EasyChair Preprint

№ 7734

Post-Modern GMRES

Stephen Thomas, Erin Carson, Miro Rozloznik, Arielle Carr and Kasia Swirydowicz

April 8, 2022

POST-MODERN GMRES

STEPHEN THOMAS*, ERIN CARSON†, MIRO ROZLOŽNíK‡, ARIELLE CARR§, AND KASIA SWIRYDOWICZ¶

Abstract. The GMRES algorithm of Saad and Schultz (1986) for nonsymmetric linear systems relies on the Arnoldi expansion for the Krylov basis. The algorithm computes the QR factorization of the matrix $B = [\, \mathbf{r}_0, AV_m \,]$. Despite an $\mathcal{O}(\varepsilon)\kappa(B)$ loss of orthogonality, the modified Gram-Schmidt (MGS) formulation was shown to be backward stable in the seminal papers by Paige, et al. (2006) and Paige and Strakoš (2002). Classical Gram-Schmidt (CGS) exhibits an $\mathcal{O}(\varepsilon)\kappa^2(B)$ loss of orthogonality, whereas DCGS-2 (CGS with delayed reorthogonalization) reduces this to $\mathcal{O}(\varepsilon)$ in practice (without a formal proof). We present a post-modern (viz not classical) GMRES algorithm based on Ruhe (1983) and the low-synch algorithms of Swirydowicz et al (2020) that achieves $\mathcal{O}(\varepsilon)\, \|A\mathbf{v_k}\|_2/h_{k+1,k}$ loss of orthogonality. By projecting the vector Av_m with Gauss-Seidel onto the orthogonal complement of the space spanned by the computed Krylov vectors V_m where $\tilde{V}_m^T\tilde{V}_m = I + L_m + L_m^T$, we can further demonstrate that the loss of orthogonality closely follows $\mathcal{O}(\varepsilon)$. For a broad class of matrices, unlike MGS-GMRES, significant loss of orthogonality does not occur and the relative residual no longer stagnates for highly non-normal systems. The Krylov vectors remain linearly independent and the smallest singular value of $\tilde{V}_m$ is close to one. We also demonstrate that Henrici's departure from normality of the lower triangular matrix $T_m \approx (\, \tilde{V}_m^T\tilde{V}_m \,)^{-1}$ in the Gram-Schmidt projector $P = I - V_m T_m V_m^T$ is an appropriate quantity for detecting the loss of orthogonality.

1. Introduction. The purpose of the present work is to derive a post-modern (viz. not classical) formulation of the GMRES algorithm that uses an orthgonalization scheme based on the iterated solution of the normal equations in the Gram-Schmidt projector, as described by Ruhe [1], and the low-synch algorithms introduced by Swirydowicz et al. [2]. The essential idea developed here is to project the vector $A\tilde{\mathbf{v}}_k$ onto the orthogonal complement of the space spanned by the computed Krylov vectors represented by the columns of $\tilde{V}_m \in \mathbb{C}^{n\times m}$, where $\tilde{V}_m^T\tilde{V}_m = I + L_m + L_m^T$ and $L_m \in \mathbb{C}^{m\times m}$ is strictly upper triangular. Ruhe [1] suggested applying LSQR, whereas Björck [3, pg. 312] recommended conjugate gradients. Instead, we apply two Gauss-Seidel iterations and note that Higham and Knight [4] proved norm-wise backward stability for such stationary iterations. We demonstrate that the loss of orthogonality may then be bounded by $\mathcal{O}(\varepsilon)$ *without any need for reorthogonalization* in the Arnoldi-QR algorithm. Unlike the relative residual for MGS-GMRES, the stagnation shown in Paige and Strakoš [5] does not occur and yet the seminal backward stability result of Paige et al. [6] still applies.

In the present study, linear systems of the form $A\mathbf{x} = \mathbf{b}$ with A an $n \times n$ real-valued matrix, are solved with Krylov subspace methods. Here, let $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$ denote the initial residual with initial guess $\mathbf{x}_0$. Inside GMRES, the Arnoldi QR algorithm is applied to generate an orthonormal basis for the Krylov subspace $\mathcal{K}_m(\, A,\, \mathbf{r}_0\,)$ spanned by the columns of the $n \times m$ matrix, V_m, where $m \ll n$, and produces the $(m+1) \times m$ Hessenberg matrix, $H_{m+1,m}$, in the Arnoldi expansion such that

$$AV_m = V_{m+1}H_{m+1,m}.$$

The Arnoldi algorithm produces a QR factorization of $B = [\, \mathbf{r}_0,\, AV_m \,]$ and the columns of V_{m+1} form an orthogonal basis for the Krylov subspace $\mathcal{K}_m$ [6]. When the Krylov vectors are orthogonalized via the finite precision MGS algorithm, their loss of orthogonality is related in a straightforward way to the convergence of GMRES. Orthogonality among the Krylov vectors is effectively maintained until the norm-wise relative backward error approaches the machine precision as discussed in Paige and Strakoš [5] and Paige et al. [6]. The growth of the condition number of B is related to the norm-wise relative backward error

$$\beta(\mathbf{x}^{(k)}) = \frac{\|\mathbf{r}^{(k)}\|_2}{\|\mathbf{b}\|_2 + \|A\|_\infty\|\mathbf{x}^{(k)}\|_2}$$

and in particular, it is observed in exact arithmetic that $\beta(\mathbf{x}^{(k)})\, \kappa([\, \mathbf{r}_0,\, A\mathbf{v}_k \,]) = \mathcal{O}(1)$.

The orthogonality of the columns determines the numerical rank of the Krylov basis. However, in finite-precision arithmetic, V_m may "lose" orthogonality and this loss, as measured by $\|I - V_m^TV_m\|_F$,

*National Renewable Energy Laboratory, Golden, CO
†Charles University, Prague, CZ
‡Czech Academy of Sciences, Institute of Mathematics, Prague, CZ
§Lehigh University, Bethlehem, PA
¶Pacific Northwest Laboratory, WA

may deviate substantially from machine precision, $\mathcal{O}(\varepsilon)$. When linear independence is completely lost, the relative residual may stagnate at a certain level above $\mathcal{O}(\varepsilon)$ and this occurs when $\|S_m\|_2 = 1$, where $S_m = (I + L_m^T)^{-1} L_m^T$ and L_m is the $m \times m$ strictly lower triangular part of $\tilde{V}_m^T \tilde{V}_m$.

The development of low-synchronization Gram-Schmidt and generalized minimal residual algorithms by Świrydowicz et al. [2] and Bielich et al. [7] was largely driven by applications that need stable, yet scalable solvers. Both the modified (MGS) and classical Gram-Schmidt algorithms with delayed re-orthogonalization (DCGS-2) are stable for a GMRES solver. Although the DCGS-2 results in an $\mathcal{O}(\varepsilon)$ loss of orthogonality, which suffices for GMRES to converge, stability has not been proven formally. Paige et al. [6] demonstrate that despite $\mathcal{O}(\varepsilon)\kappa(B)$ loss of orthogonality, MGS-GMRES is backward stable for the solution of linear systems. Here, the condition number of the matrix B is given by $\kappa(B) = \sigma_{\max}(B)/\sigma_{\min}(B)$, where $\sigma_{\max}(B)$ and $\sigma_{\min}(B)$ are the maximum and minimum singular values of the matrix B, respectively.

An inverse compact WY modified Gram-Schmidt algorithm is presented in [2] and is based upon the application of the projector

$$P = I - V_m \, T_m \, V_m^T, \quad T_m \approx (\, \tilde{V}_m^T \tilde{V}_m \,)^{-1}$$

where $\tilde{V}_m$ is again $n \times m$, I is the identity matrix of dimension n, and T_m is an $m \times m$ lower triangular matrix. To obtain a low-synch MGS algorithm, or one MPI global reduction per GMRES iteration, the normalization is delayed to the next iteration. The matrix T_m is obtained from the strictly lower triangular part of $\tilde{V}_m^T \tilde{V}_m$, denoted L_m. Note that because V_m has almost orthonormal columns, the norm of L_m is small, and T_m is close to I (here, the identity matrix of dimension m).

A Neumann series expansion for the inverse of the lower triangular matrix, T_m, results from the compact WY form of the projector P, Thomas et al. [8]. A post-modern GMRES (PM-GMRES) formulation based upon the matrix polynomial $T_m^{(2)}$ associated with two Gauss-Seidel iterations for the normal equations

$$V_{m-1}^T V_{m-1} \mathbf{r}_{1:m-1,m} = V_{m-1}^T \, A \, \mathbf{v}_m \tag{1.1}$$

can be derived, with the projector

$$P^{(2)} = I - V_m \, T_m^{(2)} \, V_m^T, \quad T_m^{(2)} = I - L_m - L_m^T + L_m^2 + L_m^T L_m + L_m L_m^T - \cdots, \tag{1.2}$$

where the rows of L_m are constructed from the matrix-vector products $V_{m-1}^T \mathbf{v}_{m-1}$. The sum is finite because the matrix L_m is nilpotent, as originally noted by Ruhe [1]. The loss of orthogonality then follows $\mathcal{O}(\varepsilon) \, \|A\tilde{\mathbf{v}}_\mathbf{m}\|_2 / h_{m+1,m}$. For extremely ill-conditioned and non-normal matrices, the convergence history of the PM-GMRES algorithm has been found to be identical to the original MGS-GMRES algorithm introduced by Saad and Schultz [9], with the exception that the (implicit) relative residual continues to decrease monotonically and never stagnates until reaching the level $\mathcal{O}(\varepsilon)$ and thus matches the Householder HH-GMRES of Walker [10].

Contributions. In this paper, we present a new formulation of the MGS-GMRES algorithm of Saad and Schultz [9], and prove backward stability of the solutions, thus extending the results of Paige et al. [6]. The computed Krylov vectors maintain orthogonality to machine precision level by projection onto their orthogonal complement and this is accomplished with two Gauss-Seidel iterations in the low-synchronization Gram-Schmidt algorithms of Swirydowicz et al. [2]. The triangular matrix T_m is an approximation of the matrix $(\tilde{Q}_m^T \tilde{Q}_m)^{-1}$ and is recognized as a Neuman series. Two Gauss-Seidel iterations results in $T_m^{(2)}$ that is symmetric to $\mathcal{O}(\varepsilon)$. This matrix was split and applied across two iterations to achieve $\mathcal{O}(\varepsilon)$ orthogonality for DCGS-2. Giraud et al. [11] demonstrated how a rank-k correction could be applied in an *à posteriori* step to improve orthogonality by computing the polar decomposition of $\hat{Q}_{k-1}$, the matrix exhibited by Björck and Paige [12]. The algorithms described herein allow us to maintain the orthogonality of the normalized $\tilde{Q}_{k-1}$ at each iteration or 'on-the-fly' instead of as a post-processing step.

Our paper is organized as follows, low synchronization Gram-Schmidt algorithms are reviewed in Section 2 and two Gauss-Seidel iterations are applied to solve the normal equations $Q_{k-1}^T Q_{k-1} \mathbf{r}_{1:k-1,k} =$

$Q_{k-1}^T \mathbf{a}_k$. A rounding error analysis of the new Gram-Schmidt algorithm is presented in Section 3, leading to bounds on the representation error and the orthogonality of the columns of $\bar{Q}_{k-1}$. Section 4 extends these results to the Arnoldi-QR algorithm and post-modern GMRES. The relationship with Henrici's departure from normality is explored in Section 5. Finally, numerical experiments on challenging problems studied over the past thirty-five years are presented in Section 6.

Notation Lowercase bold letters denote a column vector and uppercase letters are matrices (e.g. $\mathbf{v}$ and A, respectively). a_{ij} represents the (i, j) scalar entry of a matrix A, and $\mathbf{a}_j$ denotes the j^{th} column of A. Where appropriate, A_j is the j–th column of a matrix. Superscripts indicate the approximate solution and corresponding residual (e.g. $\mathbf{x}^{(k)}$ and $\mathbf{r}^{(k)}$) of an iterative method at step k. Throughout this article, the notation U_k (or L_k) and U_s (or L_s) will explicitly refer to *strictly upper/lower triangular matrices*.[1] Vector notation indicates a subset of the rows and/or columns of a matrix; e.g. $V_{1:k+1,1:k}$ denotes the first $k+1$ rows and k columns of the matrix V and the notation $V_{:,1:k}$ represents the entire row of the first k columns of V. $H_{m+1,m}$ represents an $(m+1) \times m$ matrix, and in particular H refers to a Hessenberg matrix. In cases where standard notation in the literature is respected that may otherwise conflict with the aforementioned notation, this will be explicitly indicated. Bars denote computed quantities such as $\bar{Q}_{k-1}$. While $\tilde{Q}_{k-1}$ indicates a correctly (properly) normalized matrix as was introduced in Björck and Paige [12].

2. Low-synchronization Gram-Schmidt Algorithms. Krylov linear system solvers are often required for extreme scale physics simulations on parallel machines with many-core accelerators. Their strong-scaling is limited by the number and frequency of global reductions in the form of `MPI_AllReduce` and these communication patterns are expensive [13]. Low-synchronization algorithms are based on Ruhe [1], and are designed such that they require only one reduction per iteration to normalize each vector and apply projections. The Gram-Schmidt projector applied to $\mathbf{a}_k$, the k-th column of A, in the $A = QR$ factorization can be written as

$$P\mathbf{a}_k = \mathbf{a}_k - Q_{k-1}\mathbf{r}_{1:k-1,k} = \mathbf{a}_k - Q_{k-1}\,(\,Q_{k-1}^T Q_{k-1}\,)^{-1}\,Q_{k-1}^T\mathbf{a}_k,$$

where the vector $\mathbf{r}_{1:k-1,k}$ is a solution of the normal equations

$$Q_{k-1}^T Q_{k-1}\mathbf{r}_{1:k-1,k} = Q_{k-1}^T\mathbf{a}_k. \tag{2.1}$$

Ruhe [1] established that the iterated MGS algorithm employs a *multiplicative* Gauss-Seidel relaxation scheme with matrix splitting $\tilde{Q}_{k-1}^T\tilde{Q}_{k-1} = M_{k-1} - N_{k-1}$, where $M_{k-1} = I + L_{k-1}$ and $N_{k-1} = -L_{k-1}^T$ and $M_{k-1}^{-1} = T_{k-1}^{(1)}$. The iterated CGS is an *additive* Jacobi relaxation.

The inverse compact WY form for MGS was derived in Świrydowicz et al. [2], with strictly lower triangular matrix L_{k-1} . Specifically, these inverse compact WY algorithms batch the inner-products together and compute one row of L_{k-1} as

$$L_{k-1,:} = (\,Q_{k-1}^T\,\mathbf{q}_{k-1}\,)^T. \tag{2.2}$$

The resulting projector $P^{(1)}$ is given by

$$P^{(1)} = I - Q_{k-1}\,T_{k-1}^{(1)}\,Q_{k-1}^T, \quad T_{k-1}^{(1)} = (\,I + L_{k-1}\,)^{-1}$$

and corresponds to one Gauss-Seidel iteration for the normal equations (2.1). The MGS algorithm with two Gauss-Seidel iterations is given as Algorithm 2.1 below.

[1] We note that the distinction between these two notations in crucial. For U_k, the size of the strictly upper triangular matrix changes with k, whereas the size of U_s remains fixed.

Algorithm 2.1 Inverse compact WY MGS Algorithm with Normalization Lag

Input: Matrices Q_{k-1}, and R_{k-1}, $A_{k-1} = Q_{k-1}R_{k-1}$; column vector $\mathbf{a}_k$; matrix L_{k-2}
Output: Q_k and R_k, such that $A_k = Q_k R_k$; L_{k-1}, $\mathbf{w}_k$

1: if $k = 1$ return
2: $[\, L_{:,k-1}^T,\ \mathbf{r}_k\,] = Q_{k-1}^T[\mathbf{q}_{k-1}\ \mathbf{a}_k]$ ▷ Global synchronization
3: $\mathbf{r}_{k-1,k-1} = \|\mathbf{w}_{k-1}\|_2$
4: $\mathbf{q}_{k-1} = \mathbf{w}_{k-1}/\mathbf{r}_{k-1,k-1}$ ▷ Lagged normalization
5: $\mathbf{r}_{1:k-1,k}^{(0)} = \mathbf{r}_{1:k-1,k}/\mathbf{r}_{k-1,k-1}$ ▷ Scale for Arnoldi
6: $L_{:,k-1}^T = L_{:,k-1}^T/\mathbf{r}_{k-1,k-1}$ ▷ Scale for Arnoldi
7: $\mathbf{r}_{1:k-1,k}^{(1)} = (\, I + L_{k-1}\,)^{-1}\, \mathbf{r}_{1:k-1,k}^{(0)}$ ▷ First G-S
8: $\mathbf{r}_{1:k-1,k}^{(2)} = \mathbf{r}_{1:k-1,k}^{(1)} - (\, I + L_{k-1}\,)^{-1}\, L_{:,k-1}^T\, \mathbf{r}_{1:k-1,k}^{(1)} \approx \mathbf{r}_{1:k-1,k}^{(1)} - L_{:,k-1}^T\, \mathbf{r}_{1:k-1,k}^{(1)}$ ▷ Second G-S
9: $\mathbf{w}_k = \mathbf{a}_k - \tilde{Q}_{k-1}\, \mathbf{r}_{1:k-1,k}^{(2)}$

The representation error $\mathbf{e}_k^{(1)}$ in the computed projection after one Gauss-Seidel iteration, where $\mathbf{r}_{1:k-1,k}^{(1)}$ is a solution of the normal equations, can be expressed as

$$\mathbf{w}_k = \mathbf{a}_k - Q_{k-1}\,(\, I + L_{k-1}\,)^{-1}\, \mathbf{r}_{1:k-1,k}^{(0)} + \mathbf{e}_k^{(1)} \tag{2.3}$$

After two Gauss-Seidel iterations, the error $\mathbf{e}_k^{(2)}$ is given by

$$\mathbf{w}_k = \mathbf{a}_k - Q_{k-1}\, \mathbf{r}_{1:k-1,k}^{(1)} - Q_{k-1}\,(\, I + L_{k-1}\,)^{-1}\, L_{k-1}^T\, \mathbf{r}_{1:k-1,k}^{(1)} + \mathbf{e}_k^{(2)} \tag{2.4}$$

The correction matrix for two Gauss-Seidel iterations is found to be

$$T_{k-1}^{(2)} = M_{k-1}^{-1}\,[\, I + N_{k-1}\, M_{k-1}^{-1}\,],$$

which is no longer triangular but is close to the symmetric matrix $T_{k-1} = I - L_{k-1} - L_{k-1}^T$.

The backward error analyses of Björck [14] and Björck and Paige [12] can be applied in the case of one Gauss-Seidel iteration. In particular, the triangular solve $M_k\mathbf{x} = \mathbf{b}$ implied by Step 7 of Algorithm 2.1 with $M_k = I + L_k$ is backward stable as shown by Higham [15], where

$$(\, M_k + E_k\,)\,\mathbf{x} = \mathbf{b}, \quad \|E_k\|_2 \le \mathcal{O}(\varepsilon)\, \|M_k\|_2$$

for small $\|L_k\|_2 \le 1/2$. Therefore, it follows that the error for the computed $\bar{\mathbf{r}}_{1:k-1,k}^{(0)}$ is bounded according to

$$\bar{\mathbf{r}}_{1:k-1,k}^{(0)} = \bar{Q}_{k-1}^T\, \mathbf{a}_k + \mathbf{e}_k^{(0)}, \quad \|\mathbf{e}_k^{(0)}\|_2 \le \mathcal{O}(\varepsilon)\, \|\bar{Q}_{k-1}\|_2\, \|\mathbf{a}_k\|_2$$

as in the error analysis of the traditional MGS algorithm. The backward error $E_{k-1}^{(1)}$ for the triangular solve is bounded as follows

$$(\, M_{k-1} + E_{k-1}^{(1)}\,)\, \bar{\mathbf{r}}_{1:k-1,k}^{(1)} = \bar{\mathbf{r}}_{1:k-1,k}^{(0)}, \quad \|E_{k-1}^{(1)}\|_2 \le \mathcal{O}(\varepsilon)\, \|M_{k-1}\|_2$$

3. Backward Stability. Our next step consists of a rounding error analysis of the low-synch MGS algorithm with two Gauss-Seidel iterations. We begin with the derivation of the representation error for the QR factorization using an induction argument. Algorithm 2.1 contains the following recurrence,

$$\begin{aligned}
\mathbf{r}_{1:k-1,k}^{(0)} &= Q_{k-1}^T \mathbf{a}_k \\
\mathbf{r}_{1:k-1,k}^{(1)} &= (\, I + L_{k-1}\,)^{-1}\, \mathbf{r}_{1:k-1,k}^{(0)} = M_{k-1}^{-1}\, \mathbf{r}_{1:k-1,k}^{(0)} \\
\mathbf{r}_{1:k-1,k}^{(2)} &= \mathbf{r}_{1:k-1,k}^{(1)} - (\, I + L_{k-1}\,)^{-1}\, L_{k-1}^T\, \mathbf{r}_{1:k-1,k}^{(1)} = \mathbf{r}_{1:k-1,k}^{(1)} + M_{k-1}^{-1}\, N_{k-1}\, \mathbf{r}_{1:k-1,k}^{(1)}
\end{aligned}$$

where $\bar{\mathbf{r}}_{k-1,k-1} = \|\bar{\mathbf{r}}^{(2)}_{1:k-1,k}\|_2$ and $\bar{\mathbf{a}}_k = \bar{\mathbf{r}}^{(2)}_{1:k-1,k}/\bar{\mathbf{r}}_{k-1,k-1}$. The associated error terms and bounds are then given by

$$
\begin{aligned}
\bar{\mathbf{r}}^{(0)}_{1:k-1,k} &= \bar{Q}^T_{k-1}\mathbf{a}_k + \mathbf{e}^{(0)}_k, && \|\mathbf{e}^{(0)}_k\|_2 \le \mathcal{O}(\varepsilon)\|\bar{Q}_{k-1}\|_2\,\|\mathbf{a}_k\|_2 \\
(\, M_{k-1} + E^{(1)}_{k-1}\,)\,\bar{\mathbf{r}}^{(1)}_{1:k-1,k} &= \bar{\mathbf{r}}^{(0)}_{1:k-1,k}, && \|E^{(1)}_{k-1}\|_2 \le \mathcal{O}(\varepsilon)\|M_{k-1}\|_2 \\
\bar{\mathbf{r}}^{(1/2)}_{1:k-1,k} &= N^T_{k-1}\bar{\mathbf{r}}^{(1)}_{1:k-1,k} + \mathbf{e}^{(1)}_k, && \|\mathbf{e}^{(1)}_k\|_2 \le \mathcal{O}(\varepsilon)\|N_{k-1}\|_2\|\bar{\mathbf{r}}^{(1)}_{1:k-1,k}\|_2 \\
(\, M_{k-1} + E^{(2)}_{k-1}\,)\,\bar{\mathbf{r}}^{(3/2)}_{1:k-1,k} &= \bar{\mathbf{r}}^{(1/2)}_{1:k-1,k}, && \|E^{(2)}_{k-1}\|_2 \le \mathcal{O}(\varepsilon)\|M_{k-1}\|_2 \\
\bar{\mathbf{r}}^{(2)}_{1:k-1,k} &= \bar{\mathbf{r}}^{(1)}_{1:k-1,k} - \bar{\mathbf{r}}^{(3/2)}_{1:k-1,k} + \mathbf{e}^{(2)}_k, && \|\mathbf{e}^{(2)}_k\|_2 \le \mathcal{O}(\varepsilon)(\,\|\bar{\mathbf{r}}^{(1)}_{1:k-1,k}\|_2 + \|\bar{\mathbf{r}}^{(3/2)}_{1:k-1,k}\|_2\,)
\end{aligned}
$$

where

$$
\begin{aligned}
\bar{\mathbf{r}}^{(1)}_{1:k-1,k} &= M^{-1}_{k-1}\,\bar{Q}^T_{k-1}\,\mathbf{a}_k + M^{-1}_{k-1}\,\mathbf{f}^{(1)}_k, && \mathbf{f}^{(1)}_k = \mathbf{e}^{(0)}_k - \mathbf{E}^{(1)}_{k-1}\,\bar{\mathbf{r}}^{(1)}_{1:k-1,k} \\
\bar{\mathbf{r}}^{(3/2)}_{1:k-1,k} &= M^{-1}_{k-1}\,N_{k-1}\,\bar{\mathbf{r}}^{(1)}_{1:k-1,k} + M^{-1}_{k-1}\,\mathbf{f}^{(2)}_k, && \mathbf{f}^{(2)}_k = \mathbf{e}^{(1)}_k - E^{(2)}_{k-1}\,\bar{\mathbf{r}}^{(3/2)}_{1:k-1,k} \\
\bar{\mathbf{r}}^{(2)}_{1:k-1,k} &= M^{-1}_{k-1}\,\bar{Q}^T_{k-1}\,\mathbf{a}_k + M^{-1}_{k-1}\,N_{k-1}\,M^{-1}_{k-1}\,\bar{Q}^T_{k-1}\,\mathbf{a}_k + \mathbf{e}_k \\
\mathbf{e}_k &= M^{-1}_{k-1}\,\mathbf{f}^{(1)}_k + M^{-1}_{k-1}\,N_{k-1}\,M^{-1}_{k-1}\mathbf{f}^{(1)}_k + M^{-1}_{k-1}\,\mathbf{f}^{(2)}_k
\end{aligned}
$$

After one Gauss-Seidel iteration, the following bound holds

$$
\begin{aligned}
\|\bar{\mathbf{r}}^{(1)}_{1:k-1,k}\|_2 &\le (\, M_{k-1} + E^{(1)}_{k-1}\,)^{-1}\,(\,\bar{Q}^T_{k-1}\mathbf{a}_k + \mathbf{e}^{(0)}_k\,) \\
&\le \frac{\|\bar{Q}_{k-1}\|_2\,\|\mathbf{a}_k\|_2\,(\,1+\mathcal{O}(\varepsilon)\,)}{\sigma_{k-1}(M_{k-1}) - \|E^{(1)}_{k-1}\|_2} \le \frac{\|\bar{Q}_{k-1}\|_2\,\|\mathbf{a}_k\|_2\,(\,1+\mathcal{O}(\varepsilon)\,)\,\|M^{-1}_{k-1}\|_2}{1-\mathcal{O}(\varepsilon)\kappa(M_{k-1})}
\end{aligned}
$$

For the next step in Algorithm 2.1, the Gauss-Seidel iteration matrix appears explicitly

$$
\bar{\mathbf{r}}^{(3/2)}_{1:k-1,k} = M^{-1}_{k-1}\,N_{k-1}\,\bar{\mathbf{r}}^{(1)}_{1:k-1,k} + M^{-1}_{k-1}\,\mathbf{f}^{(2)}_k
$$

and an upper bound is given by

$$
\begin{aligned}
\|\bar{\mathbf{r}}^{(3/2)}_{1:k-1,k}\|_2 &\le \frac{\|M^{-1}_{k-1}\,N_{k-1}\|_2\,\|\bar{\mathbf{r}}^{(1)}_{1:k-1,k}\|_2 + \|\mathbf{e}^{(1)}_k\|_2}{1-\|M^{-1}_{k-1}\|_2\,\|E^{(2)}_{k-1}\|_2} \\
&\le \frac{\|M^{-1}_{k-1}\,N_{k-1}\|_2 + \mathcal{O}(\varepsilon)\|N_{k-1}\|_2}{1-\mathcal{O}(\varepsilon)\kappa(M_{k-1})}\|\bar{\mathbf{r}}^{(1)}_{1:k-1,k}\|_2
\end{aligned}
$$

In order to bound the 2-norm of the iteration matrix, we have $\|M^{-1}_{k-1}N_{k-1}\|_2 < \|M^{-1}_{k-1}\|_2\,\|N_{k-1}\|_2$, and assume that $\|N_{k-1}\|_2 < 1/2$, where $\|N_{k-1}\|_2 \le \|N_{k-1}\|_2/(\,1-\|N_{k-1}\|_2\,)$ and the result follows. The individual terms in $\mathbf{e}_k$ are then bounded as follows

$$
\begin{aligned}
\|\mathbf{f}^{(1)}_k\|_2 &\le \|\mathbf{e}^{(0)}_k\|_2 + \|E^{(1)}_{k-1}\|_2\,\|\bar{\mathbf{r}}^{(1)}_{1:k-1,k}\|_2 \\
&\le \mathcal{O}(\varepsilon)\|\bar{Q}_{k-1}\|_2\,\|\mathbf{a}_k\|_2 + \frac{\mathcal{O}(\varepsilon)\kappa(M_{k-1})\,\|\bar{Q}_{k-1}\|_2\,\|\mathbf{a}_k\|_2}{1-\mathcal{O}(\varepsilon)\kappa(M_{k-1})}
\end{aligned}
$$

from which it follows that

$$
\|M^{-1}_{k-1}\|_2\,\|\mathbf{f}^{(1)}_k\|_2 \le \frac{\mathcal{O}(\varepsilon)\kappa(M_{k-1})}{1-\mathcal{O}(\varepsilon)\kappa(M_{k-1})}\,\|\bar{Q}_{k-1}\|_2\,\|\mathbf{a}_k\|_2.
$$

For the third term in the error, a bound follows from

$$
\begin{aligned}
\|\mathbf{f}^{(2)}_k\|_2 &\le \|\mathbf{e}^{(1)}_k\|_2 + \|E^{(2)}_{k-1}\|_2\,\|\bar{\mathbf{r}}^{(3/2)}_{1:k-1,k}\|_2 \\
&\le \mathcal{O}(\varepsilon)\|N_{k-1}\|_2\,\|\bar{\mathbf{r}}^{(1)}_{1:k-1,k}\|_2 + \frac{\mathcal{O}(\varepsilon)\kappa(M_{k-1})\,(\,1+\mathcal{O}(\varepsilon)\|N_{k-1}\|_2\,\|\bar{\mathbf{r}}^{(1)}_{1:k-1,k}\|_2)}{1-\mathcal{O}(\varepsilon)\kappa(M_{k-1})}
\end{aligned}
$$

and therefore the last term is given by

$$\|M_{k-1}^{-1}\|_2 \, \|\mathbf{f}_k^{(2)}\|_2 \leq \frac{\mathcal{O}(\varepsilon)\kappa(M_{k-1})}{[\, 1 - \mathcal{O}(\varepsilon)\kappa(M_{k-1})\,]^2} \, \|\bar{Q}_{k-1}\|_2 \, \|\mathbf{a}_k\|_2.$$

The bounds for the remaining computed vectors in the representation error can now be determined.

$$\begin{aligned} M_{k-1}\, \bar{\mathbf{r}}_{1:k-1,k}^{(1)} &= \bar{Q}_{k-1}^T\, \mathbf{a}_k + \mathbf{e}_k^{(0)} - E_{k-1}^{(1)}\, \bar{\mathbf{r}}_{1:k-1,k}^{(1)} = \bar{Q}_{k-1}^T\, \mathbf{a}_k + \mathbf{f}_k^{(1)} \\ M_{k-1}\, \bar{\mathbf{r}}_{1:k-1,k}^{(3/2)} &= N_{k-1}\, \bar{\mathbf{r}}_{1:k-1,k}^{(1)} + \mathbf{e}_k^{(1)} - E_{k-1}^{(2)}\, \bar{\mathbf{r}}_{1:k-1,k}^{(3/2)} = N_{k-1}\, \bar{\mathbf{r}}_{1:k-1,k}^{(1)} + \mathbf{f}_k^{(2)} \end{aligned}$$

Finally, the columns of the computed $\bar{R}$ are bounded according to

$$\|\bar{\mathbf{r}}_{1:k-1,k}^{(2)}\|_2 \leq \|M_{k-1}^{-1}(\, I + N_{k-1}M_{k-1}^{-1}\,)\|_2 \, \|\bar{Q}_{k-1}\|_2 \, \|\mathbf{a}_k\|_2 + \|\mathbf{e}_k\|_2 \tag{3.1}$$

The computed form of the projection in Step 9 of Algorithm 2.1 can be written as

$$\bar{\mathbf{w}}_k = \mathbf{a}_k - \bar{Q}_{k-1}\, \bar{\mathbf{r}}_{1:k-1,k}^{(2)} + \mathbf{f}_k$$

where

$$\|\mathbf{f}_k\|_2 \leq \mathcal{O}(\varepsilon) \left[\, \|\mathbf{a}_k\|_2 + \|\bar{Q}_{k-1}\|_2 \, \|\bar{\mathbf{r}}_{1:k-1,k}^{(2)}\|_2 \,\right]$$

From equation (3.1), it therefore follows that

$$\begin{aligned} \|\mathbf{f}_k\|_2 &\leq \mathcal{O}(\varepsilon) \left[\, \|\mathbf{a}_k\|_2 + \|M_{k-1}^{-1}\|_2 \, \|\bar{Q}_{k-1}\|_2^2 \, \|\mathbf{a}_k\|_2 \,\right] + \mathcal{O}(\varepsilon^2) \\ &\leq \mathcal{O}(\varepsilon)\|\mathbf{a}_k\|_2 \left[\, 1 + \frac{\|\bar{Q}_{k-1}\|_2^2}{1 - \|N_{k-1}\|_2} \,\right] \leq \frac{\mathcal{O}(\varepsilon)\|\bar{Q}_{k-1}\|_2^2 \, \|\mathbf{a}_k\|_2}{1 - \|N_{k-1}\|_2} \end{aligned}$$

The representation error for the QR factorization is now determined from the following equations

$$\begin{aligned} A_{k-1} &= \bar{Q}_{k-1}\, \bar{R}_{k-1}^{(2)} - F_{k-1} \\ \mathbf{a}_k &= \bar{Q}_{k-1}\, \bar{\mathbf{r}}_{1:k-1,k}^{(2)} + \bar{\mathbf{w}}_k - \mathbf{f}_k \end{aligned}$$

Combining these into an augmented matrix form, we obtain

$$A_k = \left[\; A_{k-1}, \quad \mathbf{a}_k \;\right] = \bar{Q}_{k-1} \left[\; \bar{R}_{k-1}, \quad \bar{\mathbf{r}}_{1:k-1,k}^{(2)} \;\right] + \left[\; -F_{k-1}\, \bar{\mathbf{w}}_k, \quad -\mathbf{f}_k \;\right]$$

and the representation error is bounded as follows

$$||F_k|| = \left\|\left[\; F_{k-1}, \quad \mathbf{f}_k \;\right]\right\|_2 + \|\bar{\mathbf{w}}_k\|_2 \geq \left\|\left[\; -F_{k-1}, \quad \bar{\mathbf{w}}_k - \mathbf{f}_k \;\right]\right\|_2 \geq \sigma_k(A_k)$$

Thus, the 2-norm of the projected vector $\bar{\mathbf{w}}_k$ is bounded below by

$$\|\bar{\mathbf{w}}_k\|_2 \geq \sigma_k(A_k) - \|F_k\|_2$$

4. Loss of Orthogonality. The loss of orthogonality at the k–th iteration of the Gram-Schmidt algorithm with two Gauss-Seidel iterations is characterized by the 2-norm of the vector

$$\|\bar{Q}_{k-1}^T\, \bar{\mathbf{q}}_k \,\|_2 = \|\bar{Q}_{k-1}^T\, \bar{\mathbf{w}}_k \,\|_2 \,/\, \|\bar{\mathbf{w}}_k \,\|_2$$

From our error analysis, the projected vector is expanded as follows

$$\begin{aligned} \bar{Q}_{k-1}^T\, \bar{\mathbf{w}}_k &= \bar{Q}_{k-1}^T \left[\, \mathbf{a}_k - \bar{Q}_{k-1}\, \bar{\mathbf{r}}_{1:k-1,k}^{(2)} + \mathbf{f}_k \,\right] \\ &= \bar{Q}_{k-1}^T \left[\, \mathbf{a}_k - \bar{Q}_{k-1} M_{k-1}^{-1}\, (\, I + N_{k-1}\, M_{k-1}^{-1}\,)\, \bar{Q}_{k-1}^T\, a_{k-1} \,\right] \\ &\quad - \bar{Q}_{k-1}^T \bar{Q}_{k-1}\, \mathbf{e}_k + \bar{Q}_{k-1}^T\, \mathbf{f}_k \end{aligned}$$

We have already established that $\mathbf{f}_k$ is bounded, however, at this point the residual of the normal equations can be identified, and they are solved to $\mathcal{O}(\varepsilon)$

$$\tau = \left\| \bar{Q}_{k-1}^T \mathbf{a}_k - \bar{Q}_{k-1}^T \bar{Q}_{k-1} \bar{\mathbf{r}}_{1:k-1,k}^{(2)} \right\|_2 \leq \mathcal{O}(\varepsilon) \, \|\bar{Q}_{k-1}\|_2 \, \|\mathbf{a}_k\|_2$$

The largest singular value $\|\bar{Q}_{k-1}\|_2^2 \leq 1 + 2\|L_{k-1}\|_2$ and $1 - \|N_{k-1}\|_2 \leq 1$. Therefore, an appropriate bound on the loss of orthogonality at the k-th iteration of Algorithm 2.1 is given by

$$\|\bar{Q}_{k-1}^T \, \bar{\mathbf{q}}_k \, \|_2 \lesssim \mathcal{O}(\varepsilon) \|\mathbf{a}_k \, \|_2 \, / \, \bar{\mathbf{r}}_{k,k}^{(2)}$$

Let A be an $n \times n$ real-valued matrix, and consider the Arnoldi expansion of the matrix A. After k steps, in exact arithmetic, the algorithm produces the factorization

$$AV_k = V_{k+1} \, H_{k+1,k}, \quad V_{k+1}^T V_{k+1} = I_{k+1},$$

where $H_{k+1,k}$ is an upper Hessenberg matrix. When applied to the linear system $A\mathbf{x} = \mathbf{b}$, assume $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{r}_0 = \mathbf{b}$, $\|\mathbf{b}\|_2 = \rho$ and $\mathbf{v}_1 = \mathbf{b}/\rho$. The Arnoldi algorithm produces an orthogonal basis for the Krylov vectors spanned by the columns of the matrix V_k.

Consider the properly normalized matrix $\tilde{V}_k$ with Krylov vectors as columns. The strictly lower triangular matrix L_k is computed incrementally one row per iteration and is obtained from the loss of orthogonality relation

$$\tilde{V}_k^T \tilde{V}_k = I + L_k + L_k^T.$$

The essential result is based on the QR factorization of the matrix

$$B = [\, \mathbf{r}_0, \; A\mathbf{v}_k \,] = V_{k+1} \, [\, \mathbf{e}_1 \rho, \; H_{k+1,k} \,].$$

Our backward error analysis is now applied to the Arnoldi-QR algorithm, where the vector $A\mathbf{v}_k$ is projected onto the orthogonal complement of the computed Krylov vectors. Two Gauss-Seidel iterations reduce the error in equation (2.3)

$$\bar{\mathbf{w}}_{k+1} = A\tilde{\mathbf{v}}_k - \tilde{V}_k \, \bar{h}_{1:k,k} + \mathbf{f}_k \tag{4.1}$$

For Arnoldi-QR, we multiply A times $\mathbf{v}_k$ at iteration k. In effect, this is MGS (or our iterated Gauss-Seidel MGS) with $\mathbf{a}_k$ in Algorithm 2.1 is replaced by $A\mathbf{v}_k$. By applying the Arnoldi-QR recurrence with $A\mathbf{v}_{k-1}$ in the Gram-Schmidt algorithm, we define the column vectors $\mathbf{h}_{1:k-1,k} \equiv \mathbf{r}_{1:k-1,k}^{(2)}$, and the representation error E for the computed Arnoldi expansion

$$A\tilde{V}_m - \tilde{V}_{m+1}\bar{H}_{m+1,m} = E_m$$

is a matrix that grows by one column in size at each iteration. It is important to note that the Arnoldi expansion represents the underlying recurrence relation based on the Krylov subspace, $\mathcal{K}_m(\, A, \, \mathbf{r}_0 \,)$ and equation (4.1) is one column of the Arnoldi expansion. The bound $\|\bar{\mathbf{f}}_k\|_2 \leq \mathcal{O}(\varepsilon)\|\bar{\mathbf{a}}_k\|_2$ in the error analysis is given above and thus a bound on the loss-of-orthogonality is given by

$$\| \, I - \tilde{V}_k^T \tilde{V}_k \, \|_2 \lesssim \mathcal{O}(\varepsilon) \, \|A\tilde{\mathbf{v}}_\mathbf{k}\|_2 / \bar{h}_{k+1,k}$$

In practice, this bound is close to $\|N_k\|_2$ and is closely related to the departure from normality

$$dep(T_k^{-1})^2 = dep(\bar{Q}_{k-1}^T \bar{Q}_{k-1})^2 = \|L_k\|_F^2$$

5. Departure from Normality. A normal matrix $A \in \mathbb{C}^{n\times n}$ satisfies $A^*A = AA^*$. In the present study we consider Henrici's definition of the departure from normality

$$dep(A) = \sqrt{\|A\|_F^2 - \|D\|_F^2}, \tag{5.1}$$

where $D \in \mathbb{C}^{n\times n}$ is the diagonal matrix containing the eigenvalues of A [16] serves as a useful metric for the loss of orthogonality. While we find practical use for this metric for measuring the degree of

(non)normality of a matrix, there are of course other useful metrics to describe (non)normality. We refer the reader to [16–18] and references therein. In particular, we have that the loss of orthogonality is signaled by the departure from normality of T_k^{-1} as follows

$$dep(T_k^{-1})^2 = dep(I + L_k)^2 = \|I + L_k\|_F^2 - \|I\|_F^2 = \|I\|_F^2 + \|L_k\|_F^2 - k = \|L_k\|_F^2$$

We can then relate this to $S_k = (\, I + U_k\,)^{-1}\, U_k$ as we observe in practice that $\|S_k\|_F = \|L_k\|_F$ up to the first order in ε (noting that due to symmetry $\|L_k\|_F = \|U_k\|_F$); see Figure 2.

Ipsen [17] characterizes the convergence of GMRES in terms the (non)normality of the matrix A. A numerical measure of normality was presented by Ruhe [19], where normality implies that the eigenvalues and singular values of a matrix A are the same, namely $\sigma_i(A) = |\lambda_i(A)|$ for a certain ordering of the eigenvalues. Ruhe [19] establishes the connection between Henrici's metric and the distance between the eigenvalues and singular values

$$dep^2 = \sum_i \sigma_i^2 - \sum_i |\lambda_i|^2$$

The loss of orthogonality in GMRES is signaled by $\sigma_{\max}^2(\, I + L_{k-1}\,)$ increasing above one and this closely follows $\|S_k\|_2$.

6. Post-Modern GMRES. The MGS–GMRES orthogonalization algorithm is the QR factorization of a matrix B formed by adding a new column to $\mathbf{v}_k$ in each iteration. For the PM-GMRES Algorithm 2.1 below, note that bold-face (e.g. $\mathbf{r}_0$) denotes the residual vector at iteration k and subscripting (e.g. $\mathbf{r}_{1:k,k+1}$) denotes the corresponding element in the upper triangular matrix R.

The MGS-GMRES algorithm was proven to be backward stable for the solution of linear systems $A\mathbf{x} = \mathbf{b}$ in [6] and orthogonality is maintained to $\mathcal{O}(\varepsilon)\kappa(B)$, depending upon the condition number of the matrix $B = [\mathbf{r}_0,\ A\mathbf{v}_k]$. The normalization for the Krylov vector $\mathbf{v}_k$ at iteration k represents the delayed scaling of the vector $\mathbf{v}_{k-1}$ in the matrix-vector product $\mathbf{v}_k = A\mathbf{w}_{k-1}$. Therefore, an additional Step 8 is required in the low-synch Algorithm 6.1, $\mathbf{r}_{1:k-1,k} = \mathbf{r}_{1:k-1,k}/\mathbf{r}_{k-1,k-1}$ and $\mathbf{v}_{k-1} = \mathbf{v}_k/\mathbf{r}_{k-1,k-1}$. The diagonal element $\mathbf{r}_{k-1,k-1}$ of the R matrix, corresponds to $h_{k,k-1}$ in the Arnoldi QR factorization of the matrix B, and is updated after the MGS projection in Step 12 of the PM-GMRES Algorithm 6.1. Algorithm 6.1 applies the close to symmetric matrix $T_k^{(2)}$ in Step 11 as two Gauss-Seidel iterations in the form of a lower triangular solve and matrix-vector multiply where the matrix is dimension k.

Algorithm 6.1 Low-synchronization PM-GMRES

1: $\mathbf{r}_0 = \mathbf{b} - A\mathbf{x}_0$, $\mathbf{v}_1 = \mathbf{r}_0$.
2: **for** $k = 1, 2, \ldots$ **do**
3: $\quad \mathbf{v}_k = A\mathbf{w}_{k-1}$ ▷ Matrix-vector product
4: $\quad [\, L_{:,k-1}^T,\ \mathbf{r}_k\,] = V_{k-1}^T[\mathbf{w}_{k-1}\ \mathbf{v}_k]$ ▷ Global AllReduce
5: $\quad \mathbf{r}_{k-1,k-1} = \|\mathbf{v}_k\|_2$
6: $\quad \mathbf{r}_{1:k,k+1} = \mathbf{r}_{1:k,k+1}/\mathbf{r}_{k-1,k-1}$ ▷ Scale for Arnoldi
7: $\quad \mathbf{v}_{k-1} = \mathbf{v}_{k-1}/\mathbf{r}_{k-1,k-1}$ ▷ Scale for Arnoldi
8: $\quad L_{:,k-1}^T = L_{:,k-1}^T/\mathbf{r}_{k-1,k-1}$
9: $\quad \mathbf{r}_{1:k-1,k} = T_{k-1}^{(2)}\, \mathbf{r}_{1:k-1,k}$ ▷ projection
10: $\quad \mathbf{w}_k = \mathbf{v}_k - V_{k-1}\, \mathbf{r}_{1:k-1,k}$
11: $\quad H_k = R_{k-1}$
12: $\quad$ Apply Givens rotations to H_k
13: **end for**
14: $\mathbf{y}_m = \text{argmin}\|(\, H_m\mathbf{y}_m - \|\mathbf{r}_0\|_2\mathbf{e}_1\,)\|_2$
15: $\mathbf{x} = \mathbf{x}_0 + V_m\mathbf{y}_m$

The low-synch DCGS-2 algorithm introduced by Swirydowicz [2] was recently applied to the Arnoldi-QR factorization in the Krylov-Schur eigenvalue and GMRES solvers [7]. Although a formal backward stability analysis is not yet available, the algorithm exhibits desirable numerical characteristics including the computation of invariant subspaces of maximal size for the Krylov-Schur algorithm [20]. However, the Arnoldi algorithm was modified to account for the delayed re-orthogonalization and it was not

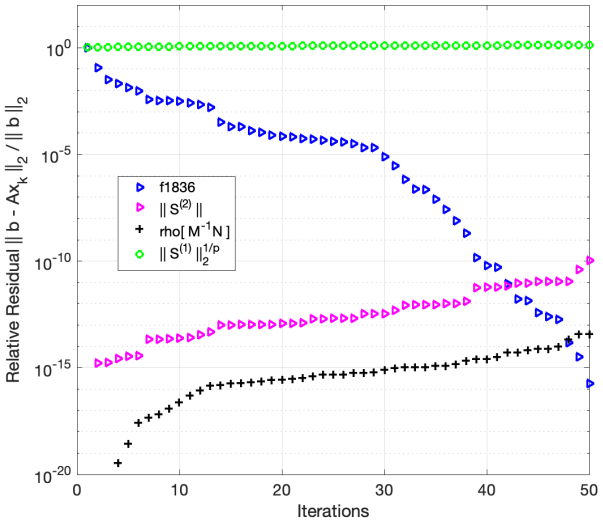

Fig. 1: GMRES residual for fs1836 matrix.

obvious how to adapt this approach to restarted GMRES. The PM-GMRES algorithm presented herein does not require any modifications to the basic Arnoldi-QR iteration.

Our first numerical experiment illustrates that the bounds derived in the previous sections are tight and properly capture the numerical behavior of the PM-GMRES algorithm. In particular, we examine the `fs1836` matrix studied by Paige and Strakoš [5]. Our Figure 1 should be compared with Figures 7.1 and 7.3 of their paper. In order to demonstrate empirically that the backward error is reduced by the iteration matrix $M_{k-1}^{-1}N_{k-1}$, the quantity $\|S^{(1)}\|_2^{1/p_k}$ is plotted, where the $\|S^{(1)}\|_2$ measures the loss of orthogonality for one Gauss-Seidel iteration in MGS, as defined by Paige et al. [6]. Recall that $p_m = -\log_{10}\rho$, and ρ the spectral radius of the matrix $M_m^{-1}N_m$. The result is a constant value of one (1), indicating that two Gauss-Seidel iterations would be sufficient to reduce the loss of orthogonality to $\mathcal{O}(\varepsilon)$. The spectral radius and the metric $\|S^{(2)}\|_2$ for two Gauss-Seidel iterations are also plotted in Figure 1. Unlike the relative residual in Figure 7.1 of Paige and Strakoš [5] which stagnates at 1e−7 before reaching $\mathcal{O}(\varepsilon)$, the relative residual for PM-GMRES continues to decrease monotonically. The matrix in our tests was also column-scaled, as in the earlier work. We employed a diagonal matrix D_c with the max-norm of each column as the diagonal elements.

7. Computational Complexity and Parallel Communication. The GMRES algorithm of Saad and Schultz [9] has a thirty-five year history and several variations or alternative formulations of the basic algorithm have been proposed over that time frame. A comprehensive review of these developments is presented by Zou [21]. In particular, pipelined s-step and block algorithms have been proposed which are better able to hide latency in parallel implementations and are described in Yamazki et al. [22]. In the case of the DCGS2 algorithm, the symmetric correction matrix T_{k-1} was derived in Appendix 1 of [2] and is given by

$$T_{k-1} = I - L_{k-1} - L_{k-1}^T$$

This form of the correction matrix was employed in the s-step and pipelined GMRES.When the matrix T_{k-1} is split into $I - L_{k-1}$ and L_{k-1}^T and applied across two iterations of the DCGS2 algorithm, the resulting loss of orthogonality is $\mathcal{O}(\varepsilon)$ in this case. Indeed, it was conjectured in Bielich et al. [7] that two iterations of DCGS2 are needed to achieve $\mathcal{O}(\varepsilon)$ orthogonal vectors, however, our results demonstrate that one MGS iteration is sufficient without an additional projection step.

The low-synchronization modified Gram-Schmidt and GMRES algorithms described in Swirydowicz et al. [2] improve parallel strong-scaling by employing one global reduction for each iteration. A review of compact WY Gram Schmidt algorithms and their computational costs is given in [7]. The triangular

solve and matrix-vector multiply for the Gauss-Seidel iterations require an additional $(k-1)^2$ flops at iteration $k-1$ and thus lead to a slightly higher operation count compared to the original MGS algorithm. The matrix-vector multiply in (2.2) increases the complexity by mn^2 ($3mn^2$ total) but decreases the number of global reductions from $k-1$ at iteration k to only one reduction when combined with the lagged normalization of a Krylov vector. The above costs can be compared with the DCGS2 algorithm with delayed reorthogonalization which requires $4mn^2$ flops.

Block generalizations of the DGCS2 and CGS2 algorithm are presented in Carson et al. [23,24]. The authors generalize the Pythagorean trick to block form and derive `BCGS-PIO` and `BCGS-PIP` algorithms with the more favorable communication patterns described herein. An analysis of the backward stability of the these block Gram-Schmidt algorithms is also presented.

8. Numerical Results. Numerically challenging test problems for GMRES have been proposed and analyzed over the past 35 years. These include both symmetric and non-symmetric matrices. Simoncini and Szyld [25] introduced a symmetric, diagonal matrix with real eigenvalues, causing MGS-GMRES to stagnate. Highly-non-normal matrices from Walker [10] were used to explore the convergence characteristics of the Householder HH-GMRES and then the non-normal fs1836 from Paige et al. [6] and west0132 from Paige and Strakoš [5] encounter stagnation. In addition to these, the impcol_e matrix from Greenbaum et al. [26], reaches the $\mathcal{O}(\varepsilon)$ relative residual level level on the final iteration, unless it stagnates. Matrices with complex eigenvalues forming a disc inside the unit circle such as the Helmert matrix from Liesen and Tichy [27], are also evaluated. Results from a very large fluid mechanics pressure continuity solver with AMG preconditioner and a circuit simulation with the ADD32 matrix from Rozložník, Strakoš and Tuma [28] are also presented.

8.1. Ill-conditioned Diagonal Matrix. Simoncini and Szyld [25] consider several difficult ill-conditioned problems that can lead to stagnation of the GMRES relative residual before converging to the level of machine precision $\mathcal{O}(\varepsilon)$. In example 5.5, they construct $A = \text{diag}([1e-4, 2, 3, \ldots, 100])$, a diagonal matrix, while the right-hand side is $b = \text{randn}(100, 1)$, normalized so that $b = 1$. The condition number of this matrix is 1e+6.

With the MGS-GMRES algorithm, the relative residual stagnates at the level 1e−12 after 75 iterations, when $\|S\|_2 = 1$ indicating that the Krylov vectors are not linearly independent. In the case of the PM-GMRES algorithm, the convergence history is plotted in Figure 2, where it can be observed that the relative residual continues to decrease monotonically. Furthermore. the upper bound $\mathcal{O}(\varepsilon)\ \|A\tilde{\mathbf{v}}_\mathbf{k}\|_2/h_{k+1,k}$ is plotted along with Henrici's departure from normality $dep(T^{-1})$. The latter overlaps with the metric $\|S\|_2$ and indicates that a significant loss of orthogonality does not occur.

8.2. Ill-Conditioned Symmetric and Non-Symmetric Matrices. Figures 1.1 and 1.2 from Greenbaum et al. [26] describe the results for STEAM1 (the HH and MGS implementations, respectively). Similarly, Figures 1.3 and 1.4 correspond to IMPCOLE. They emphasize that the behaviour illustrated in these figures represents typical behaviour of the MGS and HH-GMRES. The condition number of the system matrix was $\kappa(A) = 2.855 \times 10^7$ for STEAM1 and $\kappa(A) = 7.102 \times 10^6$ for IMPCOLE.

Greenbaum et al. [26] observe that although orthogonality of the MGS vectors is not maintained near the machine precision, as for the Householder implementation, the norms of the computed residuals of the MGS-GMRES are almost identical to those of the HH-GMRES, until the smallest singular value of the matrix V_m begins to depart from the value one. At that point the MGS-GMRES residual norm begins to stagnate close to its final precision level. This observation is demonstrated on the numerical examples for matrices STEAM1 ($N = 240$, symmetric positive definite matrix used in oil recovery simulations) and IMPCOLE ($N = 225$, nonsymmetric matrix from modelling of the hydrocarbon separation problem) (Figures 1.1-1.4). In both experiments $x = (1, ..., 1)^T$, $b = Ax$ and $x_0 = 0$.

The convergence histories for the PM-GMRES algorithm applied to these matrices are plotted in Figures 7 and Steam1. The impcol_e matrix was also diagonally column scaled as in other tests. A significant loss of orthogonality is not observed until the last iteration at convergence. Otherwise the computed metric $\|S\|_2$ and the associated bound remain at $\mathcal{O}(\varepsilon)$.

8.3. Highly Non-Normal Matrices. Bidiagonal matrices with a δ off-diagonal were studied by Embree [29]. These are non-normal matrices where $0 < \delta \leq 1$ and also defective. The pseudo-spectra [18] of these matrices are discs in the complex plane. Our PM-GMRES algorithm leads to convergence after

16 iterations without stagnation and orthogonality is maintained to machine precision as plotted in Figure 3.

Walker [10] employed the highly non-normal matrix below to compare the Gram-Schmidt and Householder implementations of GMRES. The element α controls both the condition $\kappa(A)$ and departure from normality $dep(A)$ of the matrix.

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 & \alpha \\ 0 & 2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & n \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

For large values of α, Walker found that the MGS-GMRES relative residual would stagnate and that the CGS algorithm led to instability. Furthermore, it was found that even CGS-2 with re-orthogonalization exhibited some instability near convergence. The Householder HH-GMRES maintains $\mathcal{O}(\varepsilon)$ orthogonality as measured by $\|I - \mathbf{v}_k^T \mathbf{v}_k\|_F$ and reduces the relative residual to machine precision.

In our experiments, the value $\alpha = 2000$ leads to a matrix with $\kappa(A) = 4e + 5$. The departure from normality, based on Henrici's metric, is large $dep(A) = 2000$. The convergence history for PM-GMRES is displayed in Figure 4 where the matrix A has been column scaled. The loss of orthogonality as measured by $\|S\|_2$ remains close to $O(\varepsilon)$ and our upper bound is tight for this problem.

Paige and Strakoš [5] study two highly non-normal matrices, FS1836 and WEST0132. In all their experiments $b = (1, \ldots, 1)^T$. Results for the matrix FS1836 with $n = 183$, $\|A\|_2 \approx 1.2 \times 10^9$, $\kappa(A) \approx 1.5 \times 10^{11}$.For the matrix WEST0132 with n = 132, $\|A\|_2 \approx 3.2 \times 10^5$, $\kappa(A) \approx 6.4 \times 10^{11}$. The MGS-GMRES algorithm is employed in all the experimental results reported by the authors. Their Figure 7.1 indicates that the relative residual for FS1836 stagnates at 1e−7 at iteration 43 when orthogonality is lost. The relative residual for the WEST0132 matrix also stagnates at the 1e−7 level after 130 iterations.

These results contrast with our Figures 5 and 8. In both cases the relative residuals continue to decrease monotonically and metric $\|S\|_2$ either grows slowly or remains close to machine precision. For both cases, the matrices have been column-scaled by a diagonal matrix D_c containing the column max-norms.

8.4. Complex Eigenvalues in a Disc. For their final experiment, Liesen and Tichy [27] use the Helmert matrix generated by the Matlab command `gallery('orthog',18,4)`. Helmert matrices occur in a number of practical problems, for example in applied statistics. Their matrix is orthogonal, and the eigenvalues cluster around -1, as in the right part of their Figure 4.4. The worst-case GMRES residual norm decreases quickly throughout the iterations and stagnates at the 12–th iteration, where the relative residual remains at 1e−10. From the PM-GMRES convergence history plotted in Figure 10, the loss of orthogonality as measured by $\|S\|_2$ remains close to machine precision and the relative residual does not stagnate. The bound $\mathcal{O}(\varepsilon)\ \|A\tilde{\mathbf{v}}_\mathbf{k}\|_2/h_{k+1,k}$ is an excellent predictor of the metric $\|S\|_2$.

8.5. Nalu-Wind Model. Nalu-Wind solves the incompressible Navier-Stokes equations, with a pressure projection scheme. The governing equations are discretized in time with a BDF-2 integrator, where an outer Picard fixed-point iteration is employed to reduce the nonlinear system residual at each time step. Within each time step, the Nalu-Wind simulation time is often dominated by the time required to setup and solve the linearized governing equations. The pressure systems are solved using PM-GMRES with an AMG preconditioner, where a polynomial Gauss-Seidel smoother is now applied as described in Mullowney et al. [30]. Hence, Gauss-Seidel is a compute time intensive component, when employed as a smoother within an AMG V-cycle.

The McAlister experiment for wind-turbine blades is an unsteady RANS simulation of a fixed-wing, with a NACA0015 cross section, operating in uniform inflow. Resolving the high-Reynolds number boundary layer over the wing surface requires resolutions of $\mathcal{O}(10^{-5})$ normal to the surface resulting in grid cell aspect ratios of $\mathcal{O}(40,000)$. These high aspect ratios present a significant challenge. The simulations were performed for the wing at 12 degree angle of attack, 1 m chord length, denoted c, 3.3 aspect ratio, i.e., $s = 3.3c$, and square wing tip. The inflow velocity is $u_\infty = 46$ m/s, the density is $\rho_\infty = 1.225$ kg/m^3, and dynamic viscosity is $\mu = 3.756 \times 10^{-5}$ kg/(m s), leading to a Reynolds number, $Re = 1.5 \times 10^6$. Due to the complexity of mesh generation, only one mesh with approximately 3 million grid points was generated.

The smoother is hybrid block-Jacobi with two sweeps of polynomial Gauss-Seidel applied locally on an MPI rank and then Jacobi smoothing for globally shared degrees of freedom. The coarsening rate for the wing simulation is roughly $4\times$ with eight levels in the V-cycle for hypre [31]. Operator complexity C is close to 1.6 indicating more efficient V-cycles with aggressive coarsening, however, an increased number of GMRES iterations are required compared to standard coarsening. The convergence history is plotted in Figure 9, where the loss of orthogonality is completely flat and close to machine precision.

8.6. Circuit simulation. Rozložník et al. [28] study a typical linear system arising in circuit simulation (the matrix from a 32-bit adder design). In exact arithmetic the Arnoldi vectors are orthogonal. However, in finite precision computation the orthogonality is lost, which may potentially affect both the convergence rate and the ultimate attainable accuracy of the computed approximate solution. In their Figure 3, the authors have plotted the loss of orthogonality of the computed Krylov vectors for different implementations of the GMRES method (MGS, Householder and CGS). The equivalent results for the PM-GMRES algorithm are plotted in Figure 11, where the loss of orthogonality is identical to the Householder HH-GMRES solver.

9. Conclusions. The essential contribution of our work was to derive a post-modern (viz. not classical) formulation of the GMRES algorithm that employs an iterated solution of the normal equations appearing in the Gram-Schmidt projector, as described by Ruhe [1], and the low-synchronization algorithms introduced by Swirydowicz et al. [2]. The essential idea developed here was to project the vector $A\mathbf{v}_k$ onto the orthogonal complement of the space spanned by the computed Krylov vectors represented by the columns of $\tilde{V}_m \in \mathbb{C}^{n\times m}$,.

The insights of Ruhe [1] led to the conclusion that the iterated modified Gram-Schmidt algorithm was in fact a Gauss-Seidel iteration for the normal equations $Q_{k-1}^T Q_{k-1}\, r = Q_{k-1}^T a$. The Gram-Schmidt projector is then given by $Pa = a - Q_{k-1}\, T_{k-1}^{(1)}\, Q_{k-1}^T a$, with lower triangular matrix $T_{k-1}^{(1)} \approx (\, \tilde{Q}_{k-1}^T \tilde{Q}_{k-1}\,)^{-1}$. Swirydowicz et al. [2] identified the ICWY form of MGS with $T_{k-1}^{(1)} = (\, I + L_{k-1}\,)^{-1}$, where the strictly lower triangular matrix L_{k-1} comes from the loss of orthogonality relation

$$\tilde{Q}_{k-1}^T \tilde{Q}_{k-1} = I + L_{k-1} + L_{k-1}^T.$$

where $\tilde{Q}_{k-1}$ is the correctly (properly) normalized matrix as described in Björck and Paige [12]. The matrix $T_{k-1}^{(1)}$ was also present, but not yet defined, in the error analysis of Björck [14], in his Lemma 5.1, as given by

$$\bar{Q} = \tilde{Q}\,(\, I + U\,)$$

In effect, the low-synch MGS algorithm presented in [2] represents one Gauss-Seidel iteration to construct the projector. The iteration can be applied twice, resulting in a correction matrix that is close to a symmetric matrix

$$T_{k-1}^{(2)} = M_{k-1}^{-1}\,[\, I + N_{k-1}\, M_{k-1}^{-1}\,] = (\, I + L_{k-1}\,)^{-1} - L_{k-1}^T$$

We have employed the iterated low-synch MGS algorithm without a re-orthogonalization step for the Arnoldi-QR algorithm which forms the basis of a post-modern GMRES.

When applied in the context of the Arnoldi-QR algorithm, two iterations of Gauss-Seidel relaxation have the effect of embedding the iteration matrix $M_{k-1}^{-1} N_{k-1}$ into the columns of the representation error for the Arnoldi expansion $AV_m = V_{m+1} H_{m+1,m}$, thereby reducing the bound on the loss of orthogonality of the Krylov vectors to $\mathcal{O}(\varepsilon)\, \|A\tilde{\mathbf{v}}_\mathbf{k}\|_2 / h_{k+1,k}$ where $B = [\mathbf{r}_0,\, AV_m]$. This is related to recent work on the iterative solution of triangular linear systems using Jacobi iterations. The Jacobi iterations can diverge for highly non-normal matrices. Here, the departure from normality $dep(T_{k-1}^{-1})$ is an indicator of the loss of orthogonality. Both of these indicate a loss of numerical rank for the Krylov vectors with the smallest singular value decreasing from one. Our numerical experiments, on challenging problems proposed over the past thirty-five years, demonstrate that the relative residual does not stagnate above the level $O(\varepsilon)$. Furthermore, the loss of orthogonality remains bounded and close to machine precision.

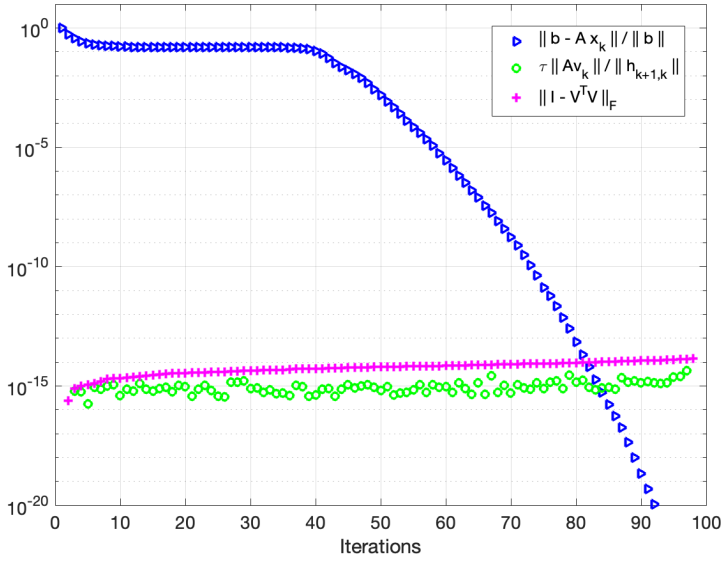

Fig. 2: GMRES relative residual for Simoncini matrix using two Gauss-Seidel iterations.

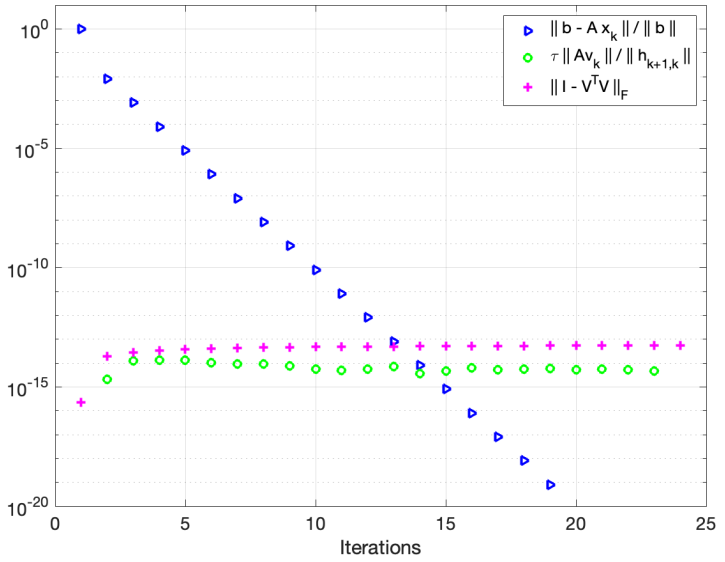

Fig. 3: GMRES relative residual for Embree bidiagonal δ matrix.

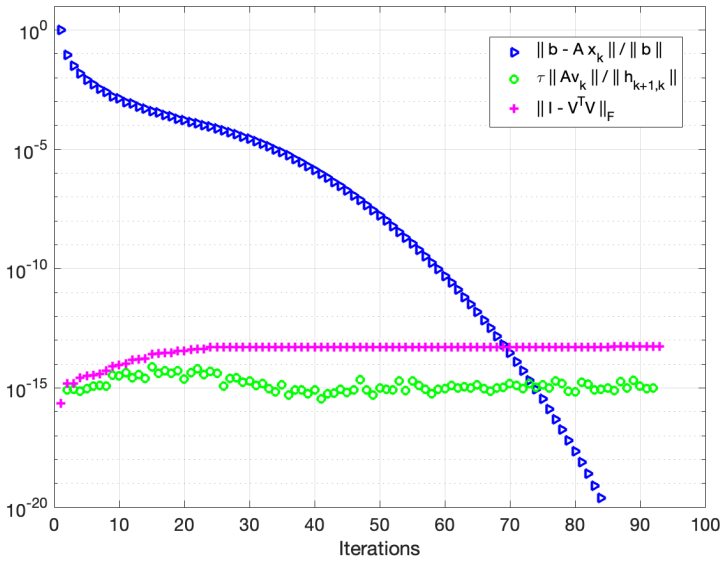

Fig. 4: GMRES relative residual for Walker matrix.

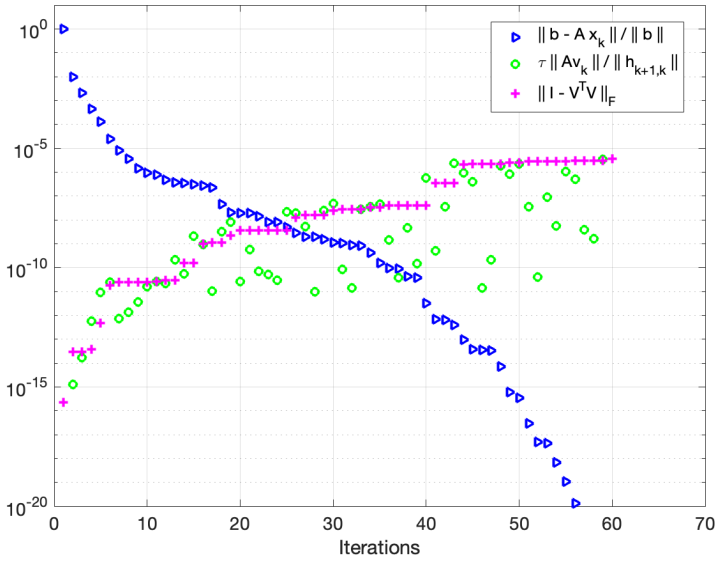

Fig. 5: GMRES residual for fs1863 matrix .

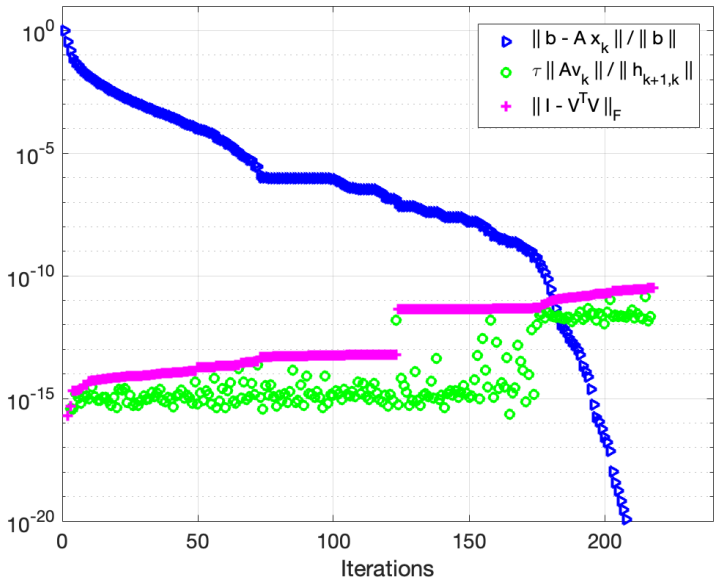

Fig. 6: GMRES residual for steam1 matrix .

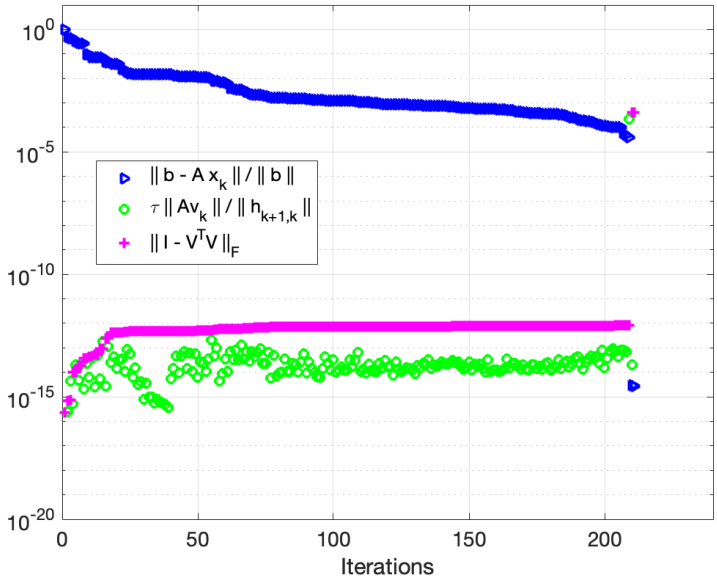

Fig. 7: GMRES residual for impcol_e matrix .

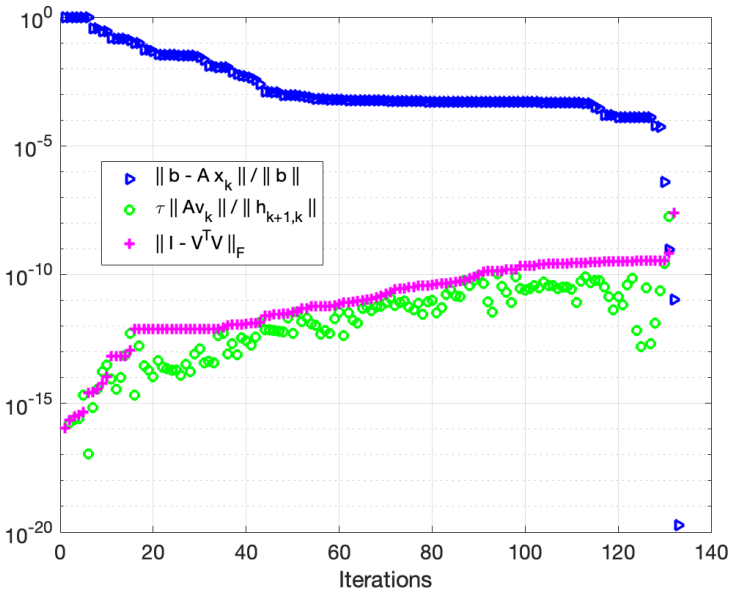

Fig. 8: GMRES residual for west0132 matrix .

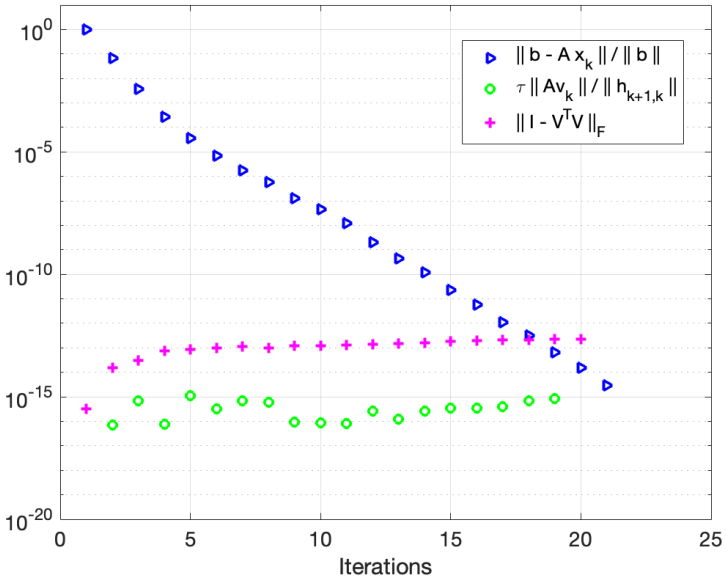

Fig. 9: Pressure-continuity GMRES+AMG

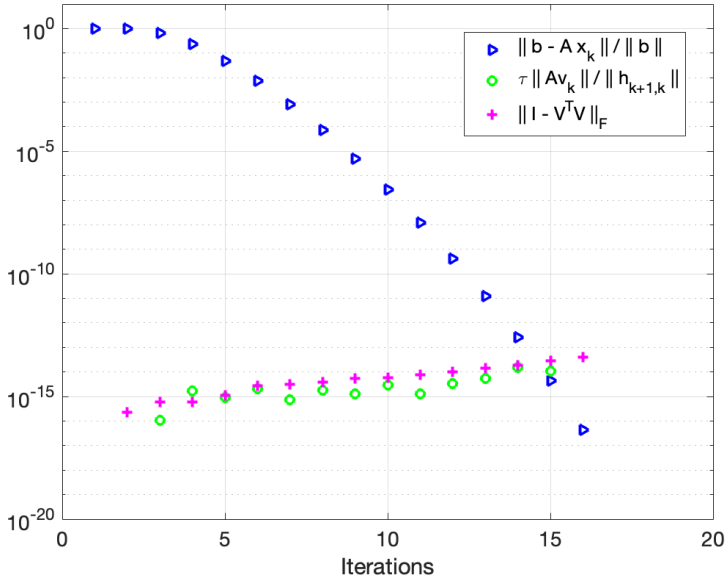

Fig. 10: GMRES residual for Helmert matrix

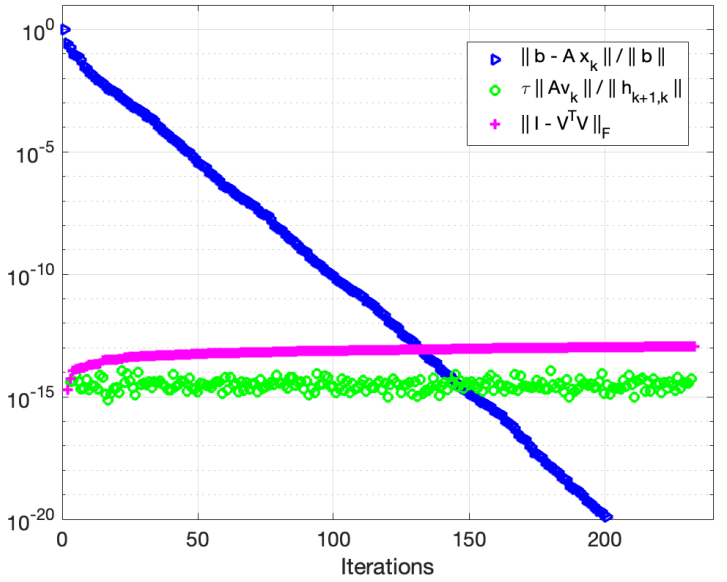

Fig. 11: GMRES residual for Add32 matrix. Tracks HH-GMRES.

REFERENCES

[1] A. Ruhe, Numerical aspects of Gram-Schmidt orthogonalization of vectors, Linear Algebra and its Applications 52 (1983) 591–601.

[2] K. Swirydowicz, J. Langou, S. Ananthan, U. Yang, S. Thomas, Low synchronization Gram-Schmidt and generalized minimal residual algorithms, Numerical Linear Algebra with Applications 28 (2020) 1–20.

[3] Å. Björck, Numerics of Gram-Schmidt orthogonalization, Linear Algebra and Its Applications 197 (1994) 297–316.

[4] N. J. Higham, P. A. Knight, Componentwise error analysis for stationary iterative methods, in: C. D. Meyer, R. J. Plemmons (Eds.), Linear Algebra, Markov Chains, and Queueing Models, IMA Volumes in Mathematics and its Applications, Springer–Verlag, 1993, pp. 29—-46.

[5] C. C. Paige, Z. Strakoš, Residual and backward error bounds in minimum residual Krylov subspace methods, SIAM Journal on Scientific and Statistical Computing 23 (6) (2002) 1899–1924.

[6] C. C. Paige, M. Rozložník, Z. Strakoš, Modified Gram-Schmidt (MGS), least squares, and backward stability of MGS-GMRES, SIAM Journal on Matrix Analysis and Applications 28 (1) (2006) 264–284.

[7] D. Bielich, J. Langou, S. Thomas, K. Świrydowicz, I. Yamazaki, E. Boman, Low-synch Gram-Schmidt with delayed reorthogonalization for Krylov solvers, Parallel Computing (2021).

[8] S. Thomas, A. Carr, P. Mullowney, R. Li, K. Swirydowicz, Neuman series in GMRES and algebraic multigrid, SIAM Journal of Matrix Analysis and Applications (2022).

[9] Y. Saad, M. H. Schultz, GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, SIAM Journal on scientific and statistical computing 7 (3) (1986) 856–869.

[10] H. F. Walker, Implementation of the GMRES method using Householder transformations, SIAM Journal on Scientific and Statistical Computing 9 (1) (1988) 152–163.

[11] L. Giraud, S. Gratton, J. Langou, A rank-k update procedure for reorthogonalizing the orthogonal factor from modified Gram-Schmidt, SIAM J. Matrix Analysis and Applications 25 (4) (2004) 1163–1177.

[12] Å. Björck, C. C. Paige, Loss and recapture of orthogonality in the modified Gram-Schmidt algorithm, SIAM Journal on Matrix Analysis and Applications 13 (1992) 176–190.

[13] S. Lockhart, D. J. Gardner, C. S. Woodward, S. Thomas, L. N. Olson, Performance of low synchronization orthogonalization methods in Anderson accelerated fixed point solvers, in: Proceedings of the 2022 SIAM Conference on Parallel Processing for Scientific Computing (PP), pp. 49–59.

[14] A. Björck, Solving least squares problems by Gram–Schmidt orthogonalization, BIT 7 (1967) 1–21.

[15] N. J. Higham, The accuracy of solutions to triangular systems, SIAM Journal on Numerical Analysis 26 (5) (1989) 1252–1265.

[16] P. Henrici, Bounds for iterates, inverses, spectral variation and fields of values of non-normal matrices, Numerische Mathematik 4 (1) (1962) 24–40.

[17] I. C. Ipsen, A note on the field of values of non-normal matrices, Tech. rep., North Carolina State University. Center for Research in Scientific Computation (1998).

[18] L. Trefethen, M. Embree, The behavior of nonnormal matrices and operators, Spectra and Pseudospectra (2005).

[19] A. Ruhe, On the closeness of eigenvalues and singular values for almost normal matrices, Linear Algebra and its Applications 11 (1) (1975) 87–93.

[20] G. W. Stewart, A Krylov–Schur algorithm for large eigenproblems, SIAM Journal on Matrix Analysis and Applications 23 (3) (2001) 601–614.

[21] Q. Zou, GMRES algorithms over 35 years (2021). arXiv:2110.04017.

[22] I. Yamazaki, S. Thomas, M. Hoemmen, E. G. Boman, K. Świrydowicz, J. J. Elliott, Low-synchronization orthogonalization schemes for s-step and pipelined Krylov solvers in Trilinos, in: Proceedings of the 2020 SIAM Conference on Parallel Processing for Scientific Computing, SIAM, 2020, pp. 118–128. doi:10.1137/1.9781611976137.11.

[23] E. Carson, K. Lund, M. Rozložník, S. Thomas, Block Gram-Schmidt algorithms and their stability properties, Linear Algebra and its Applications 638 (2022) 150–195.

[24] E. Carson, K. Lund, M. Rozložník, The stability of block variants of classical gram–schmidt, SIAM Journal on Matrix Analysis and Applications 42 (3) (2021) 1365–1380.

[25] V. Simoncini, D. Szyld, Theory of inexact Krylov subspace methods and applications to scientific computing, Siam Journal on Scientific Computing (01 2003).

[26] A. Greenbaum, M. Rozložník, Z. Strakoš, Numerical behaviour of the modified Gram-Schmidt GMRES implementation, BIT 37 (3) (1997) 706–719.

[27] P. Liesen, Jörg amd Tichý, The worst-case GMRES for normal matrices, BIT Numerical Mathematics 44 (2004) 79–98.

[28] M. Rozložník, Z. Strakoš, M. Tuma, On the role of orthogonality in the GMRES method, in: K. G. Jeffery, J. Král, M. Bartosek (Eds.), SOFSEM '96: Theory and Practice of Informatics, 23rd Seminar on Current Trends in Theory and Practice of Informatics, Milovy, Czech Republic, November 23-30, 1996, Proceedings, Vol. 1175 of Lecture Notes in Computer Science, Springer, 1996, pp. 409–416.

[29] M. Embree, How descriptive are GMRES convergence bounds?, Tech. Rep. Tech. Rep. 99/08, Mathematical Institute, University of Oxford, UK (1999).

[30] P. Mullowney, R. Li, S. Thomas, S. Ananthan, A. Sharma, A. Williams, J. Rood, M. A. Sprague, Preparing an incompressible-flow fluid dynamics code for exascale-class wind energy simulations, in: Proceedings of the ACM/IEEE Supercomputing 2021 Conference, ACM, 2021, pp. 1–11.

[31] R. D. Falgout, J. E. Jones, U. M. Yang, The design and implementation of hypre, a library of parallel high performance preconditioners, in: A. M. Bruaset, A. Tveito (Eds.), Numerical Solution of Partial Differential Equations on Parallel Computers, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 267–294.