Preconditioning of the reduced system associated with the restricted additive Schwarz method

François Pacull[1] and Damien Tromeur-Dervout[1]

It is of interest to solve large scale sparse linear systems on distributed computers, using Krylov subspace methods along with domain decomposition methods. If accurate subdomain solutions are used, the restricted additive Schwarz preconditioner allows a reduction to the interface via the Schur complement, which leads to an unpreconditioned reduced operator for the interface unknowns. Our purpose is to form a preconditioner for this interface operator by approximating it as a low-rank correction of the identity matrix. To this end, we use a sequence of orthogonal vectors and their image under the interface operator, which are both available after some iterations of the generalized minimal residual method.

The framework of study is purely algebraic and general real sparse non-symmetric and indefinite matrices are considered. The linear system to solve is:

$$Au = f \tag{1}$$

with $A \in \mathbb{R}^{n \times n}$, $u \in \mathbb{R}^n$ and $f \in \mathbb{R}^n$.

Next, we set up the classical notations and terminologies from the algebraic Schwarz literature.

1 Notations

We denote by $\mathcal{V} = \{1, \cdots, n\}$ the set of vertices and by $\mathcal{E}$ the set of edges of the connectivity graph of A: $\mathcal{G} \equiv \mathcal{G}(A) = (\mathcal{V}, \mathcal{E})$. In the present work, we assume that the structure of A is not too far from being symmetric, which is common for matrices issued from partial differential equations. For this reason, edges from $\mathcal{E}$ are made of unordered pairs of vertices from $\mathcal{V}$, and the

[1]Institut Camille Jordan CNRS UMR 5208, Université Lyon 1, Université de Lyon, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne cedex, France. francois.pacull@univ-lyon1.fr, damien.tromeur-dervout@univ-lyon1.fr

graph $\mathcal{G}$ is said to be unoriented: given two vertices i and j from $\mathcal{V}$, the edge (i,j) belongs to $\mathcal{E}$ if and only if $A_{i,j} \neq 0$ or $A_{j,i} \neq 0$.

Given a subset $\mathcal{S} \subset \mathcal{V}$, the induced subgraph $\mathcal{G}|_{\mathcal{S}}$ consists of the vertices $\mathcal{S}$ and the edges $\mathcal{E}|_{\mathcal{S}} = \left\{(i,j) \in \mathcal{E} \ / \ (i,j) \in \mathcal{S}^2\right\} \subset \mathcal{E}$.

Two vertices are said to be adjacent if they share an edge in $\mathcal{E}$. Given a subset $S \subset \mathcal{V}$, the adjacent set $adj(S)$ contains all the vertices that are adjacent to at least one vertex of $\mathcal{S}$, but which do not belong to $\mathcal{S}$.

This allows the definition of overlapping and non-overlapping partitions of $\mathcal{V}$, as used by the algebraic Schwarz preconditioners.

A set $\mathcal{P}_0 = \{\mathcal{V}_{i,0}\}_{1 \leqslant i \leqslant p}$ of subsets $\mathcal{V}_{i,0} \subset \mathcal{V}$ is called a *non-overlapping partition* of $\mathcal{V}$ if:

- no element of $\mathcal{P}_0$ is empty,
- the elements of $\mathcal{P}_0$ are pairwise disjoint,
- the union of the elements of $\mathcal{P}_0$ is equal to $\mathcal{V}$.

A set $\mathcal{P}_\Delta = \{\mathcal{V}_{i,\Delta}\}_{1 \leqslant i \leqslant p}$ of subsets $\mathcal{V}_{i,\Delta} \subset \mathcal{V}$ is called an *overlapping partition* of $\mathcal{V}$ associated with the non-overlapping partition $\mathcal{P}_0$ if, for $1 \leqslant i \leqslant p$:

- $\mathcal{V}_{i,0}$ is a subset of $\mathcal{V}_{i,\Delta}$,
- each vertex from the overlap subset $\mathcal{V}_{i,\Delta} \setminus \mathcal{V}_{i,0}$ is connected to at least one vertex of $\mathcal{V}_{i,0}$ within the subgraph $\mathcal{G}|_{\mathcal{V}_{i,\Delta}}$.

By *connected*, it is meant that there exists a path made of successive adjacent vertices.

The parts or subsets of the partitions are referred to as *subdomains*. The techniques for partitioning a graph and growing subdomains from $\mathcal{P}_0$ to $\mathcal{P}_\Delta$ are beyond the scope of this paper. Let us only say that a common strategy for partitioning a graph is to minimize the number of edges that straddle across the non-overlapping subdomains, while creating p equal size subsets. Also, a basic technique to get an overlapping partition is to add the adjacent vertices of each subdomain: $\mathcal{V}_{i,\Delta} = \mathcal{V}_{i,0} \cup adj\left(\mathcal{V}_{i,0}\right)$. This is what is used for all the computations presented hereafter. Note that this process could be performed recursively in order to further extend the overlap: $\mathcal{V}_{i,\Delta} \leftarrow \mathcal{V}_{i,\Delta} \cup adj\left(\mathcal{V}_{i,\Delta}\right)$.

While all the notations and definitions given above allow the description of the Restricted Additive Schwarz (RAS) preconditioner, another subset of vertices is required in order to study the reduction of the unknowns: $\mathcal{V}_0^E$, the *external interface vertices* (we follow the terminology of [1]) of $\mathcal{P}_0$, which is the union of the p subdomain adjacency sets: $\mathcal{V}_0^E = \cup_{i=1}^p adj\left(\mathcal{V}_{i,0}\right)$.

All the other vertices are referred to as the *irrelevant* vertices (see [2]): $\mathcal{V}_0^I = \mathcal{V} \setminus \mathcal{V}_0^E$. A simple example of a non-overlapping partition, with marked external interface vertices, is shown on **Fig. 1**.

If $n_{i,0} = |\mathcal{V}_{i,0}|$, for $1 \leqslant i \leqslant p$, we denote by $R_{i,0} \in \mathbb{R}^{n_{i,0} \times n}$ the restriction operator from $\mathbb{R}^n$ onto the subspace associated with $\mathcal{V}_{i,0}$. Similarly, if $n_{i,\Delta} =$

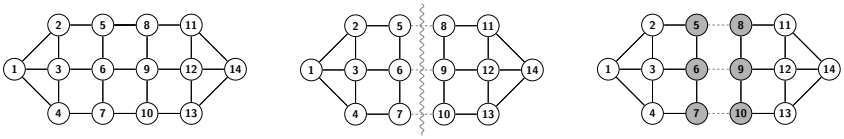

Fig. 1 Example of a connectivity graph (left); non-overlapping partition with 2 subdomains (middle); external interface vertices of the non-overlapping partition (right)

$|\mathcal{V}_{i,\Delta}|$, we denote by $R_{i,\Delta} \in \mathbb{R}^{n_{i,\Delta} \times n}$ the restriction operator from $\mathbb{R}^n$ onto the subspace associated with $\mathcal{V}_{i,\Delta}$. The special restriction operator $\widetilde{R}_{i,\Delta}$ is defined as follows: $\widetilde{R}_{i,\Delta} = R_{i,\Delta} R_{i,0}^T R_{i,0}$.

If $n_0^E = |\mathcal{V}_0^E|$ and $n_0^I \equiv n - n_0^E$, we denote respectively by $R_0^E \in \mathbb{R}^{n_0^E \times n}$ and $R_0^I \in \mathbb{R}^{n_0^I \times n}$ the restriction operators from $\mathbb{R}^n$ onto the subspaces associated with $\mathcal{V}_0^E$ and $\mathcal{V}_0^I$.

Also, the diagonal operator I_0^I stands for ${R_0^I}^T R_0^I$. Note that the transpose of the restriction operators are the corresponding prolongation operators. Finally, the subdomain operators, assumed to be non-singular, are denoted by: $A_{i,\Delta} = R_{i,\Delta} A R_{i,\Delta}^T$.

Despite the fact that the methods related to the reduction to the interface are well-known in the community, we are not aware of a detailed description of this reduction in the right RAS preconditioning case: the next section provides this. We refer to [3] for the left RAS preconditioning case. Anyhow, convergence behaviors are alike when setting the preconditioner either on the left or the right side of A, for Krylov subspace methods, since $M_\Delta^{-1} A$ and $A M_\Delta^{-1}$ are similar operators.

2 Reduction to the interface of the right preconditioned system

The preconditioned operator writes, by definition (see [4]):

$$A M_\Delta^{-1} = A \sum_{i=1}^{p} \widetilde{R}_{i,\Delta}^T A_{i,\Delta}^{-1} R_{i,\Delta}$$

Let us start by stating the following proposition.

Proposition 1. *For any overlapping partitioning $\mathcal{P}_\Delta$ associated with the disjoint subsets $\mathcal{P}_0$, we have:*

$$I_0^I A M_\Delta^{-1} = I_0^I$$

The proof is rather lengthy compared to this paper format. Because it only involves classical algebra, it is left to the reader.

If we come back to the system (1), the right preconditioned version is the following:

$$AM_{\Delta}^{-1}\hat{u} = f, \qquad u = M_{\Delta}^{-1}\hat{u}$$

By introducing a global permutation matrix $P_0^T = [{R_0^I}^T {R_0^E}^T]$, which reorders the unknowns such that those from the external interface of the non-overlapping partition $\mathcal{V}_0^E$ are second, and by permuting this latter system, we get: $P_0 A M_{\Delta}^{-1} P_0^T P_0 \hat{u} = P_0 f$.

We denote respectively by $\hat{x}_0 = R_0^I \hat{u}$ and $\hat{y}_0 = R_0^E \hat{u}$ the irrelevant and external interface unknowns. **Proposition 1** yields:

$$\begin{bmatrix} I & 0 \\ R_0^E A M_{\Delta}^{-1} {R_0^I}^T & R_0^E A M_{\Delta}^{-1} {R_0^E}^T \end{bmatrix} \begin{Bmatrix} \hat{x}_0 \\ \hat{y}_0 \end{Bmatrix} = \begin{Bmatrix} R_0^I f \\ R_0^E f \end{Bmatrix} \tag{2}$$

Let G_{Δ} denote the operator $I - R_0^E A M_{\Delta}^{-1} {R_0^E}^T$. With $\hat{x}_0 = R_0^I f$ and h_{Δ} standing for $R_0^E \left(I - A M_{\Delta}^{-1} I_0^I\right) f$, Eq. (2) can be reduced to the following system:

$$\left(I - G_{\Delta}\right) \hat{y}_0 = h_{\Delta} \tag{3}$$

In order to solve the unpreconditioned system (3) with the Generalized Minimal RESidual (GMRES) method, given an initial guess $\hat{y}_0^{(0)} = \hat{y}_0^{init}$, the evaluation of the initial residual is needed:

$$r_0 = h_{\Delta} - \left(I - G_{\Delta}\right) y_0^{(0)} = R_0^E \left(f - A M_{\Delta}^{-1} \left({R_0^E}^T \hat{y}_0^{init} + I_0^I f \right) \right) \tag{4}$$

Then, in the GMRES outer loop, the following matrix-vector product is required: $w \leftarrow \left(I - G_{\Delta}\right) v_j = R_0^E A M_{\Delta}^{-1} {R_0^E}^T v_j$. This implies solving local problems on each $\mathcal{V}_{i,\Delta}$ subdomain with a right-hand side that is not zero-valued only in $\mathcal{V}_{i,\Delta} \cap \mathcal{V}_0^E$.

Finally, once the iterative method converged to $\hat{y}_0^{(\infty)}$, the solution of the system (1) is recovered as follows:

$$u^{(\infty)} = M_{\Delta}^{-1} \hat{u}^{(\infty)} = M_{\Delta}^{-1} \left({R_0^E}^T \hat{y}_0^{(\infty)} + I_0^I f \right)$$

The algorithm is described on the right of **Fig. 2**. At each iteration, we can monitor the global system's residual norm from the interface residual norm. Using **Proposition 1**, it is easy to check that:

$$\| f - A M_{\Delta}^{-1} \left({R_0^E}^T \hat{y}_0 + I_0^I f \right) \|_2 = \| h_{\Delta} - (I - G_{\Delta}) \hat{y}_0 \|_2$$

If solving the interface system instead of the global one represents only a slight modification of the GMRES algorithm (described on the left of **Fig. 2**), the advantage lies in the size of the system, n_0^E against n, and thus the floating point operation count and memory usage of the GMRES method. The difference between the respective convergence behaviors is not significant,

1: Initial guess: $u_0 \in \mathbb{R}^n$
2: $r_0 \leftarrow f - Au_0$
3: $\beta \leftarrow \|r_0\|_2$, $v_1 \leftarrow r_0/\beta$
4: **for** $j = 1, \cdots, m$ **do**
5: $\quad w \leftarrow AM_\Delta^{-1} v_j$
6: $\quad$ **for** $i = 1, \cdots, j$ **do**
7: $\quad\quad H_{i,j} \leftarrow (w, v_i)$
8: $\quad\quad w \leftarrow w - H_{i,j} v_i$
9: $\quad$ **end for**
10: $\quad H_{j+1,j} \leftarrow \|w\|_2$
11: $\quad v_{j+1} \leftarrow w/H_{j+1,j}$
12: **end for**
13: $V_m = [v_1, \cdots, v_m]$
14: $\overline{H}_m = \{H_{i,j} / 1 \leqslant i \leqslant m+1, 1 \leqslant j \leqslant m\}$
15: $z_m \leftarrow argmin_z \|\beta e_1 - \overline{H}_m z\|_2$
16: $u \leftarrow u_0 + M_\Delta^{-1} V_m z_m$

1: Initial guess: $\hat{y}_0 \in \mathbb{R}^{n_0^E}$
2: $r_0 \leftarrow R_0^E \left(f - AM_\Delta^{-1} \left({R_0^E}^T \hat{y}_0 + I_0^I f \right) \right)$
3: $\beta \leftarrow \|r_0\|_2$, $v_1 \leftarrow r_0/\beta$
4: **for** $j = 1, \cdots, m$ **do**
5: $\quad w \leftarrow R_0^E AM_\Delta^{-1} {R_0^E}^T v_j$
6: $\quad$ **for** $i = 1, \cdots, j$ **do**
7: $\quad\quad H_{i,j} \leftarrow (w, v_i)$
8: $\quad\quad w \leftarrow w - H_{i,j} v_i$
9: $\quad$ **end for**
10: $\quad H_{j+1,j} \leftarrow \|w\|_2$
11: $\quad v_{j+1} \leftarrow w/H_{j+1,j}$
12: **end for**
13: $V_m = [v_1, \cdots, v_m]$
14: $\overline{H}_m = \{H_{i,j} / 1 \leqslant i \leqslant m+1, 1 \leqslant j \leqslant m\}$
15: $z_m \leftarrow argmin_z \|\beta e_1 - \overline{H}_m z\|_2$
16: $u \leftarrow M_\Delta^{-1} \left({R_0^E}^T (\hat{y}_0 + V_m z_m) + I_0^I f \right)$

Fig. 2 GMRES solvers for the global (left) and the interface (right) unknowns

as shown on an example in **Fig. 3**. Indeed, we can see from Eq. (2) that the spectrum of AM_Δ^{-1} is equal to the spectrum of $I - G_\Delta$ augmented with n_0^I one-valued eigenvalues.

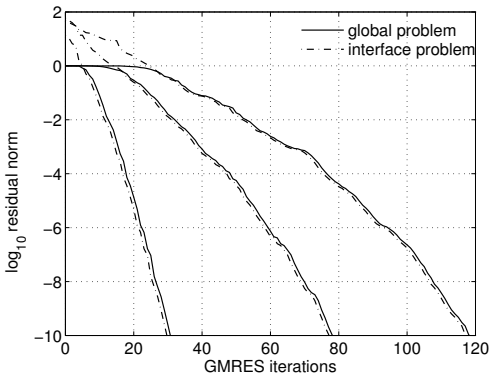

Fig. 3 Full GMRES convergence of the global and interface systems. The GT01R matrix and RHS from the UF sparse matrix collection [5] is used, with a zero initial guess. From left to right, the domain is divided into 2, 4 and 8 parts. The number of primary unknowns is 7980, while the number of interface unknowns is 420, 1260, and 2940 respectively from left to right

The next section is devoted to the preconditioning applied to the reduced system (3).

3 Preconditioning the reduced system

The main difficulty with the preconditioning of $I - G_\Delta$ is that G_Δ is rather dense, as shown with an example in **Table 1**.

Matrix	A	G_Δ		
Number of parts p	N.A.	2	4	8
Size	7980	420	1260	2940
Density (%)	0.68	66.06	51.10	25.72

Table 1 Density percentage of matrices A and G_Δ. The GT01R matrix and RHS from the UF sparse matrix collection [5] is used.

The cost of an approximate inverse approach appears to be prohibitive regarding computational time and memory. Our motivation is to only use the matrix-vector product $(I - G_\Delta)v$ in order to build the preconditioning strategy.

If we have a set V_q of q orthonormal vectors of size n_0^E, we can approach G_Δ using the orthogonal projection matrix $V_q V_q^T$: $\widetilde{G}_\Delta \equiv V_q V_q^T G_\Delta V_q V_q^T$. If we note W_q the image of V_q under G_Δ, and $\widehat{G}_\Delta \equiv V_q^T G_\Delta V_q = V_q^T W_q \in \mathbb{R}^{q \times q}$, we get by Woodbury matrix identity:

$$(I - \widetilde{G}_\Delta)^{-1} = I - V_q (I - \widehat{G}_\Delta^{-1})^{-1} V_q^T = I + V_q ((I - \widehat{G}_\Delta)^{-1} - I) V_q^T \quad (5)$$

$$= I + V_q ((I - V_q^T W_q)^{-1} - I) V_q^T \quad (6)$$

As we can see on **Fig. 2**, we already have an orthonormal basis $V_q = [1, \cdots, v_q]$ after q iterations of the GMRES algorithm. Also, the image of each vector of V_q under $(I - G_\Delta)$ is computed on line 5 of the algorithm on the right side. Thus we need to store theses images $W_q = (I - G_\Delta)V_q$ in order to build the preconditioner (6) and use it subsequently. Some results of this strategy are shown on **Fig. 5**: in this approach, the preconditioner is build first and then kept throughout the GMRES process. On the whole, we observe a trade-off between the number of GMRES iterations used to build the preconditioner q and those saved thanks to the preconditioning. If we note m_{conv} the number of GMRES iterations required to reach a given tolerance, we observe that $q + m_{conv}$ remains constant for all values of q.

Indeed, this interface preconditioner appears to be a cheap and efficient way to avoid stagnation when restarting by keeping some of the most recent convergence information. This is why we tested it on the GMRES(q) technique. At each restart, a new preconditioner is built using the just computed V_q basis. This preconditioner is only used for the subsequent q GMRES iterations. The left-preconditioned GMRES(q) algorithm for the interface unknowns is described on **Fig. 4**.

Some results are shown on **Fig. 6**: this preconditioned restarted GMRES method appears to be robust, while avoiding the growth of memory and orthogonalization time of the full GMRES approach.

Actually, by plugging the equality $I - \widehat{G}_\Delta = V_q^T (I - G_\Delta) V_q$ into Eq. (5), it appears that this preconditioner is related to the preconditioner by deflation from [6], but with a fixed-size approximate invariant subspace that is fully renewed at each restart.

Choose $tol > 0$
Initial guess: $\hat{y}_0 \in \mathbb{R}^{n_0^E}$
$restart \leftarrow 0$
$convergence \leftarrow false$
repeat
  $r_0 \leftarrow R_0^E(f - AM_\Delta^{-1}({R_0^E}^T\hat{y}_0 + I_0^I f))$
  if $restart = 1$ **then**
    $r_0 \leftarrow (I + V_q(Q - I)V_q^T)r_0$
  end if
  $\beta \leftarrow \|r_0\|_2$, $v_1 \leftarrow r_0/\beta$
  for $j = 1, \cdots, q$ **do**
    $w \leftarrow R_0^E A M_\Delta^{-1} {R_0^E}^T v_j$
    $w_j \leftarrow w$
    if $restart = 1$ **then**
      $w \leftarrow (I + V_q(Q - I)V_q^T)w$
    end if
    for $i = 1, \cdots, j$ **do**
      $H_{i,j} \leftarrow (w, v_i)$
      $w \leftarrow w - H_{i,j}v_i$
    end for
    $H_{j+1,j} \leftarrow \|w\|_2$
    $v_{j+1} \leftarrow w/H_{j+1,j}$
  end for
  $V_q = [v_1, \cdots, v_q]$
  $W_q = [w_1, \cdots, w_q]$
  $\overline{H}_q = \{H_{i,j} \ / \ 1 \leqslant i \leqslant q+1, 1 \leqslant j \leqslant q\}$
  $z_q = argmin_z \|\beta e_1 - \overline{H}_q z\|_2$
  $\hat{y}_q \leftarrow \hat{y}_0 + V_q z_q$
  if $\|\beta e_1 - \overline{H}_q z_q\|_2 < tol$ **then**
    $convergence \leftarrow true$
  else
    $restart \leftarrow 1$
    $Q \leftarrow (I - V_q^T W_q)^{-1}$
    $\hat{y}_0 \leftarrow \hat{y}_q$
  end if
until $convergence$
$u \leftarrow M_\Delta^{-1}({R_0^E}^T\hat{y}_q + I_0^I f)$

Fig. 4 GMRES(q) solver for the interface unknowns with a variable left preconditioner

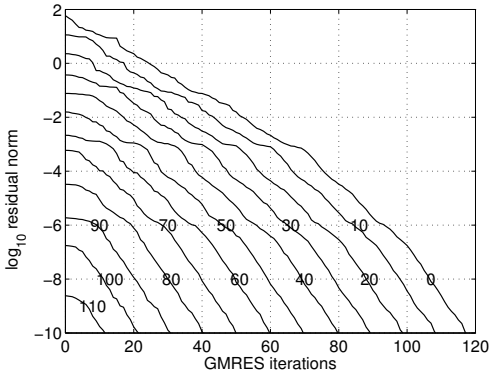

Fig. 5 Full GMRES convergence of the interface system preconditioned on the right side. The GT01R matrix and RHS from the UF sparse matrix collection [5] is used. The initial guess is the outcome of the preconditioner building process. The domain is divided into 8 parts. The value of q, the size of the approximation space for $\widetilde{G}_\Delta$, ranges from 0 (right) to 110 (left).

4 Conclusion

At first, we saw that if accurate subdomain solutions are employed, the right RAS preconditioned system can be reduced to a system of interface equations. The interface unknowns are located at the external interface vertices of the non-overlapping partition. Then, our purpose was to approximate the reduced interface operator as a low-rank correction of the identity matrix, using a sequence of Arnoldi vectors and their image. As might be expected, it is observed that the total cost of the linear solver regarding the number of applications of the Schwarz preconditioner remains approximately constant: what is gained by using an unvarying interface preconditioner is counter-

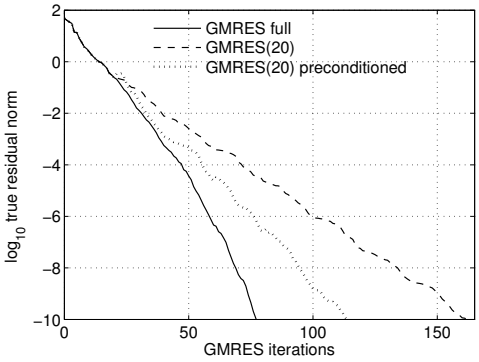

Fig. 6 GMRES convergence of the interface system for three different strategies: full GMRES, GMRES(20) and a left-preconditioned GMRES(20). The GT01R matrix and RHS from the UF sparse matrix collection [5] is used. The domain is divided into 4 parts.

balanced with its building cost. However, this technique becomes beneficial when the restarted variant of the Krylov subspace method is used along with a new interface preconditioner at each restart. A link with the deflation preconditioner was also presented.

Acknowledgments

This work has been supported by the French National Agency of Research, through the ANR-MONU12-0012 H2MNO4 project. The authors wish to express their thanks to Stéphane Aubert for some stimulating conversations and suggestions.

References

[1] Yousef Saad and Maria Sosonkina. Distributed Schur complement techniques for general sparse linear systems. *SIAM J. Sci. Comput.*, 21(4):1337–1356 (electronic), 1999/00.

[2] Erik Brakkee and Peter Wilders. A domain decomposition method for the advection-diffusion equation. Technical report, Delft University of Technology, 1994.

[3] François Pacull and Stéphane Aubert. GMRES acceleration of restricted Schwarz iterations. In *Domain Decomposition Methods in Science and Engineering XXI*, pages 725–732. Springer - Lecture Notes in Computational Science and Engineering, Vol. 98, 2014.

[4] Xiao-Chuan Cai and Marcus Sarkis. A restricted additive Schwarz preconditioner for general sparse linear systems. *SIAM J. Sci. Comput.*, 21(2):792–797 (electronic), 1999.

[5] Timothy A. Davis and Yifan Hu. The University of Florida sparse matrix collection. *ACM Trans. Math. Software*, 38(1):Art. 1, 25, 2011.

[6] Jocelyne Erhel, Kevin Burrage, and Bert Pohl. Restarted GMRES preconditioned by deflation. *J. Comput. Appl. Math.*, 69(2):303–318, 1996.