FLEXIBLE GMRES WITH DEFLATED RESTARTING

L. GIRAUD*, S. GRATTON†, X. PINEL‡, AND X. VASSEUR‡

CERFACS TECHNICAL REPORT TR/PA/09/111

Abstract. In many situations, it has been observed that significant convergence improvements can be achieved in preconditioned Krylov subspace methods by enriching them with some spectral information. On the other hand effective preconditioning strategies are often designed where the preconditioner varies from one step to the next so that a flexible Krylov solver is required. In this paper, we present a new numerical technique for non-symmetric problems that combines these two features. We illustrate the numerical behavior of the new solver both on a set of small academic test examples as well as on large industrial simulation arising in wave propagation simulations.

Key words. Krylov subspace methods, flexible preconditioning, deflation, implicit restarting

AMS subject classifications. 65F10, 65N22, 15A06

1. Introduction. The solution of large linear systems is a basic kernel in many large scale simulations and preconditioned Krylov subspace methods are among the most popular linear solvers. For non-symmetric problems the GMRES [25] method is often chosen because of its robustness [18, 19] and because the Euclidean norm of the system residual is non-increasing along the iterations. In order to make the GMRES method affordable from a memory and floating point operation-count point of view, a restarting or a truncation [26] process has to be implemented. In the classical restarted GMRES approach, the initial guess at restart is chosen to be the best (for the residual norm) known iterate, enforcing thereby the non-increase of the residual norm even when a restart is performed. In such a situation the restart is performed with only one vector. In addition, it has been observed that reusing part of the current Krylov space (and not only one vector) for the construction of iterates in the next cycle of GMRES might significantly improve the convergence. In many approaches, some estimate of the invariant subspace is searched in the Krylov subspace and reused in the next restart either by augmenting the space [3, 14, 23], by deflating over the subspace [16] or by ensuring some orthogonality properties with respect to that space [20]. One of the most recent work in this field based on a deflation approach is GMRES-DR [16]. This method reduces to GMRES, when no deflation is applied, but may provide a much faster convergence than GMRES for well chosen deflation spaces as described in [16].

The methods mentioned above suppose that the preconditioner is a given matrix M that is not allowed to change along the iterations. However, there are situations where this is not true anymore, as e.g. in domain decomposition methods, when approximate solvers are considered for the interior problems (see references in [29, Sect. 4.4] or in [31, Sect. 4.3]). This approach is notably used when the size of the local subproblems is so large that obtaining an approximate solution using an iterative method is computationally more interesting than using a direct method. If the domain decomposition preconditioner is based on the use of approximate solvers, its application is not a linear operation in general, and *flexible methods*, such as the Flexible GMRES method (see the FGMRES method in [22]), are designed to handle this situation.

In this paper, we present a new approach that combines flexible iterations and a restarting strategy that exploits some spectral information. The paper is organized as follows: in Section 2, we briefly present Flexible GMRES [22] and GMRES with deflated restarting [16]. In Section 3 we describe the minimum residual norm subspace method that allows deflated restarting and flexible preconditioning. Section 4 is devoted to numerical experiments where both academic and real life problems are considered to illustrate the numerical features of the new solver.

2. Problem setting. For the sake of generality we describe in this paper approaches for complex-valued linear systems, although everything also specializes to real arithmetic calculation. Let $A \in \mathbb{C}^{n\times n}$ be a square nonsingular $n \times n$ non-symmetric complex matrix, and $b \in \mathbb{C}^n$ be a

*INRIA Bordeaux Sud-Ouest, 351 cours de la Libération, F-33400 Talence, France and CERFACS, 42, Avenue Gaspard Coriolis, F-31057 Toulouse Cedex 1, France

†INPT-IRIT, University of Toulouse and ENSEEIHT, 2 rue Camichel, BP 7122, F-31071 Toulouse cedex 7, France

‡CERFACS, 42, Avenue Gaspard Coriolis, F-31057 Toulouse Cedex 1, France

vector, that both define the linear system

$$Ax = b. \tag{2.1}$$

Among the possible subspace methods to solve the linear system (2.1) we consider the variants that search for an approximation associated with a minimum residual norm property such as in GMRES. In this section we briefly present two existing minimum residual norm subspace methods that allow either variable preconditioning or deflated restarting respectively. Both methods will play an important role in the novel subspace method that allows flexible preconditioning and deflated restarting simultaneously. First we introduce some notations and definitions used through the paper.

2.1. Notations and definitions.

2.1.1. Notations. We denote $\|.\|$ the Euclidean norm, $I_k \in \mathbb{C}^{k\times k}$ the identity matrix of dimension k and $0_{i\times j} \in \mathbb{C}^{i\times j}$ the zero rectangular matrix with i rows and j columns. ^{T} denotes the transpose operation, whereas ^{H} represents the Hermitian transpose operation. Given a vector $d \in \mathbb{C}^k$ with components d_i, $D = \mathrm{diag}(d_1, \cdots, d_k)$ is the diagonal matrix $D \in \mathbb{C}^{k\times k}$ such that $D_{ii} = d_i$. Given a matrix Q we denote q_j its $j-th$ column. Finally $e_m \in \mathbb{C}^m$ denotes the m-th Cartesian basis vector of $\mathbb{C}^m$. Regarding the algorithmic part (Algorithms 1-6), we adopt Matlab-like notations in the presentation. For instance $Q(i,j)$ denotes the q_{ij} entry of matrix Q and $Q(1:m, 1:j)$ refers to the submatrix made of the first m rows and first j columns of Q.

2.1.2. Definitions. Definition 2.1 introduces the fundamental relation that will characterize the class of methods that allow flexible preconditioning and/or deflated restarting.

DEFINITION 2.1. *Flexible Arnoldi relation. The subspace methods investigated in this paper will satisfy the following relation named later on flexible Arnoldi relation:*

$$AZ_\ell = V_{\ell+1}\bar{H}_\ell$$

where $Z_\ell \in \mathbb{C}^{n\times\ell}$, $V_{\ell+1} \in \mathbb{C}^{n\times(\ell+1)}$ such that $V_{\ell+1}^H V_{\ell+1} = I_{\ell+1}$ and $\bar{H}_\ell \in \mathbb{C}^{(\ell+1)\times\ell}$. These methods will compute an approximation of the solution in a ℓ-dimensional affine space $x_0 + Z_\ell\, y_\ell$ where $y_\ell \in \mathbb{C}^\ell$. In certain cases, $\bar{H}_\ell$ will be an upper Hessenberg matrix.

Finally we recall the definition of a harmonic Ritz pair [17, 28] since this notion plays an important role when considering deflated restarting.

DEFINITION 2.2. *Harmonic Ritz pair. Consider a subspace $\mathcal{U}$ of $\mathbb{C}^n$. Given a matrix $B \in \mathbb{C}^{n\times n}$, $\lambda \in \mathbb{C}$ and $y \in \mathcal{U}$, (λ, y) is a harmonic Ritz pair of B with respect to $\mathcal{U}$ if and only if*

$$By - \lambda\, y \perp B\,\mathcal{U}$$

or equivalently, for the canonical scalar product,

$$\forall w \in \mathrm{range}(B\,\mathcal{U}) \quad w^H\,(By - \lambda\, y) = 0.$$

We call y a harmonic Ritz vector associated with the harmonic Ritz value λ.

2.2. Flexible GMRES. In many large-scale scientific and industrial applications one might not be able to consider a fixed preconditioner at each step of the subspace method. This happens for instance when block preconditioners, including domain decomposition techniques, are considered where the blocks are too large to be handled by a direct solver. In such a situation, an iterative solver has to be implemented to solve linear systems involving these blocks. Consequently the preconditioner varies from one step to the next and *flexible* Krylov solvers have been developed to manage this issue [22, 33].

In [22] Saad has proposed a minimum residual norm subspace method based on the GMRES approach that allows variable preconditioning. We denote M_j the nonsingular matrix that represents the preconditioner at step j of the method. Algorithm 1 depicts the FGMRES(m) method where the approximation subspace is not allowed to be larger than a prescribed dimension noted

m in the rest of the paper. Starting from an initial guess x_0, it is based on the flexible Arnoldi relation with $Z_m \in \mathbb{C}^{n\times m}$, $V_{m+1} \in \mathbb{C}^{n\times(m+1)}$ and the upper Hessenberg matrix $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$

$$AZ_m = V_{m+1}\bar{H}_m \quad \text{with} \quad V_{m+1}^H V_{m+1} = I_{m+1}. \tag{2.2}$$

An approximate solution $x_m \in \mathbb{C}^n$ is then found by minimizing the residual norm $\|b-A(x_0+Z_m y)\|$ over the space $x_0 + \text{range}(Z_m)$, the corresponding residual being $r_m = b - Ax_m \in \mathbb{C}^n$ with $r_m \in \text{range}(V_{m+1})$. An optimality property similar to the one that defines GMRES is thus obtained. We note however that no general convergence results are available since the subspace of approximants range(Z_m) is no longer a standard Krylov subspace. We refer the reader to [22, 24] for the analysis of the breakdown in FGMRES. Finally, we note that the additional storage cost of FGMRES(m) over GMRES(m) only concerns the extra memory required to store Z_m; i.e., m additional vectors of length n.

Algorithm 1 Flexible GMRES(m)

1: *Initialization:* Choose $m > 0$, $tol > 0$, $x_0 \in \mathbb{C}^n$. Let $r_0 = b - Ax_0$, $\beta = \|r_0\|$, $c = [\beta, 0_{1\times m}]^T$ where $c \in \mathbb{C}^{m+1}$, $v_1 = r_0/\beta$.

Loop

2: *Computation of V_{m+1}, Z_m and $\bar{H}_m$ (see Algorithm 2):* Apply m steps of the Arnoldi method with flexible preconditioning ($z_j = M_j^{-1}v_j, 1 \le j \le m$) to obtain $V_{m+1} \in \mathbb{C}^{n\times(m+1)}$, $Z_m \in \mathbb{C}^{n\times m}$ and the upper Hessenberg matrix $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$ such that:

$$AZ_m = V_{m+1}\bar{H}_m \quad \text{with} \quad V_{m+1}^H V_{m+1} = I_{m+1}.$$

3: *Minimum norm solution:* Compute the minimum norm solution $x_m \in \mathbb{C}^n$ in the affine space $x_0 + \text{range}(Z_m)$; that is, $x_m = x_0 + Z_m y^*$ where $y^* = \underset{y\in\mathbb{C}^m}{\text{argmin}} \|c - \bar{H}_m y\|$.

4: *Check the convergence criterion:* If $\|c - \bar{H}_m y^*\|/\|b\| \le tol$, exit

5: *Settings:* Set $x_0 = x_m$, $r_0 = b - Ax_0$, $\beta = \|r_0\|$, $c = [\beta, 0_{1\times m}]^T$, $v_1 = r_0/\beta$.

End of loop

Algorithm 2 Flexible GMRES(m): computation of V_{m+1}, Z_m and $\bar{H}_m$

1: **for** $j = 1, m$ **do**
2: $z_j = M_j^{-1}v_j$
3: $w = Az_j$
4: **for** $i = 1, j$ **do**
5: $h_{i,j} = w^H v_i$
6: $w = w - h_{i,j}v_i$
7: **end for**
8: $h_{i+1,j} = \|w\|$, $v_{j+1} = w/h_{i+1,j}$
9: **end for**
10: Define $Z_m = [z_1, \cdots, z_m]$, $V_{m+1} = [v_1, \cdots, v_{m+1}]$, $\bar{H}_m = \{h_{i,j}\}_{1\le i\le m+1, 1\le j\le m}$

2.3. GMRES with deflated restarting. Krylov subspace methods with standard restarting implement a scheme where the maximal dimension of the approximation subspace is fixed (m here). The method is restarted in order to control both the memory requirements and the computational cost of the orthogonalization scheme of the method. In the case of GMRES(m) it means in practice that the orthonormal basis built is thrown away. Since some information is discarded at the restart, the convergence is expected to be slower compared to full GMRES.

Nevertheless more sophisticated procedures have been proposed to enhance convergence properties of restarted Krylov subspace methods. Basically these methods fall in the category of augmented or deflated methods and we refer the reader to [27, Sections 8 and 9] for a review and detailed references. In this paper we focus on GMRES with deflated restarting, one of those methods, referred to as GMRES-DR [16]. This method aims at using spectral information at a restart

mainly to improve the convergence of restarted GMRES. We consider a fixed right preconditioning matrix noted M and suppose that an Arnoldi relation of type $AM^{-1}V_m = V_{m+1}\bar{H}_m$ holds. We note that $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$ has the following form

$$\bar{H}_m = \begin{bmatrix} H_m \\ h_{m+1,m} e_m^T \end{bmatrix}$$

where $H_m \in \mathbb{C}^{m\times m}$ is supposed to be nonsingular. A subspace of dimension k (with $k < m$) spanned by harmonic Ritz vectors (and not only the approximate solution with minimum residual norm) is retained in the restarting scheme. We denote $Y_k = V_m G_k$ these harmonic Ritz vectors where $Y_k = [y_1, \ldots, y_k] \in \mathbb{C}^{n\times k}$ and $G_k = [g_1, \ldots, g_k] \in \mathbb{C}^{m\times k}$. As originally proposed by Morgan [16], the vectors g_j and associated Ritz values λ_j (with $1 \leq j \leq k$) are obtained as solutions of the following eigenvalue problem

$$(H_m + |h_{m+1,m}|^2 H_m^{-H} e_m e_m^T) g_j - \lambda_j g_j = 0. \tag{2.3}$$

Next, the QR factorization of the following $(m+1)\times(k+1)$ matrix

$$\begin{bmatrix} \begin{bmatrix} G_k \\ 0_{1\times k} \end{bmatrix} & V_{m+1}^H \; r_0 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} G_k \\ 0_{1\times k} \end{bmatrix} & c - \bar{H}_m y^* \end{bmatrix} \quad \text{with} \quad r_0 = V_{m+1}(c - \bar{H}_m y^*)$$

is performed where $c \in \mathbb{C}^{m+1}$ and $y^* \in \mathbb{C}^m$. This allows to compute new matrices $V_{k+1}^{new} \in \mathbb{C}^{n\times(k+1)}$ and $\bar{H}_k^{new} \in \mathbb{C}^{(k+1)\times k}$ such that

$$\begin{aligned} AM^{-1}V_k^{new} &= V_{k+1}^{new} \; \bar{H}_k^{new}, \\ V_{k+1}^{new\,H} \; V_{k+1}^{new} &= I_{k+1}, \\ \text{range}([Y_k, r_0]) &= \text{range}(V_{k+1}^{new}) \end{aligned}$$

where $\bar{H}_k^{new}$ is a $(k+1)\times k$ rectangular matrix. GMRES-DR then carries out $m-k$ Arnoldi steps with fixed preconditioning and starting vector v_{k+1}^{new} to eventually build V_{m+1} and $\bar{H}_m$. At the end of the GMRES cycle with deflated restarting we have a final relation similar to the Arnoldi relation (2.2) with $V_{m+1} \in \mathbb{C}^{n\times(m+1)}$ and $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$

$$AM^{-1}V_m = V_{m+1}\bar{H}_m \quad \text{with} \quad V_{m+1}^H \; V_{m+1} = I_{m+1}$$

where $\bar{H}_m$ is *no longer* upper Hessenberg after the first cycle. An approximate solution $x_m \in \mathbb{C}^n$ is then found by minimizing the residual norm $\|b - A(x_0 + M^{-1}V_m y)\|$ over the space $x_0 + M^{-1}\text{range}(V_m)$, the corresponding residual being $r_m = b - Ax_m \in \mathbb{C}^n$ with $r_m \in \text{range}(V_{m+1})$. An optimality property is thus also obtained.

We refer the reader to [16, 21] for further comments on the algorithm and computational details. This approach has been proved efficient on many academic examples [16]. We note that GMRES with deflated restarting is equivalent to GMRES with eigenvectors [14] and to implicitly restarted GMRES [15]. Details of the method are given in Algorithms 3 and 4 respectively. GMRES-DR(m,k) does require only $m-k$ matrix vector products and preconditioning operations per cycle while GMRES(m) needs m. Finally we note that Krylov subspace methods with deflated restarting have been exclusively developed in the case of a fixed preconditioner. In Section 3 we extend the GMRES-DR method to the case of variable preconditioning.

Algorithm 3 Right-preconditioned GMRES with deflated restarting: GMRES-DR(m,k)

1: *Initialization:* Choose $m > 0$, $k > 0$, $tol > 0$, $x_0 \in \mathbb{C}^n$. Let $r_0 = b - Ax_0$; $\beta = \|r_0\|$, $c = [\beta, 0_{1\times m}]^T \in \mathbb{C}^{m+1}$, $v_1 = r_0/\beta$.

2: *Computation of V_{m+1} and $\bar{H}_m$:* Apply m steps of the Arnoldi procedure with right preconditioning to obtain $V_{m+1} \in \mathbb{C}^{n\times(m+1)}$ and the upper Hessenberg matrix $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$ such that:

$$AM^{-1}V_m = V_{m+1}\bar{H}_m \quad \text{with} \quad V_{m+1}^H V_{m+1} = I_{m+1}.$$

Loop

3: *Minimum norm solution:* Compute the minimum norm solution $x_m \in \mathbb{C}^n$ in the affine space $x_0 + M^{-1}\text{range}(V_m)$; that is, $x_m = x_0 + M^{-1}V_m y^*$ where $y^* = \underset{y\in\mathbb{C}^m}{\text{argmin}} \|c - \bar{H}_m y\|$. Set $x_0 = x_m$ and $r_0 = b - Ax_0$.

4: *Check the convergence criterion:* If $\|c - \bar{H}_m y^*\|/\|b\| \leq tol$, exit

5: *Computation of V_{k+1}^{new} and $\bar{H}_k^{new}$:* see Algorithm 4. At the end of this step the following relations hold:

$$AM^{-1}V_k^{new} = V_{k+1}^{new}\bar{H}_k^{new} \text{ with } V_{k+1}^{new\,H}V_{k+1}^{new} = I_{k+1} \text{ and } r_0 \in \text{range}(V_{k+1}^{new}).$$

6: *Arnoldi procedure:* Set $V_{k+1} = V_{k+1}^{new}$, $\bar{H}_k = \bar{H}_k^{new}$ and apply $(m-k)$ steps of the Arnoldi procedure with right preconditioning and starting vector v_{k+1} to build $V_{m+1} \in \mathbb{C}^{n\times(m+1)}$ and $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$ such that:

$$AM^{-1}V_m = V_{m+1}\bar{H}_m \quad \text{with} \quad V_{m+1}^H V_{m+1} = I_{m+1}.$$

7: *Setting:* Set $c = V_{m+1}^H r_0$.

End of loop

Algorithm 4 GMRES-DR(m,k): computation of V_{k+1}^{new} and $\bar{H}_k^{new}$

1: *Input:* A, V_{m+1} such that $AM^{-1}V_m = V_{m+1}\bar{H}_m$ and $c - \bar{H}_m y^*$ such that $r_0 = V_{m+1}(c - \bar{H}_m y^*)$.

2: *Settings:* Define $h_{m+1,m} = \bar{H}_m(m+1, m)$, $H_m \in \mathbb{C}^{m\times m}$ as $H_m = \bar{H}_m(1:m, 1:m)$.

3: *Compute k harmonic Ritz vectors:* Compute k independent eigenvectors g_i of the matrix $H_m + |h_{m+1,m}|^2 H_m^{-H} e_m e_m^T$. Set $G_k = [g_1, \ldots, g_k] \in \mathbb{C}^{m\times k}$.

4: *Augmentation of G_k:* Define $G_{k+1} \in \mathbb{C}^{(m+1)\times(k+1)}$ as

$$G_{k+1} = \left[\begin{bmatrix} G_k \\ 0_{1\times k} \end{bmatrix}, c - \bar{H}_m y^*\right].$$

5: *Orthonormalization of the columns of G_{k+1}:* Perform a QR-factorization of G_{k+1} as $G_{k+1} = P_{k+1}\Gamma_{k+1}$. Define $P_k \in \mathbb{C}^{m\times k}$ as $P_k = P_{k+1}(1:m, 1:k)$.

6: *Settings and final relation:* Set $V_{k+1}^{new} = V_{m+1}P_{k+1}$ and $\bar{H}_k^{new} = P_{k+1}^H \bar{H}_m P_k$. At the end of this step the following relations are satisfied:

$$AM^{-1}V_m P_k = V_{m+1}P_{k+1}P_{k+1}^H \bar{H}_m P_k \quad ;\text{i.e.,} \quad AM^{-1}V_k^{new} = V_{k+1}^{new}\ \bar{H}_k^{new}$$

where $\bar{H}_k^{new}$ is generally a dense matrix.

3. Flexible GMRES with deflated restarting. In this section we present the new subspace method that allows deflated restarting and variable preconditioning simultaneously. We suppose that a flexible Arnoldi relation holds ($AZ_m = V_{m+1}\bar{H}_m$) and analyze one cycle of this method.

3.1. Analysis of a cycle. We discuss now the two main points related to the extension of GMRES-DR in a flexible setting: what is the harmonic Ritz information recovered at restart and is it still possible as in GMRES-DR to restart at low computational cost the flexible Arnoldi relation?

Both questions will be answered in this section.

3.1.1. Harmonic Ritz formulation. Proposition 1 presents the harmonic Ritz formulation used in the flexible variant of GMRES with deflated restarting.

PROPOSITION 1. *Flexible GMRES with deflated restarting relies on the computation of k harmonic Ritz vectors $Y_k = V_m G_k$ with $Y_k \in \mathbb{C}^{n\times k}$ and $G_k \in \mathbb{C}^{m\times k}$, where each harmonic Ritz pair $(\lambda_j, V_m g_j)$ satisfies*

$$(H_m + |h_{m+1,m}|^2 H_m^{-H} e_m e_m^T) g_j - \lambda_j g_j = 0.$$

Y_k correspond to harmonic Ritz vectors of $AZ_m V_m^H$ with respect to range(V_m).

Proof. The proposed eigenvalue problem - which is the same as in GMRES with deflated restarting (see relation (2.3)) - can be also written in a compact form as

$$\bar{H}_m^H \bar{H}_m g_j - \lambda_j \bar{H}_m^H \begin{pmatrix} g_j \\ 0 \end{pmatrix} = 0. \tag{3.1}$$

Since V_{m+1} has orthonormal columns the following relations hold for the two terms of Equation (3.1)

$$\begin{aligned} \bar{H}_m^H \bar{H}_m &= (AZ_m)^H (AZ_m), \\ \bar{H}_m^H \begin{pmatrix} g_j \\ 0 \end{pmatrix} &= \bar{H}_m^H V_{m+1}^H V_{m+1} \begin{pmatrix} g_j \\ 0 \end{pmatrix} = (AZ_m)^H V_m g_j. \end{aligned}$$

Consequently the eigenvalue problem (3.1) becomes

$$\forall w \in \text{range}(AZ_m) \quad w^H (AZ_m g_j - \lambda_j V_m g_j) = 0, \tag{3.2}$$

or equivalently since $V_m^H V_m = I_m$

$$\forall w \in \text{range}(AZ_m V_m^H V_m) \quad w^H (AZ_m V_m^H V_m g_j - \lambda_j V_m g_j) = 0.$$

Thus following Definition 2.2, Y_k correspond to harmonic Ritz vectors of $AZ_m V_m^H$ with respect to range(V_m). When a fixed preconditioning is used, it is straightforward to deduce that GMRES-DR relies on harmonic Ritz vectors of $AM^{-1}V_m^H$ with respect to range(V_m). Due to relation (3.2) we also note that the harmonic residual vectors $AZ_m V_m^H V_m g_j - \lambda_j V_m g_j \in$ range(V_{m+1}) are orthogonal to a subspace of dimension m spanned by the columns of AZ_m.

□

In Lemma 3.1 we detail a useful relation satisfied by the harmonic Ritz vectors.

LEMMA 3.1.

In Flexible GMRES with deflated restarting, the harmonic Ritz vectors are given by $Y_k = V_m G_k$ with corresponding harmonic Ritz values λ_k. $G_k \in \mathbb{C}^{m\times k}$ satisfies the following relation:

$$AZ_m G_k = V_{m+1} \left[\begin{bmatrix} G_k \\ 0_{1\times k} \end{bmatrix}, \rho_m \right] \begin{bmatrix} \text{diag}(\lambda_1, \cdots, \lambda_k) \\ \alpha_{1\times k} \end{bmatrix} \tag{3.3}$$

where $\rho_m \in \mathbb{C}^{m+1}$ is such that $r_0 = V_{m+1}\rho_m = V_{m+1}(c - \bar{H}_m y^*)$ and $\alpha_{1\times k} = [\alpha_1, \cdots, \alpha_k] \in \mathbb{C}^{1\times k}$.

Proof. The harmonic residual vectors $AZ_m V_m^H V_m g_i - \lambda_i V_m g_i$ and the residual vector r_0 all reside in a subspace of dimension $m+1$ (spanned by the columns of V_{m+1}) and are orthogonal to the same subspace of dimension m (spanned by the columns of AZ_m subspace of range(V_{m+1})), so they must be collinear. Consequently there exist k coefficients noted $\alpha_i \in \mathbb{C}$ with $1 \le i \le k$ such that

$$\forall i \in \{1, \cdots, k\} \; AZ_m g_i - \lambda_i V_m g_i = \alpha_i r_0 = \alpha_i V_{m+1} \rho_m. \tag{3.4}$$

Setting $\alpha_{1\times k} = [\alpha_1, \cdots, \alpha_k] \in \mathbb{C}^{1\times k}$, the collinearity expression (3.4) can be written in matrix form

$$AZ_m G_k = V_{m+1} \left[\begin{bmatrix} G_k \\ 0_{1\times k} \end{bmatrix}, \rho_m \right] \begin{bmatrix} \text{diag}(\lambda_1, \cdots, \lambda_k) \\ \alpha_{1\times k} \end{bmatrix}.$$

□

3.1.2. Flexible Arnoldi relation. Let us further denote by $G_k = P_k\Gamma_k$ the QR-factorization of G_k, where $P_k \in \mathbb{C}^{m\times k}$ has orthonormal columns and $\Gamma_k \in \mathbb{C}^{k\times k}$ is a nonsingular upper triangular matrix. We denote $G_{k+1} \in \mathbb{C}^{(m+1)\times(k+1)}$ the following matrix that appears in Lemma 3.1:

$$G_{k+1} = \left[\begin{bmatrix} G_k \\ 0_{1\times k}\end{bmatrix}, \rho_m\right]. \tag{3.5}$$

Proposition 2 shows that a flexible Arnoldi relation can be recovered at low computational cost when restarting with some harmonic information; i.e., without involving any matrix-vector product with A as in [5].

PROPOSITION 2.

At each restart of Flexible GMRES with deflated restarting, the flexible Arnoldi relation

$$AZ_k^{new} = V_{k+1}^{new}\bar{H}_k^{new}$$

holds with

$$Z_k^{new} = Z_m P_k,$$

$$V_{k+1}^{new} = V_{m+1}P_{k+1},$$

and

$$\bar{H}_k^{new} = P_{k+1}^H \bar{H}_m P_k.$$

Proof.

After orthogonalization of the vector ρ_m against the columns of $\begin{bmatrix} P_k \\ 0_{1\times k}\end{bmatrix}$ we obtain the unit norm vector $p_{k+1} \in \mathbb{C}^{m+1}$ that satisfies

$$p_{k+1} = \bar{p}_{k+1}/\|\bar{p}_{k+1}\| \quad \text{with} \quad \bar{p}_{k+1} = \rho_m - \begin{bmatrix} P_k \\ 0_{1\times k}\end{bmatrix}\begin{bmatrix} P_k \\ 0_{1\times k}\end{bmatrix}^H \rho_m.$$

We note $a = \|\bar{p}_{k+1}\|$ and $u_{k\times 1} \in \mathbb{C}^k$ the following quantity $u_{k\times 1} = \begin{bmatrix} P_k \\ 0_{1\times k}\end{bmatrix}^H \rho_m$ respectively. Thus

$$\rho_m = \left[\begin{bmatrix} P_k \\ 0_{1\times k}\end{bmatrix},\ p_{k+1}\right]\begin{bmatrix} u_{k\times 1} \\ a\end{bmatrix}.$$

Consequently the QR factorization of $G_{k+1} = P_{k+1}\Gamma_{k+1}$ can be written as

$$\left[\begin{bmatrix} G_k \\ 0_{1\times k}\end{bmatrix},\ \rho_m\right] = \left[\begin{bmatrix} P_k \\ 0_{1\times k}\end{bmatrix},\ p_{k+1}\right]\begin{bmatrix} \Gamma_k & u_{k\times 1} \\ 0_{1\times k} & a\end{bmatrix}.$$

From relation (3.3) of Lemma 3.1 we deduce

$$AZ_m P_k = V_{m+1}P_{k+1}\Gamma_{k+1}\begin{bmatrix} \operatorname{diag}(\lambda_1,\cdots,\lambda_k) \\ \alpha_{1\times k}\end{bmatrix}\Gamma_k^{-1}. \tag{3.6}$$

Using the flexible Arnoldi relation $AZ_m = V_{m+1}\bar{H}_m$ and $P_{k+1}^H P_{k+1} = I_{k+1}$ we obtain

$$P_{k+1}^H \bar{H}_m P_k = \Gamma_{k+1}\begin{bmatrix} \operatorname{diag}(\lambda_1,\cdots,\lambda_k) \\ \alpha_{1\times k}\end{bmatrix}\Gamma_k^{-1}.$$

If we denote $Z_k^{new} = Z_m P_k$, $V_{k+1}^{new} = V_{m+1}P_{k+1}$ and

$$\bar{H}_k^{new} = \Gamma_{k+1}\begin{bmatrix} \operatorname{diag}(\lambda_1,\cdots,\lambda_k) \\ \alpha_{1\times k}\end{bmatrix}\Gamma_k^{-1} = P_{k+1}^H\bar{H}_m P_k,$$

Equation (3.6) can be written in the following flexible Arnoldi relation

$$AZ_k^{new} = V_{k+1}^{new} \bar{H}_k^{new}.$$

□

Next, setting $Z_k = Z_k^{new}$, $V_{k+1} = V_{k+1}^{new}$ and $\bar{H}_k = \bar{H}_k^{new}$ respectively flexible GMRES with deflated restarting then carries out $(m-k)$ flexible Arnoldi steps with flexible preconditioning and starting vector v_{k+1} leading to

$$A\ Z_m = V_{m+1}\ \bar{H}_m,$$

where $Z_m \in \mathbb{C}^{n\times m}$, $V_{m+1} \in \mathbb{C}^{n\times(m+1)}$ and $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$.

3.2. Algorithm and computational aspects. Details of flexible GMRES with deflated restarting are depicted in Algorithms 5 and 6 respectively. We will call this algorithm FGMRES-DR(m,k) and compare this method with both FGMRES(m) and GMRES-DR(m,k) from a computational and storage point of view.

Algorithm 5 Flexible GMRES with deflated restarting: FGMRES-DR(m,k)

1: *Initialization:* Choose $m > 0$, $k > 0$, $tol > 0$, $x_0 \in \mathbb{C}^n$. Let $r_0 = b - Ax_0$; $\beta = \|r_0\|$, $c = [\beta, 0_{1\times m}]^T \in \mathbb{C}^{m+1}$, $v_1 = r_0/\beta$.

2: *Computation of V_{m+1}, Z_m and $\bar{H}_m$:* Apply m steps of the Arnoldi procedure with *flexible* preconditioning to obtain $V_{m+1} \in \mathbb{C}^{n\times(m+1)}$, $Z_m \in \mathbb{C}^{n\times m}$ and the upper Hessenberg matrix $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$ such that:

$$AZ_m = V_{m+1}\bar{H}_m \quad \text{with} \quad V_{m+1}^H V_{m+1} = I_{m+1}.$$

Loop

3: *Minimum norm solution:* Compute the minimum norm solution $x_m \in \mathbb{C}^n$ in the affine space $x_0 + \text{range}(Z_m)$; that is, $x_m = x_0 + Z_m y^*$ where $y^* = \underset{y\in\mathbb{C}^m}{\text{argmin}} \|c - \bar{H}_m y\|$. Set $x_0 = x_m$ and $r_0 = b - Ax_0$.

4: *Check the convergence criterion:* If $\|c - \bar{H}_m y^*\|/\|b\| \leq tol$, exit

5: *Computation of V_{k+1}^{new}, Z_k^{new} and $\bar{H}_k^{new}$:* see Algorithm 6. At the end of this step the following relations hold:

$$AZ_k^{new} = V_{k+1}^{new}\bar{H}_k^{new} \text{ with } V_{k+1}^{new\,H} V_{k+1}^{new} = I_{k+1} \text{ and } r_0 \in \text{range}(V_{k+1}^{new}). \tag{3.7}$$

6: *Arnoldi procedure:* Set $V_{k+1} = V_{k+1}^{new}$, $Z_k = Z_k^{new}$, $\bar{H}_k = \bar{H}_k^{new}$ and apply $(m-k)$ steps of the Arnoldi procedure with *flexible* preconditioning and starting vector v_{k+1} to build $V_{m+1} \in \mathbb{C}^{n\times(m+1)}$, $Z_m \in \mathbb{C}^{n\times m}$ and $\bar{H}_m \in \mathbb{C}^{(m+1)\times m}$ such that:

$$AZ_m = V_{m+1}\bar{H}_m \quad \text{with} \quad V_{m+1}^H V_{m+1} = I_{m+1}.$$

7: *Setting:* Set $c = V_{m+1}^H r_0$.

End of loop

3.2.1. Computational cost. We summarize now in Table 3.1 the main computational costs associated with each generic cycle of FGMRES(m), GMRES-DR(m,k) and FGMRES-DR(m,k). We have only included the costs proportional to the size of the original problem n which is supposed to be much larger than m and k. We denote op_A and op_M the floating point operation counts for the matrix-vector product and the preconditioner application respectively. The main computational differences are in the calculation of V_{k+1} and Z_k when comparing FGMRES and FGMRES-DR. In FGMRES-DR those vectors are computed using dense matrix-matrix operations efficiently implemented in BLAS-3 libraries, while in FGMRES-DR they are obtained through a sequence of matrix-vector products, possibly sparse, depending on the nature of A and the preconditioners.

For deflating variants, the reduction of this total cost is still possible. The right-hand side c of the least-squares problem is computed as $c = V_{m+1}^H r_0$ which involves $2n(m+1)$ operations as

Algorithm 6 FGMRES-DR(m,k): computation of V_{k+1}^{new}, Z_k^{new} and $\bar{H}_k^{new}$

1: *Input:* A, Z_m, V_{m+1} such that $AZ_m = V_{m+1}\bar{H}_m$ and $c - \bar{H}_m y^*$ such that $r_0 = V_{m+1}(c - \bar{H}_m y^*)$.
2: *Settings:* Define $h_{m+1,m} = \bar{H}_m(m+1,m)$, $H_m \in \mathbb{C}^{m\times m}$ as $H_m = \bar{H}_m(1:m, 1:m)$.
3: *Compute k harmonic Ritz vectors.* Compute k independent eigenvectors g_i of the matrix $H_m + |h_{m+1,m}|^2 H_m^{-H} e_m e_m^T$. Set $G_k = [g_1, \ldots, g_k] \in \mathbb{C}^{m\times k}$.
4: *Augmentation of G_k:* Define $G_{k+1} \in \mathbb{C}^{(m+1)\times(k+1)}$ as

$$G_{k+1} = \left[\begin{bmatrix} G_k \\ 0_{1\times k} \end{bmatrix}, c - \bar{H}_m y^* \right]. \tag{3.8}$$

5: *Orthonormalization of the columns of G_{k+1}:* Perform a QR-factorization of G_{k+1} as $G_{k+1} = P_{k+1}\Gamma_{k+1}$. Define $P_k \in \mathbb{C}^{m\times k}$ as $P_k = P_{k+1}(1:m, 1:k)$.
6: *Settings and final relation:* Set $V_{k+1}^{new} = V_{m+1}P_{k+1}$, $Z_k^{new} = Z_m P_k$ and $\bar{H}_k^{new} = P_{k+1}^H \bar{H}_m P_k$, so that the following relations are satisfied:

$$AZ_m P_k = V_{m+1}P_{k+1}P_{k+1}^H \bar{H}_m P_k \quad ;\text{i.e.,} \quad AZ_k^{new} = V_{k+1}^{new}\bar{H}_k^{new} \tag{3.9}$$

where $\bar{H}_k^{new}$ is generally a dense matrix.

Computation of	FGMRES(m)	GMRES-DR(m,k)	FGMRES-DR(m,k)
$V_m(:, 1:k+1)$	$kop_A + nk(2k+5)$	$2n(m+1)(k+1)$	$2n(m+1)(k+1)$
$Z_m(:, 1:k)$	kop_M	-	$2nmk$
$V_m(:, k+2:m+1)$	$(m-k)op_A +$ $n(m-k)(2m+2k+5)$	$(m-k)(op_A + op_M) +$ $n(m-k)(2m+2k+5)$	$(m-k)op_A +$ $n(m-k)(2m+2k+5)$
$Z_m(:, k+1:m)$	$(m-k)op_M$	-	$(m-k)op_M$
c	2n	$2n(m+1)$	$2n(m+1)$

TABLE 3.1
Computational cost of a generic cycle of FGMRES(m), GMRES-DR(m,k) and FGMRES-DR(m,k).

shown in Table 3.1. This cost can be first reduced by observing that the residual r_0 belongs to the subspace spanned by the columns of V_{k+1}, consequently only its first $(k+1)$ entries are non-zero. These quantities can be obtained by computing $V_{k+1}^H r_0$ and it only requires $2n(k+1)$ operations. This has been notably investigated in [21]. The calculation of c can be even more reduced as described in Proposition 3.

PROPOSITION 3. *The first $(k+1)$ components of the right-hand side c of the next least-squares problem are given by the last column of Γ_{k+1}, the triangular factor of the QR factorization of the matrix G_{k+1} defined in relation (3.5).*

Proof. In Proposition 2 we have shown that $\rho_m = P_{k+1}\begin{bmatrix} u_{k\times 1} \\ a \end{bmatrix}$. Consequently $r_0 = V_{m+1}\rho_m = V_{k+1}^{new}\begin{bmatrix} u_{k\times 1} \\ a \end{bmatrix}$. Thus the right-hand side of the new least-squares problem is given by

$$c = V_{m+1}^H r_0 = V_{m+1}^H V_{k+1}^{new}\begin{bmatrix} u_{k\times 1} \\ a \end{bmatrix} = \begin{bmatrix} u_{k\times 1} \\ a \\ 0_{(m-k)\times 1} \end{bmatrix}.$$

□

We note that Proposition 3 holds for both GMRES-DR(m,k) and FGMRES-DR(m,k).

3.2.2. Storage requirements. Regarding storage, we have only included the storage proportional to the size of the original problem n which is supposed to be much larger than m and k.

Standard. With this convention FGMRES-DR(m,k) requires the storage of Z_m, V_{m+1} and at most $k+1$ additional vectors to store in turn V_{k+1}^{new} and Z_k^{new}. Thus FGMRES-DR(m,k) requires the storage of $(2m+k+2)$ vectors of length n.

Buffered. If an extra memory block of $buff\ size$ can be allocated, a blocked matrix-matrix product can be implemented to perform $V_{k+1}^{new} = V_{m+1}P_{k+1}$ and $Z_k^{new} = Z_m P_k$, that computes these matrices block-row by block-row before overwriting the result in the data structure allocated for V_{m+1} (Z_m respectively). The definition of this block size can be governed by the BLAS-3 performance of the targeted computer.

Economic. A reduction of storage is however still possible. It can indeed be remarked that Z_k^{new} and V_{k+1}^{new} can overwrite Z_k and V_{k+1}. This can be accomplished by performing the matrix multiplications $V_{k+1} \leftarrow V_{m+1}P_{k+1}$ and $Z_k \leftarrow Z_m P_k$ of Step 6 in Algorithm 6 *in place*, i.e., within the arrays V_{m+1} and Z_m. Here we have exploited the fact that multiplications involving triangular factors can be done in place. It is therefore advisable to perform a LU factorization with complete pivoting of P_{k+1} to obtain a very good approximation $\Pi P_{k+1}\Sigma = LU$, and then, to perform successively the operations $X \leftarrow XL$ and $X \leftarrow XU$ and the corresponding permutations e.g. for X being V. This approach leads to a storage of $(2m+1)$ vectors of length n only. It is clearly saving a lot of memory when k is close to m, but may introduce additional round-off errors that can hopefully be monitored by inspecting the quantity $\dfrac{\|\Pi P_k\Sigma - LU\|}{\|P_k\|}$.

Table 3.2 summarizes the requirements related to the storage for both GMRES-DR(m,k) and FGMRES-DR(m,k). We note that the economic variant of FGMRES-DR(m,k) needs the same amount of memory as FGMRES(m) and that flexible variants require m additional vectors with respect to non flexible variants.

Strategy	GMRES-DR(m,k)	FGMRES-DR(m,k)
Standard	$n(m+k+2)$	$n(2m+k+2)$
Buffered	$n(m+1)+buff\ size$	$n(2m+1)+buff\ size$
Economic	$n(m+1)$	$n(2m+1)$

TABLE 3.2
Storage required for GMRES-DR(m,k) and FGMRES-DR(m,k).

4. Numerical experiments. In this section we investigate the numerical behavior of the FGMRES-DR(m,k) algorithm on both academic and realistic applications. We consider the case of both sparse or dense matrices in either real or complex arithmetic. All the examples include a detailed comparison with FGMRES(m). This allows us to show the effects of incorporating the deflation strategy in the flexible preconditioning framework.

In the following experiments, the right-hand sides are computed as $b = A\mathbf{1}$ where $\mathbf{1}$ is the vector of appropriate dimension with all components equal to one. A zero initial iterate x_0 is considered as an initial guess and the following stopping criterion is used:

$$\frac{\|b - A\,x_\ell\|}{\|b\|} \leq 10^{-12} \tag{4.1}$$

where ℓ represents the step when the iterations are stopped. The numerical tests in Sections 4.1 and 4.2 were performed on a personal computer running Linux (Intel Pentium IV, 2.4 Ghz with 2 GB of memory) using Matlab version 7.1 (release 14). The numerical results shown in Section 4.3 were obtained on one processor of a Cray-XD1 computer (AMD Opteron 2.4 Ghz with 2 GB of memory) using a Fortran implementation. This code was compiled by the Portland Group compiler suite with the best optimization options and linked with the vendor BLAS and LAPACK subroutines, optimized for AMD architectures.

4.1. Harwell-Boeing and Matrix Market test problems. In order to illustrate the numerical behavior of FGMRES-DR(m,k), we first consider a few test matrices from the Harwell-Boeing [11] and Matrix Market [2] libraries so that any reader could reproduce these experiments. The sparse matrices named Sherman4, Saylor4 and Young1c have been chosen. Sherman4 and

Saylor4 are real matrices, whereas Young1c is a complex-valued one. They represent challenging sparse matrices coming from realistic applications (reservoir modelling, acoustics) that are often used to analyze the behaviour of numerical algorithms. For those experiments, the preconditioner consists in five steps of preconditioned full GMRES, where the preconditioner is based on an ILU(0) factorization. In the case of Sherman4 only, the inner solver corresponds to five steps of unpreconditioned full GMRES.

In Table 4.1, we depict the total number of matrix-vector products performed in the inner and outer parts of the solver (Mv) and the total number of dot products (dot) for several flexible methods. We also display the ratios of total memory and total floating point operations where the reference is the corresponding quantity of the full FGMRES method; i.e.,

$$r_{ops} = \frac{flops(Krylov\ solver)}{flops(full\ FGMRES)} \text{ and } r_{mem} = \frac{mem(Krylov\ solver)}{mem(full\ FGMRES)}, \tag{4.2}$$

where we assume that the memory allocated for full FGMRES is exactly what is needed to store Z_ℓ and $V_{\ell+1}$, ℓ being the step where convergence is achieved.

In order to illustrate the possible benefit of using the economic implementation presented in Section 3.2.2 we effectively consider different combinations of restart parameters and harmonic Ritz values for the flexible methods. Indeed the performance of FGMRES-DR(5,3) can be compared with FGMRES(5) if the economic variant is implemented or with FGMRES(7) if a standard implementation is considered (see Table 3.2). The total amount of floating point operations spent in matrix-vector products, dot products, preconditioning and basis orthogonalization has been computed for each solution method, excluding however the cost of the ILU(0) factorization that is identical for each proposed method. We have also indicated the results related to full FGMRES as a reference solution method; i.e., when memory is not constrained. It can be noticed that flexible methods with deflated restarting enables a faster convergence than those with standard restarting. It also results in a faster calculation since a significant amount of floating point operations is saved. Moreover we can also note that the performances of FGMRES-DR(10,5) in terms of floating point operations are close to those of full flexible GMRES especially when considering the Sherman4 and Saylor4 matrices. Those results also highlight the benefit of using deflated restarting as it enables a significant saving in memory.

	Sherman4				Saylor4				Young1c			
	Mv	dot	r_{ops}	r_{mem}	Mv	dot	r_{ops}	r_{mem}	Mv	dot	r_{ops}	r_{mem}
FGMRES-DR(5,3)	373	1288	1.41	0.14	115	384	1.10	0.30	1633	5698	2.60	0.08
FGMRES(5)	1273	3813	3.56	0.14	409	1221	3.22	0.30	6145	18430	7.41	0.08
FGMRES(7)	877	2771	2.54	0.19	295	931	2.39	0.41	5095	16126	6.33	0.11
FGMRES-DR(10,5)	247	951	1.02	0.27	109	396	1.08	0.57	967	3831	1.71	0.15
FGMRES(10)	979	3331	2.97	0.27	175	590	1.46	0.57	3619	12351	4.69	0.15
FGMRES(13)	649	2358	2.06	0.35	145	517	1.25	0.73	3205	11742	4.33	0.19
full FGMRES	229	1311	1.00	1.00	109	441	1.00	1.00	421	3535	1.00	1.00

Table 4.1

Performance of FGMRES(m) and FGMRES-DR(m,k) to satisfy the convergence threshold (4.1); Mv is the total number of matrix vector products, dot the total number of dot products and r_{ops} and r_{mem} are the ratios of floating point operations and memory respectively where the reference method is full FGMRES (see Equation (4.2)).

4.2. Two-dimensional Helmholtz problem. Our goal in this section is to illustrate the performance of FGMRES with deflated restarting on a simple two-dimensional partial differential equation model problem. In order to illustrate the effect of the part of the spectrum targeted by the deflation, we report first in Figures 4.3 and 4.4 on a numerical example where the preconditioner is fixed so that FGMRES-DR reduces to GMRES-DR. This fixed preconditioner approach enables us to display in Figure 4.2 (Figure 4.1) the complete spectrum of the preconditioned matrix (original matrix respectively). We then investigate a variable preconditioner and apply the same strategies in the selection of the deflated eigenvectors for all the other presented results. We consider a model

wave propagation problem in a two-dimensional homogeneous medium:

$$-\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right)-\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial y}\right)-\sigma^2\,u=f \qquad \text{in} \qquad \Omega=]0,1[^2 \tag{4.3}$$

with homogeneous Dirichlet boundary conditions $u = 0$ on the boundary $\partial\Omega$. The unknown u represents the pressure field in the frequency domain and σ the constant wavenumber. A second order finite difference discretization scheme of the Helmholtz equation (4.3) is used on a equidistant Cartesian grid of step size h with the following dispersion stability condition $\sigma\, h = 0.625$ [7] being satisfied. One V(1,1) cycle of a geometric multigrid method [32] is used as a preconditioner. This multigrid method uses a two-level hierarchy with a red-black Gauss-Seidel smoother, bilinear interpolation as prolongation and its adjoint as restriction operator. Galerkin coarse grid discretization is employed to build the coarse grid operator and a sparse direct solution method is used to solve the coarse grid systems. Numerical experiments with this two-grid preconditioner for FGMRES on two-dimensional wave propagation problems in geophysics with Robin boundary conditions have been reported in [10]. The discretization of the Helmholtz equation on a 64×64 grid with $\sigma = 40$ leads to a real-valued sparse symmetric indefinite matrix A, whose spectrum is shown in Figure 4.1. There are 117 negative eigenvalues for this choice of wavenumber and step size. The spectrum of the preconditioned operator AM is also shown in Figure 4.2. It exhibits both positive and negative isolated real eigenvalues and a cluster of eigenvalues around $(1, 0)$.

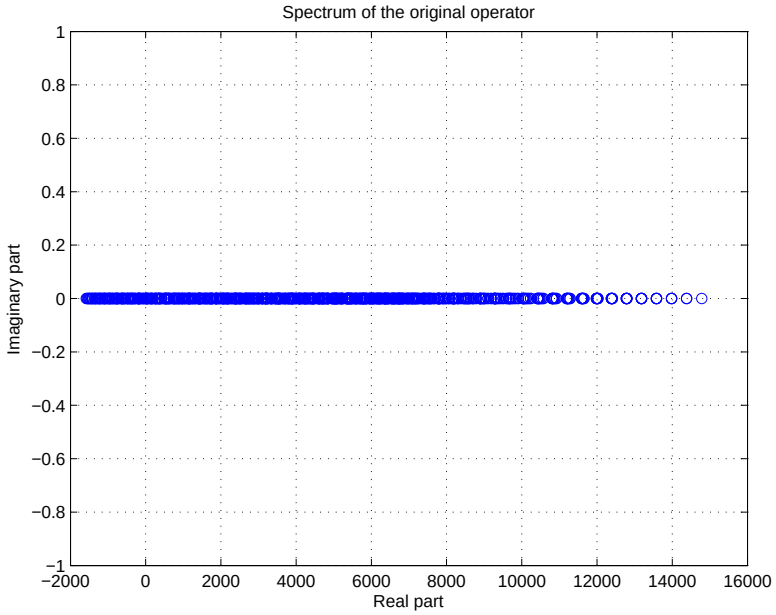

FIG. 4.1. *Spectrum of the original Helmholtz operator. Case of* $h = 1/64$ *with* $\sigma = 40$.

The harmonic Ritz vectors corresponding to the k eigenvalues λ_i of smallest magnitude of the matrix $H_m + h^2_{m+1,m} H_m^{-H} e_m e_m^T$ have been chosen in Algorithm 6 (step 3). In the sequel we call this strategy SMALLEST. However any combination of k harmonic vectors may be selected. Thus we have considered two other possibilities. The first one selects the k eigenpairs corresponding to the eigenvalues of largest magnitude. It is called LARGEST. The second deflation strategy retains the k eigenvectors associated with the eigenvalues such as $|1-\lambda_i|$ is of largest magnitude. With this latter choice we aim at selecting eigenvalues located away from a cluster around the eigenvalue of the "ideal" preconditioned operator AM with $M^{-1} = A$. This possibly allows simultaneous deflation of eigenvalues of both smallest and largest magnitude. We call this strategy CLUSTER. We investigate the influence of the different deflation strategies (SMALLEST, LARGEST and CLUSTER respectively) and compare FGMRES-DR(m,k) with FGMRES(m) for different values of the restart parameter m. Table 4.2 gives the number of approximate eigenpairs k that led to the smallest number of

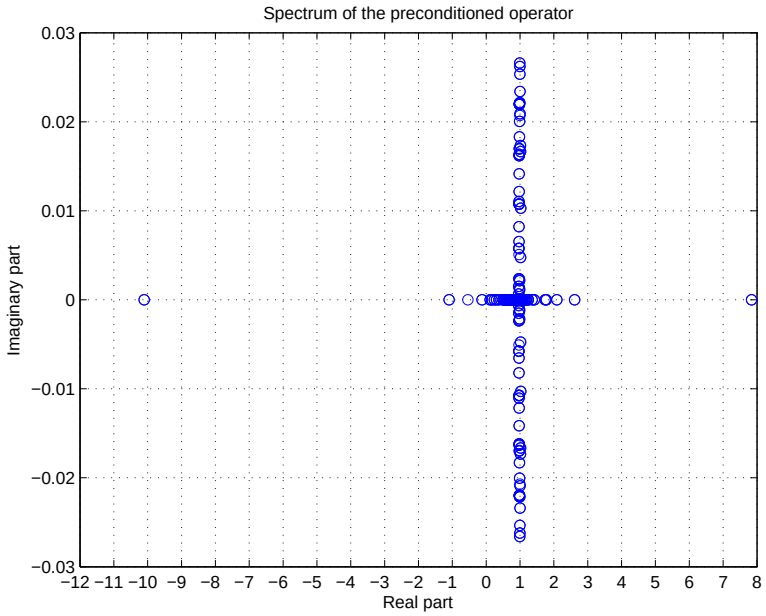

FIG. 4.2. *Spectrum of the preconditioned operator, when a V(1,1) cycle of a multigrid method is used as a preconditioner. Case of $h = 1/64$ with $\sigma = 40$. Note the different scalings on Figures 4.1 and 4.2*

iterations ℓ to satisfy the stopping criterion (4.1) for each deflation strategy. These results have been obtained by running FGMRES-DR(m,k) with $1 \leq k < m$ for each value of the restart parameter m. If the goal is to minimize the number of iterations, numerical results show that the CLUSTER deflation strategy is almost the most efficient on this application (see bold values in Table 4.2) leading sometimes to a significant reduction. The total amount of floating point operations including the cost of preconditioning has been computed for each solution method. We choose FGMRES(m) as a reference solution method and report the following normalized quantity in Table 4.2:

$$r_{ops} = \frac{flops(Strat)}{flops(FGMRES(m))} \tag{4.4}$$

where *Strat* denotes the FGMRES-DR(m,k) solution method with a given deflation strategy among SMALLEST, LARGEST and CLUSTER. Consequently values of r_{ops} less than 1 indicate which solution methods are expected to be more efficient than FGMRES(m) in terms of computational work. In this table, we see that the SMALLEST deflation strategy yields the best performance with respect to floating point operations on this application (see italic values in Table 4.2). The CLUSTER deflation strategy tends to favour values of k close to the restart parameter m to be most effective. A possible explanation is that this deflation strategy captures first the set of few outlier eigenvalues and then the set of real eigenvalues close to zero (see Figure 4.2) - also captured by the SMALLEST deflation strategy. Figures 4.3 and 4.4 show a typical convergence history on this wave propagation problem for two different settings of (m,k).

The choice of the two-grid components has led to a fixed preconditioner. This allowed us to compute the spectrum of the preconditioned operator shown in Figure 4.2. The efficiency of FGMRES with deflated restarting has been shown on this simple model problem. This is of primary interest for three-dimensional wave propagation applications, where the coarse grid systems of the two-grid method can not be handled any more by a sparse direct solution method due to excessive memory requirements. Iterative methods are then required to solve the coarse grid systems only approximately. A non constant preconditioner is then obtained which requires the use of *flexible* Krylov subspace methods. The study of preconditioned FGMRES-DR for such three-dimensional wave propagation applications is beyond the scope of this paper and will be analyzed in the near future. Nevertheless we give an illustration of the potential benefits of FGMRES-DR on the two-

	FGMRES		Smallest			Largest			Cluster		
m	ℓ	r_{ops}	k	ℓ	r_{ops}	k	ℓ	r_{ops}	k	ℓ	r_{ops}
10	492	1.00	6	**161**	*0.55*	4	355	0.98	7	175	0.71
12	194	1.00	2	97	*0.55*	4	181	1.25	9	**76**	0.87
14	148	1.00	5	75	*0.66*	3	138	1.11	10	**58**	0.80
16	152	1.00	8	61	*0.60*	6	127	1.16	11	**50**	0.62
18	124	1.00	6	54	*0.55*	2	103	0.90	16	**45**	1.25
20	101	1.00	7	52	*0.65*	4	85	0.98	16	**41**	0.90

Table 4.2

Wave propagation problem ($h = 1/64$, $\sigma = 40$). Case of a constant two-grid preconditioner. On each line is shown the iso-memory performance of FGMRES and FGMRES-DR; ℓ is the number of iterations required to satisfy the stopping criterion (4.1) and r_{ops} the ratio of total floating point operations v.s. FGMRES(m) (see Equation (4.4)). Best values of ℓ are marked in bold, while best values of r_{ops} are marked in italic.

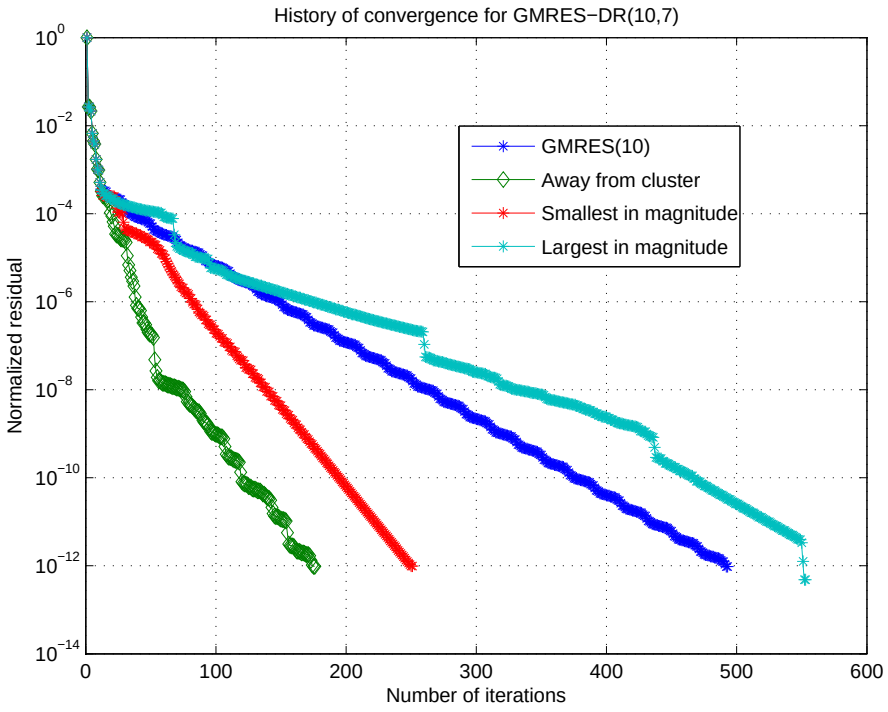

Fig. 4.3. *Convergence history of the scaled residual with respect to number of iterations on the wave propagation problem ($h = 1/64$, $\sigma = 40$). Experiments with GMRES-DR(10,7). Case of a constant two-grid preconditioner.*

dimensional model problem (4.3) when such inexact coarse grid solution method is used. As an example of approximate coarse grid solver, we consider now the use of an iterative method to solve the coarse grid system to a loose tolerance of 0.15 on the normalized residual. Table 4.3 reports the results for the two promising deflation strategies Smallest and Cluster in this setting. The same conclusions as in the constant preconditioner case hold: FGMRES with deflated restarting is efficient. This case study illustrates that there are possibly better choices than selecting the harmonic Ritz vectors corresponding to the harmonic Ritz values of smallest magnitude. If the goal is to minimize the number of matrix-vector products the Cluster policy is the most efficient on that problem.

4.3. Three-dimensional Maxwell's equations in the frequency domain. The boundary element method has become a popular tool in computational electromagnetics for the solution of Maxwell's equations in the frequency domain. These simulations are very demanding in terms of computer resources, and require fast and efficient numerical methods. Using the equivalence

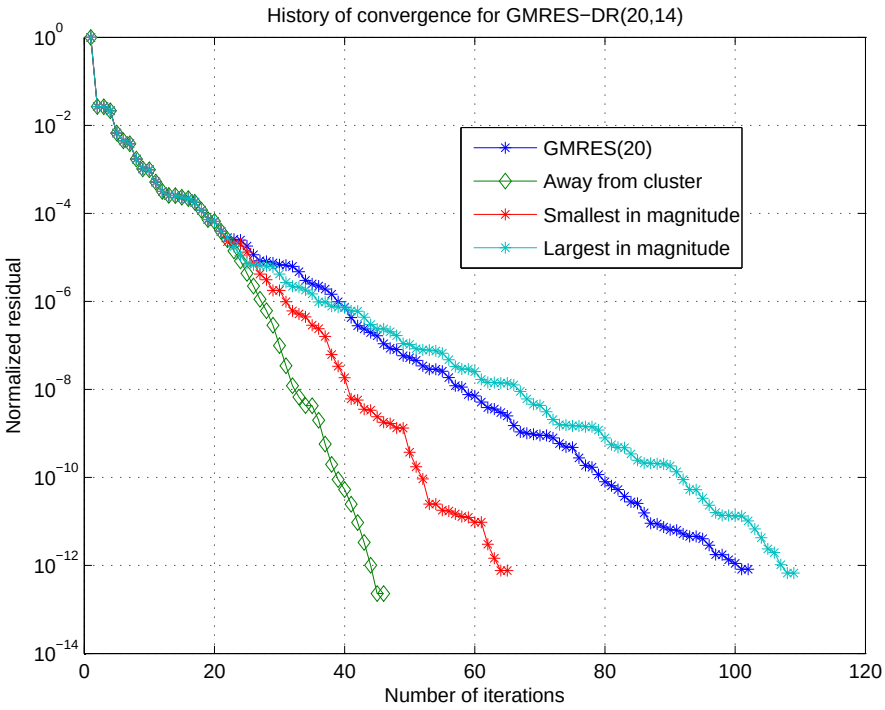

FIG. 4.4. *Convergence history of the scaled residual with respect to number of iterations on the wave propagation problem ($h = 1/64$, $\sigma = 40$). Experiments with GMRES-DR(20,14). Case of a constant two-grid preconditioner.*

	FGMRES		SMALLEST			CLUSTER		
m	ℓ	r_{ops}	k	ℓ	r_{ops}	k	ℓ	r_{ops}
10	488	1.00	3	**115**	*0.28*	8	125	0.66
12	208	1.00	2	96	*0.50*	10	**87**	1.25
14	141	1.00	8	75	*0.87*	12	**74**	1.78
16	156	1.00	8	67	*0.68*	13	**54**	0.80
18	124	1.00	6	57	*0.60*	16	**47**	0.86
20	106	1.00	7	54	*0.66*	16	**46**	0.90

TABLE 4.3

Wave propagation problem ($h = 1/64$, $\sigma = 40$). Case of a non constant two-grid preconditioner. On each line is shown the iso-memory performance of FGMRES and FGMRES-DR; ℓ is the number of iterations required to satisfy the stopping criterion (4.1) and r_{ops} the ratio of total floating point operations v.s. FGMRES(m) (see Equation (4.4)). Best values of ℓ are marked in bold, while best values of r_{ops} are marked in italic.

principle, Maxwell's equations can be recast in the form of integral equations. The discretization is performed on the surface of the object and gives rise to a linear system, where the matrix is dense and complex. Such a linear system can be solved without explicitly forming the matrix A thanks to the fast multipole method (FMM) approximation [8, 9, 13, 30]. In this framework, the features of the fast multipole techniques can be further exploited to design an inner-outer scheme [4]. An accurate FMM is used within the outer solver as it governs the final accuracy of the computed solution. The inner solver, that acts as a flexible preconditioner, consists in a few steps of full GMRES preconditioned by a sparse approximate inverse preconditioner [1, 6] and uses a less accurate FMM.

In this section, we consider a complex geometry that corresponds to an air intake of an aerospace industry object. Such a cavity is known to be particularly challenging to solve. The dimension of the linear system is 16 950 for the frequency considered in that example.

In Figure 4.5 we depict the convergence history for both FGMRES and FGMRES-DR where

the inner solver is one restart of GMRES(30) with a sparse approximate inverse preconditioner based on Frobenius norm minimization. The restart parameters of FGMRES and FGMRES-DR are chosen so that both solvers use the same amount of storage that corresponds to 51 vectors of length n. For this implementation the trick based on the LU decomposition with complete pivoting of P_k at restarting was not implemented in the prototype code. Based on a previous work [12], where a deflating preconditioning technique targeting the smallest eigenvalues in magnitude was very successful, we select the same part of the spectrum for these experiments. The history is plotted at the iteration when the methods start generating different iterates; that is after the smallest restart considered for FGMRES-DR.

It can be seen that FGMRES-DR converges significantly faster than regular FGMRES, especially when the number of deflated directions is increased. As it could be expected if too many directions are deflated the performance deteriorates (see $k = 11$ v.s. $k = 13$ in the graph). The convergence remains worse than full FGMRES but FGMRES-DR is much less memory consuming as $r_{mem} = 0.47$. The gain would become larger if more accurate solutions were expected. On that large electromagnetics calculation, the extra $\mathcal{O}(k)$ operations are completely negligible and the saving in iteration count directly results in a computational time saving. For instance for a scaled residual norm lower than 10^{-11} on one processor of a Cray-XD1 computer, the CPU time is about 5 hours 47 minutes with FGMRES(25) and only about 3 hours 19 minutes for FGMRES-DR(19,11).

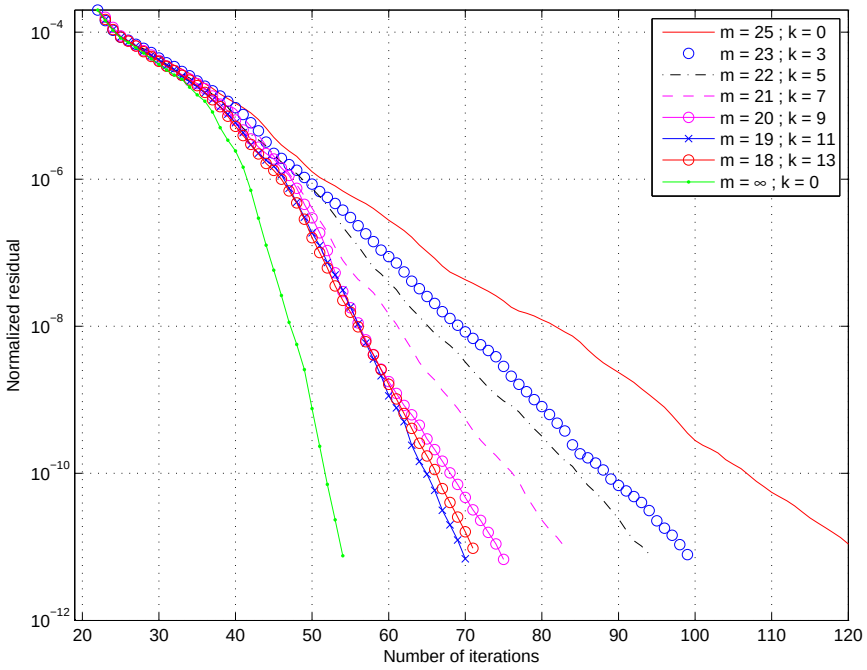

FIG. 4.5. *Convergence history of the scaled residual with respect to the iteration for the electromagnetics application.*

5. Concluding remarks. There are many situations in scientific computing where variable preconditioners have to be considered for the iterative solution of a linear system. In that framework we have proposed a novel algorithm that attempts to combine the numerical features of GMRES-DR and the flexibility of FGMRES. The new algorithm, referred to as FGMRES-DR, inherits from the attractive numerical properties of its two parents. We have shown, on a set of small test examples as well as on two real life applications in wave propagation that, after the first restart of the method, FGMRES-DR may outperform FGMRES; the benefit obtained is problem dependent. As for the GMRES-DR algorithm, the eigenvalues of smallest magnitude are often considered as good candidates for the restarting procedure. However, any other part of the spectrum can be considered; the best suited choice is again problem-dependent and heuristics could be based on the analysis of the effect of the preconditioner on the system matrix or on the location of all the

harmonic Ritz values available at restart.

Acknowledgments. We would like to thank the referees for their careful readings and many valuable suggestions that helped us to improve our manuscript.

REFERENCES

[1] G. Alléon, M. Benzi, and L. Giraud. Sparse approximate inverse preconditioning for dense linear systems arising in computational electromagnetics. *Numerical Algorithms*, 16:1–15, 1997.
[2] R. F. Boisvert, R. Pozo, K. Remington, R. Barrett, and J. J. Dongarra. The Matrix Market: A web resource for test matrix collections. In Ronald F. Boisvert, editor, *Quality of Numerical Software, Assessment and Enhancement*, pages 125–137, London, 1997. Chapman & Hall.
[3] C. Le Calvez and B. Molina. Implicitly restarted and deflated GMRES. *Numerical Algorithms*, 21:261–285, 1999.
[4] B. Carpentieri, I. S. Duff, L. Giraud, and G. Sylvand. Combining fast multipole techniques and an approximate inverse preconditioner for large electromagnetism calculations. *SIAM J. Scientific Computing*, 27(3):774–792, 2005.
[5] A. Chapman and Y. Saad. Deflated and augmented Krylov subspace techniques. *Numerical Linear Algebra with Applications*, 4:43–66, 1996.
[6] K. Chen. An analysis of sparse approximate inverse preconditioners for boundary integral equations. *SIAM J. Matrix Analysis and Applications*, 22(3):1058–1078, 2001.
[7] G. Cohen. *Higher-order numerical methods for transient wave equations.* Springer-Verlag, 2002.
[8] E. Darve. The fast multipole method (I) : Error analysis and asymptotic complexity. *SIAM J. Numerical Analysis*, 38(1):98–128, 2000.
[9] E. Darve. The fast multipole method: Numerical implementation. *J. Comp. Phys.*, 160(1):195–240, 2000.
[10] I. S. Duff, S. Gratton, X. Pinel, and X. Vasseur. Multigrid based preconditioners for the numerical solution of two-dimensional heterogeneous problems in geophysics. *International Journal of Computer Mathematics*, 84-88:1167–1181, 2007.
[11] I. S. Duff, R. G. Grimes, and J. G. Lewis. Users' guide for the Harwell-Boeing sparse matrix collection (Release I). Technical Report TR/PA/92/86, CERFACS, 1992.
[12] L. Giraud, S. Gratton, and E. Martin. Incremental spectral preconditioners for sequences of linear systems. *Applied Numerical Mathematics*, 57(11-12):1164–1180, 2007.
[13] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *Journal of Computational Physics*, 73:325–348, 1987.
[14] R. B. Morgan. A restarted GMRES method augmented with eigenvectors. *SIAM J. Matrix Analysis and Applications*, 16:1154–1171, 1995.
[15] R. B. Morgan. Implicitly restarted GMRES and Arnoldi methods for nonsymmetric systems of equations. *SIAM J. Matrix Analysis and Applications*, 21(4):1112–1135, 2000.
[16] R. B. Morgan. GMRES with deflated restarting. *SIAM J. Scientific Computing*, 24(1):20–37, 2002.
[17] C. C. Paige, B. N. Parlett, and H. A. Van der Vorst. Approximate solutions and eigenvalue bounds from Krylov subspaces. *Numerical Linear Algebra with Applications*, 2:115–134, 1995.
[18] C. C. Paige, M. Rozložník, and Z. Strakos. Modified Gram-Schmidt (MGS), least squares, and backward stability of MGS-GMRES. *SIAM Journal on Matrix Analysis and Applications*, 28(1):264–284, 2006.
[19] C. C. Paige and Z. Strakos. Residual and backward error bounds in minimum residual Krylov subspace methods. *SIAM Journal on Scientific Computing*, 23(6):1898–1923, 2002.
[20] M. Parks, E. de Sturler, G. Mackey, D.D. Johnson, and S. Maiti. Recycling Krylov subspaces for sequences of linear systems. *SIAM J. Scientific Computing*, 28(5):1651–1674, 2006.
[21] S. Röllin and W. Fichtner. Improving the accuracy of GMRES with deflated restarting. *SIAM J. Scientific Computing*, 30(1):232–245, 2007.
[22] Y. Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM J. Scientific and Statistical Computing*, 14:461–469, 1993.
[23] Y. Saad. Analysis of augmented Krylov subspace methods. *SIAM J. Matrix Analysis and Applications*, 18(2):435–449, 1997.
[24] Y. Saad. *Iterative Methods for Sparse Linear Systems.* SIAM, Philadelphia, 2003. Second edition.
[25] Y. Saad and M. H. Schultz. GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Scientific and Statistical Computing*, 7:856–869, 1986.
[26] Y. Saad and K. Wu. DQGMRES: a direct quasi-minimal residual algorithm based on incomplete orthogonalization. *Numerical Linear Algebra Appl*, 3:3–329, 1996.
[27] V. Simoncini and D. B. Szyld. Recent computational developments in Krylov subspace methods for linear systems. *Numerical Linear Algebra with Applications*, 14:1–59, 2007.
[28] G. L. G. Sleijpen and H. A. Van der Vorst. A Jacobi–Davidson iteration method for linear eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 17(2):401–425, 1996.
[29] B. F. Smith, P. E. Bjørstad, and W. D. Gropp. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations.* Cambridge University Press, 1996.
[30] G. Sylvand. *La Méthode Multipôle Rapide en Electromagnétisme : Performances, Parallélisation, Applications.* PhD thesis, Ecole Nationale des Ponts et Chaussées, 2002.
[31] A. Toselli and O. Widlund. *Domain Decomposition methods - Algorithms and Theory.* Springer Series on Computational Mathematics, vol. 34, Springer, 2004.
[32] U. Trottenberg, C. W. Oosterlee, and A. Schüller. *Multigrid.* Academic Press Inc., 2001.

[33] H. A. van der Vorst and C. Vuik. GMRESR: A family of nested GMRES methods. *Numerical Linear Algebra with Applications*, 1:369–386, 1994.