-type

residual for Arnoldi-type with inexact inner products

inexactness of the inexact inner products ϵ

residual gaps $\|res_{AT}^{(e)} - res_{AT}\|$

upper bound of $\|res_{AT}^{(e)}\|$

Figure 4. Arnoldi-type method with inexact inner products for the Grcar matrix and the tridiagonal matrix.

From Figures 4 and 5, we still can see that the residual gaps $\|res_{AT}^{(e)} - res_{AT}\|$ decrease with decline of inexactness of inexact inner products ϵ, which illustrates that the influence on Arnoldi-type method is similar with GMRES and FOM in Figures 2 and 3. In addition, when inexactness of inner products ϵ declines to some degree, the upper bound of $\|res_{AT}^{(e)}\|$ is very close to $\|res_{AT}^{(e)}\|$. More importantly, residual for exact Arnodli-type almost coincides with residual for Arnoldi-type with inexact inner products, which means that we may be able to lower the precision of inner products computations appropriately to improve computational efficiency of the Arnoldi-type method for solving large scale problems.

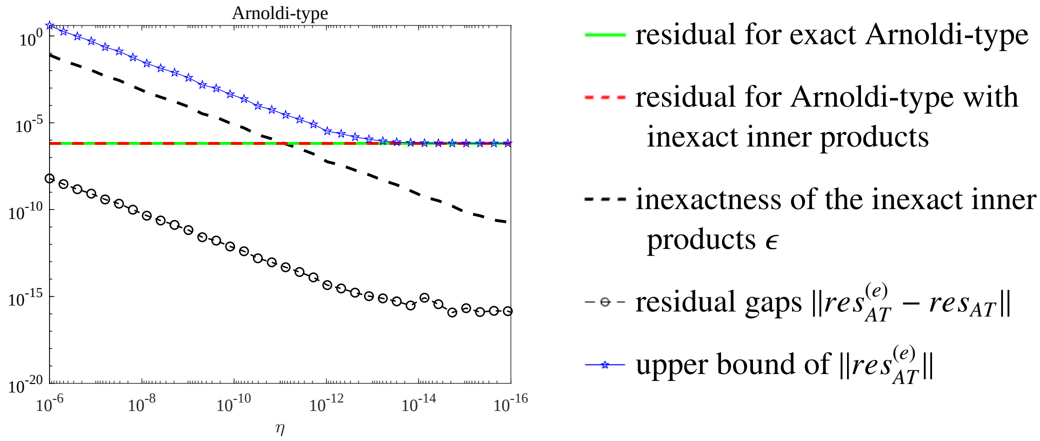


Figure 5. Arnoldi-type method with inexact inner products for the Web-stanford matrix.

5. Conclusions

In the present paper, we have explored a new analysis framework to discuss how inexact inner products influence the convergence and final accuracy of several Arnoldi-based Krylov subspace methods. By introducing a transition matrix and investigating its properties, we have defined a quantity in order to measure the inexactness of inexact inner products. Then, we analyzed the perturbation of the upper-Hessenberg matrix $H_m^{(e)}$ as shown in Proposition 3.1. Based on this analysis, we have developed the theoretical results for the GMRES, FOM, and Arnoldi-type methods as discussed in Propositions 3.3–3.5. Numerical results in Section 4 illustrate the reliability and sharpness of our theoretical results. In particular, we deduced that the residual gaps are small especially for the Arnoldi process based Krylov subspace methods with inexact inner products. A consequence stemming from this fact is that we could consider diminishing the precision of inner products arithmetic to some extent to improve the computational efficiency for Arnoldi process based Krylov subspace methods. This will be considered in a future research.

Author contributions

Meng Su proposed the theory; Chun Wen and Zhao-Li Shen designed the numerical experiments. Stefano Serra-Capizzano offered valuable feedback on the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors are grateful to the Editor and anonymous referees for their valuable comments and suggestions on the original manuscript, which greatly improved the quality of this article. This research is supported by the National Natural Science Foundation of China (12101433) and the Sichuan Science and Technology Program (2023NSFSC1136).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, H. van der Vorst, *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, Society for Industrial and Applied Mathematics, Philadelphia, USA, 2000.
2. A. Greenbaum, *Iterative Methods for Solving Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, 1997.
3. Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Halstead Press, New York, 1992.
4. Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS Publishing, Boston, MA, USA, 1996.
5. A. Bouras, V. Frayssé, A relaxation strategy for inexact matrix-vector products for Krylov methods, *Technical Report 15*, CERFACS, Toulouse, France, 2000.
6. A. Bouras, V. Frayssé, Inexact matrix-vector products in Krylov methods for solving linear systems: A relaxation strategy, *SIAM J. Matrix Anal. Appl.*, **26** (2005), 660–678. https://doi.org/10.1137/s0895479801384743
7. J. van den Eshof, G. L. G. Sleijpen, Inexact Krylov subspace methods for linear systems, *SIAM J. Matrix Anal. Appl.*, **26** (2004), 125–153. https://doi.org/10.1137/S0895479802403459
8. V. Simoncini, D. B. Szyld, Theory of inexact Krylov subspace methods and applications to scientific computing, *SIAM J. Sci. Comput.*, **25** (2003), 454–477. https://doi.org/10.1137/s1064827502406415
9. G. L. G. Sleijpen, J. van den Eshof, M. B. van Gijzen, Restarted GMRES with inexact matrix–vector products, in *Numerical Analysis and Its Applications: Third International Conference, NAA 2004*, Springer-Verlag Berlin Heidelberg, (2005), 494–502. https://doi.org/10.1007/978-3-540-31852-1_60
10. S. Gratton, E. Simon, D. Titley-Peloquin, P. L. Toint, A note on inexact inner products in GMRES, *SIAM J. Matrix Anal. Appl.*, **43** (2022), 1406–1422. https://doi.org/10.1137/20m1320018
11. L. Giraud, S. Gratton, J. Langou, Convergence in backward error of relaxed GMRES, *SIAM J. Sci. Comput.*, **29** (2007), 710–728. https://doi.org/10.1137/040608416
12. V. Simoncini, D. B. Szyld, Recent computational developments in Krylov subspace methods for linear systems, *Numer. Linear Algebra Appl.*, **14** (2007), 1–59. https://doi.org/10.1002/nla.499
13. A. Bouras, V. Frayssé, L. Giraud, A relaxation strategy for inner-outer linear solvers in domain decomposition methods, *Technical Report 17*, CERFACS, Toulouse, France, 2000.
14. P. Kürschner, M. A. Freitag, Inexact methods for the low rank solution to large scale Lyapunov equations, *BIT Numer. Math.*, **60** (2020), 1221–1259. https://doi.org/10.1007/s10543-020-00813-4

15. S. J. Leon, A. Björck, W. Gander, Gram-Schmidt orthogonalization: 100 years and more, *Numer. Linear Algebra Appl.*, **20** (2013), 175–188. https://doi.org/10.1002/nla.1839

16. A. Björck, Solving linear least squares problems by Gram-Schmidt orthogonalization, *BIT Numer. Math.*, **7** (1967), 1–21. https://doi.org/10.1007/bf01934122

17. A. Björck, C. C. Paige, Loss and recapture of orthogonality in the modified Gram–Schmidt algorithm, *SIAM J. Matrix Anal. Appl.*, **13** (1992), 176–190. https://doi.org/10.1137/0613015

18. J. Drkošová, A. Greenbaum, M. Rozložník, Z. Strakoš, Numerical stability of GMRES, *BIT Numer. Math.*, **35** (1995), 309–330. https://doi.org/10.1007/bf01732607

19. A. Greenbaum, M. Rozložník, Z. Strakoš, Numerical behaviour of the modified Gram–Schmidt GMRES implementation, *BIT Numer. Math*, **37** (1997), 706–719. https://doi.org/10.1007/bf02510248

20. C. C. Paige, M. Rozložník, Z. Strakoš, Modified Gram–Schmidt (MGS), least squares, and backward stability of MGS-GMRES, *SIAM J. Matrix Anal. Appl.*, **28** (2006), 264–284. https://doi.org/10.1137/050630416

21. R. A. Horn, C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 2013.

22. G. H. Golub, C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, USA, 2013.

23. A. Stathopoulos, K. Wu, A block orthogonalization procedure with constant synchronization requirements, *SIAM J. Sci. Comput.*, **23** (2002), 2165–2182. https://doi.org/10.1137/s1064827500370883

24. R. A. Horn, C. R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1994.

25. G. W. Stewart, *Introduction to Matrix Computations*, Academic Press, New York-London, 1973.

26. Y. Saad, M. H. Schultz, GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM J. Sci. Stat. Comput.*, **7** (1986), 856–869. https://doi.org/10.1137/0907058

27. C. L. Lawson, R. J. Hanson, *Solving Least Squares Problems*, Society for Industrial and Applied Mathematics, Philadelphia, USA, 1995.

28. G. H. Golub, C. Greif, An Arnoldi-type algorithm for computing page rank, *BIT Numer. Math.*, **46** (2006), 759–771. https://doi.org/10.1007/s10543-006-0091-y

29. Z. Jia, Refined iterative algorithms based on Arnoldi's process for large unsymmetric eigenproblems, *Linear Algebra Appl.*, **259** (1997), 1–23. https://doi.org/10.1016/S0024-3795(96)00238-8

30. G. Barbarino, C. Garoni, S. Serra-Capizzano, Block generalized locally Toeplitz sequences: Theory and applications in the multidimensional case, *Electron. Trans. Numer. Anal.*, **53** (2020), 113–216. https://doi.org/10.1553/etna_vol53s113

31. S. Serra-Capizzano, P. Tilli, Extreme singular values and eigenvalues of non-Hermitian block Toeplitz matrices, *J. Comput. Appl. Math.*, **108** (1999), 113–130. https://doi.org/10.1016/s0377-0427(99)00104-1

32. R. Bhatia, Matrix Analysis, *Graduate Texts in Mathematics*, Springer-Verlag, New York, USA, 1997.

33. T. A. Davis, Y. Hu, The University of Florida sparse matrix collection, *ACM Trans. Math. Software*, **38** (2011), 1–25. https://doi.org/10.1145/2049662.20496

34. L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank citation ranking: Bringing order to the web, Technical Report, Stanford University, USA, 1998.

35. G. Del Corso, A. Gullí, F. Romani, Comparison of Krylov subspace methods on the PageRank problem, *J. Comput. Appl. Math.*, **210-1/2** (2007), 159–166. https://doi.org/10.1016/j.cam.2006.10.080

36. S. Serra-Capizzano, Jordan canonical form of the Google matrix: A potential contribution to the PageRank computation, *SIAM J. Matrix Anal. Appl.*, **27** (2005), 305–312. https://doi.org/10.1137/s0895479804441407

37. R. A. Horn, S. Serra-Capizzano, A general setting for the parametric Google matrix, *Internet Math.*, **3** (2006), 385–411. https://doi.org/10.1080/15427951.2006.10129131