SIAM J. SCI. COMPUT.
Vol. 29, No. 2, pp. 710–728

CONVERGENCE IN BACKWARD ERROR OF RELAXED GMRES*

L. GIRAUD†, S. GRATTON‡, AND J. LANGOU§

Abstract. This work is the follow-up of the experimental study presented in [A. Bouras and V. Frayssé, *SIAM J. Matrix Anal. Appl.*, 26 (2005), pp. 660–678]. It is based on and extends some theoretical results in [V. Simoncini and D. B. Szyld, *SIAM J. Sci. Comput.*, 25 (2003), pp. 454–477; J. van den Eshof and G. L. G. Sleijpen, *SIAM J. Matrix Anal. Appl.*, 26 (2004), pp. 125–153]. In a backward error framework we study the convergence of GMRES when the matrix-vector products are performed inaccurately. This inaccuracy is modeled by a perturbation of the original matrix. We prove the convergence of GMRES when the perturbation size is proportional to the inverse of the computed residual norm; this implies that the accuracy can be relaxed as the method proceeds which gives rise to the terminology "relaxed GMRES." As for the exact GMRES we show under proper assumptions that only happy breakdowns can occur. Furthermore, the convergence can be detected using a byproduct of the algorithm. We explore the links between relaxed right-preconditioned GMRES and flexible GMRES (FGMRES). In particular, this enables us to derive a proof of convergence of FGMRES. Finally, we report results of numerical experiments to illustrate the behavior of the relaxed GMRES monitored by the proposed relaxation strategies.

Key words. GMRES, backward error analysis, relaxed GMRES

AMS subject classifications. 65F10, 65G99

DOI. 10.1137/040608416

1. Introduction.

We consider the solution of a linear system of equations $Ax = b$ using the GMRES [12] iterative method, where A is a nonsingular $n \times n$ matrix. In some applications, performing inexact matrix-vector products in this method may be interesting as long as the convergence of GMRES is maintained. Many situations in scientific computing may benefit from such a scheme. For instance, a natural application of this idea occurs in computational electromagnetics, where the fast multipole method provides approximations of the matrix-vector product within a user-defined accuracy; the less accurate the matrix-vector, the faster the computation. The key point is then to design a criterion to control the accuracy of the matrix-vector product so that the iterates achieve a satisfactory convergence level. Another example arises in nonoverlapping domain decomposition where the matrix-vector involving the Schur complement can be approximated.

In [4], a criterion is proposed for general systems and its numerical behavior is illustrated on a large set of numerical experiments. This work is based on some heuristic considerations and the approach is referred to as a relaxation strategy because the perturbation size can grow as the inverse of the residual norm. We denote by "relaxed GMRES" an inexact GMRES that implements a relaxation strategy. The relaxation strategy proposed in [4] attempts to ensure the convergence of the GMRES iterates

*Received by the editors May 15, 2004; accepted for publication (in revised form) August 25, 2006; published electronically April 10, 2007. This work was initiated while the three authors were working at CERFACS.

http://www.siam.org/journals/sisc/29-2/60841.html

†ENSEEIHT, 2, rue Charles Camichel, BP 7122, 31071 Toulouse Cedex 7, France (giraud@enseeiht.fr).

‡CERFACS, 42 av. Gaspard Coriolis, 31057 Toulouse cedex 1, France (gratton@cerfacs.fr).

§Department of Mathematical Sciences, The University of Colorado at Denver and Health Sciences Center, Denver, CO (Julien.Langou@cudenver.edu).

x_k within a relative normwise backward error

$$\begin{aligned} \eta_{A,b}(x_k) &= \min_{\Delta A, \Delta b}\{\tau > 0 : \|\Delta A\| \le \tau\|A\|,\ \|\Delta b\| \le \tau\|b\| \\ &\qquad \text{and } (A+\Delta A)x_k = b + \Delta b\} \\ &= \frac{\|Ax_k - b\|}{\|A\|\|x_k\| + \|b\|}, \end{aligned} \tag{1.1}$$

less than a prescribed quantity $\varepsilon > 0$. A significant number of numerical experiments are presented that illustrate the merits of the strategy but also reveal its lack of robustness; for a few examples the convergence at ε is not obtained. Nevertheless, based on [4], similar strategies have been successfully applied to the solution of heterogeneous diffusion problems using domain decomposition [3], to the preconditioning of a radiation diffusion problem [20], to the solution of electromagnetic problems [11], in lattice quantum chromodynamics [5], and in an ocean circulation model for steady barotropic flow [18]. A significant step toward a theoretical explanation of this observed behavior is proposed in [15, 19]. In these latter works, important justifications are brought to the fact that under some assumptions relaxed GMRES converges in residual norm.

The convergence of iterative solvers is often based on normwise backward error criteria [2, 6, 9]. In this paper we propose criteria to control the accuracy of the matrix-vector products and show that they ensure the convergence of GMRES with respect either to $\eta_{A,b}(x_k)$ defined in (1.1) or to

$$\begin{aligned} \eta_b(x_k) &= \min_{\Delta b}\{\tau > 0 : \|\Delta b\| \le \tau\|b\| \text{ and } Ax_k = b + \Delta b\} \\ &= \frac{\|Ax_k - b\|}{\|b\|}. \end{aligned} \tag{1.2}$$

We mention that $\eta_{A,b}$ and η_b are recommended in [2] when the concern related to the stopping criterion is discussed.

The forthcoming development shows that deriving a sufficient convergence condition of relaxed GMRES for the stopping criterion $\eta_b(x_k) \le \varepsilon$ is far simpler than for $\eta_{A,b}(x_k) \le \varepsilon$. However, we believe that both stopping criteria are of practical interest. Both may be used to account for data uncertainties. It is problem and application dependent to decide whether these uncertainties come chiefly from the right-hand side, when η_b must be used, or from the matrix and the right-hand side, when $\eta_{A,b}$ must be used.

The paper is organized as follows. In section 2 we study the convergence of relaxed GMRES. We first recall and discuss in section 2.1 some theoretical results established in [15, 19]. Section 2.2 is devoted to the convergence proofs. In particular, we show that under suitable assumptions only happy breakdowns can occur and we explain how the perturbation size can be monitored to ensure the convergence of relaxed GMRES with respect to either $\eta_{A,b}$ or η_b. We provide a stopping criterion that uses only byproducts of the algorithm and does not require an exact product by A to compute the backward errors. Numerical experiments that illustrate these theoretical results are given in section 2.3. In section 3, we study the situation when the preconditioner is perturbed. This might happen, for instance, when the application of the preconditioner is obtained by solving iteratively an auxiliary linear system. We also exploit the links between relaxed preconditioned GMRES and flexible GMRES (FGMRES) [13], and we derive a straightforward proof of convergence of FGMRES. In section 4 we consider the use of these strategies in a restarted framework. Finally, we conclude with some comments in section 5.

In this paper, the 2-norm of a vector x is denoted by $\|x\|$, and the spectral 2-norm of a matrix A is denoted by $\|A\|$. We use the notation $\sigma_{\max}(A)$ ($\sigma_{\min}(A)$) for the smallest (resp., the largest) singular value of A. The spectral condition number of a matrix is $\kappa(A) = \|A\|\|A^{-1}\| = \sigma_{\max}(A)/\sigma_{\min}(A)$.

2. GMRES with inexact matrix-vector.

2.1. Background and existing results. The GMRES method is based on the Arnoldi recursion $AV_k = V_{k+1}\bar{H}_k$, where $V_k = [v_1, \ldots, v_k]$ is an orthogonal matrix, and $\bar{H}_k$ is a $(k+1) \times k$ upper-Hessenberg matrix. We assume that we are given an initial guess x_0 for the solution $x^\star$ of the system. In this paper, we suppose $x_0 \neq x^\star$. The GMRES algorithm generates a sequence of iterates $\{x_k\}_{k=1,2,\ldots}$ such that x_k realizes the minimum of the 2-norm of the residual $r_k = b - Ax_k$ over the space $x_0 + \mathcal{K}_k(A, v_1)$. The Krylov subspace $\mathcal{K}_k(A, v_1)$ is defined by $\mathcal{K}_k(A, v_1) = \mathrm{span}(v_1, Av_1, \ldots, A^{k-1}v_1)$, where $\beta v_1 = r_0$ and $\beta = \|r_0\| \neq 0$. From the unitary invariance of the 2-norm, we obtain $x_k = x_0 + \delta x_k$, where $\delta x_k = V_k y_k$ and y_k is the solution of the linear least-squares problem $\min_y \|\bar{H}_k y - \beta e_1\|$, e_1 being the first vector of the canonical basis.

We assume that it is possible to monitor the accuracy of the matrix-vector product Av of the Arnoldi procedure. From a mathematical point of view, the inaccuracy can be modeled by introducing a perturbation matrix E, depending possibly on v, such that $(A + E)v$ is the quantity actually computed. At step k of this perturbed Arnoldi algorithm, the vector $w = (A + E_k)v_k$ is orthogonalized against the vectors v_j, $j = 1, \ldots, k$, so that the following relation holds:

$$[(A + E_1)v_1, \ldots, (A + E_k)v_k] = [v_1, \ldots, v_k, v_{k+1}]\bar{H}_k. \tag{2.1}$$

Note that this matrix perturbation approach was taken in [4, 15]. In [19] the inaccuracies are modeled by introducing the vector f_k such that $w = Av_k + f_k$. We took the former approach because it generalizes to the latter ones by setting $E_k = f_k v_k^T$ since $v_k^T v_k = 1$. In the above equality (2.1), referred to as the *inexact Arnoldi relation* in [15], $V_k = [v_1, \ldots, v_k]$ is an orthogonal matrix and $\bar{H}_k$ is a $(k+1) \times k$ upper-Hessenberg matrix. We first assume (as in [15, 19]) that the matrix-vector product occurring in the initial residual computation is exact, so that $r_0 = b - Ax_0$ and $\beta = \|b - Ax_0\|$. We define the kth iterate of the inexact method by $x_k = x_0 + \delta x_k$, where $\delta x_k = V_k y_k$ and y_k is the solution of the linear least-squares problem $\min_y \|\bar{H}_k y - \beta e_1\|$. Introducing the perturbation matrix $G_k = [E_1 v_1, \ldots, E_k v_k]$ the inexact Arnoldi relation can also be written [19] as an exact Arnoldi relation $\widetilde{A}_k V_k = (A + G_k V_k^T)V_k = V_{k+1}\bar{H}_k$, with $\widetilde{A}_k = A + G_k V_k^T$. This last equality shows that the quantities $\delta x_i, \bar{H}_i$, and v_i for $i \leq k$ generated by the inexact GMRES on $A\,\delta x = r_0$ until step k are the same as those generated by the first k steps of exact GMRES applied to the linear system $\widetilde{A}_k\,\delta x = r_0$. Using classical results on GMRES [14] this observation implies by induction that the norm of the residual $r_0 - \widetilde{A}_k\,\delta x_i$ is monotonically decreasing as i grows, where $i \leq k$. Furthermore, $\widetilde{A}_k\,\delta x_i = \widetilde{A}_i\,\delta x_i$, because $\widetilde{A}_k\,\delta x_i = (A + G_k V_k^T)V_i y_i = (A + G_i V_i^T)V_i y_i = \widetilde{A}_i\,\delta x_i$. Therefore, the norm of the *computed residual* $\tilde{r}_k = r_0 - \widetilde{A}_k\,\delta x_k$ decreases with k. Let us denote by r_k the residual $r_0 - A\,\delta x_k$ and let us define $\tilde{r}_0 = r_0$. The inexact Arnoldi recursion can also be written

$$\widetilde{A}_k V_k = V_k H_k + h_{k+1,k} v_{k+1} e_k^T.$$

Using an analogy with the terminology used for the exact GMRES algorithm, we say that a *breakdown* occurs at step m if $h_{m+1,m} = 0$. It is important to notice that, because of the orthogonality of V_k, similarly to the exact GMRES framework, such a breakdown *will occur* for $m \leq n$. In what follows, we denote by m the step where the breakdown occurs. At each step the residual gap is defined by $r_k - \tilde{r}_k = (A - \widetilde{A}_k)\, \delta x_k = G_k y_k = \sum_{i=1}^{k} y_{k,i} E_i v_i$, where $y_k = (y_{k,1}, \ldots, y_{k,k})^T \in \mathbb{R}^k$. Let ϵ be a positive real number. Assuming that the inexact method can be run until step ℓ without breakdown, it is shown in [15, Theorem 5.3] that if

$$\|E_k\| \leq \frac{\sigma_{\min}(\bar{H}_\ell)}{\ell} \frac{\epsilon}{\|\tilde{r}_{k-1}\|} \tag{2.2}$$

for $k \leq \ell$ then $\|r_\ell - \tilde{r}_\ell\| \leq \epsilon$.

This important result is easily derived from $r_\ell - \tilde{r}_\ell = \sum_{k=1}^{\ell} y_{\ell,k} E_k v_k$, $|y_{\ell,1}| \leq \|y_\ell\| = \|\bar{H}_\ell^\dagger \beta e_1\| \leq \|\bar{H}_\ell^\dagger\| \|r_0\| = \|\bar{H}_\ell^\dagger\| \|\tilde{r}_0\|$ and

$$|y_{\ell,k}| = |e_k^T \bar{H}_\ell^\dagger \beta e_1| = \left| e_k^T \bar{H}_\ell^\dagger \left(\beta e_1 - \bar{H}_\ell \begin{pmatrix} y_{k-1} \\ 0 \end{pmatrix} \right) \right| \leq \|\bar{H}_\ell^\dagger\| \|\tilde{r}_{k-1}\|,$$

where $1 < k \leq \ell$. Because the norm of the computed residual, $\|\tilde{r}_k\|$, is monotonically decreasing as k grows, $k \leq \ell$, the inequalities (2.2) and

$$\|r_k\| \leq \|r_k - \tilde{r}_k\| + \|\tilde{r}_k\| \tag{2.3}$$

ensure that the residual norm satisfies at step ℓ

$$\|r_\ell\| \leq \epsilon + \|\tilde{r}_\ell\|. \tag{2.4}$$

However, it can be seen that there is an implicit relation in (2.2) linking all the E_k's, $k \leq \ell$, together. The E_k's depend on $\bar{H}_\ell$ which itself depends on the E_k's. This means that it is not possible to implement (2.2) in an algorithm.

Another interesting feature would also be to monitor $\|\tilde{r}_\ell\|$ in (2.4) so that it can be less than any prescribed value for large enough ℓ. In this way, the residual norm $\|r_\ell\|$, which is a key ingredient of the backward errors $\eta_{A,b}(x_\ell)$ and $\eta_b(x_\ell)$, would also be controlled through (2.4). It turns out that contrary to the exact GMRES, $\|\tilde{r}_\ell\|$ is not necessarily zero at the breakdown of the inexact algorithm. For example, for $A = \mathrm{diag}(1, 2)$, $b = (1/\sqrt{2}, 1/\sqrt{2})^T$, and $x_0 = (0, 0)^T$, if we take $E_1 = 0$ and $E_2 = \mathrm{diag}(-2, 0)$, the inexact Arnoldi relation reads

$$[(A + E_1)v_1, (A + E_2)v_2] = [v_1, v_2] \begin{pmatrix} 3/2 & 3/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

We therefore have a breakdown of the inexact GMRES method, but $\|\tilde{r}_1\| = \|\tilde{r}_2\| = 1/\sqrt{10}$. Note that H_2 is singular, and consequently, the linear least-squares problem $\min_y \|\bar{H}_2 y - \beta e_1\|$ is rank deficient. This means that there is an infinite number of solutions to the linear least-squares problem and neither y_2 nor x_2 is well defined.

The purpose of the next section is to fix the problems related to the control of $\|\tilde{r}_\ell\|$ and to the possible singularity of H_m as well as to remove the implicit relation linking all the E_k's, $k \leq \ell$, together. We shall see that this will enable us to design a relaxed GMRES algorithm that is guaranteed to converge with the backward error stopping criteria $\eta_b \leq \varepsilon$ or $\eta_{A,b} \leq \varepsilon$ for any prescribed tolerance ε.

2.2. Theoretical results. We first design a strategy that ensures the convergence of the inexact GMRES iterates so that η_b can be made smaller than any prescribed tolerance; this theoretical result is closely related to the work of [15, 19]. We consider the splitting

$$\eta_b(x_k) \leq \frac{\|r_k - \tilde{r}_k\|}{\|b\|} + \frac{\|\tilde{r}_k\|}{\|b\|}$$

and set accordingly

$$\varepsilon = \varepsilon_g + \varepsilon_c, \tag{2.5}$$

where ε_g (resp., ε_c) is the targeted tolerance for the scaled residual *gap* $\frac{\|r_k-\tilde{r}_k\|}{\|b\|}$ (resp., the relative *computed* residual norm $\frac{\|\tilde{r}_k\|}{\|b\|}$).

In the framework of the inexact GMRES the situation where the computed residual $\tilde{r}_m$ is zero at the breakdown is referred to as a *happy breakdown*. Such a situation is of interest because it ensures that $\frac{\|\tilde{r}_k\|}{\|b\|}$ can be made smaller than any prescribed ε_c provided that k is large enough. The following theorem shows a possible way to control the E_k's such that only a happy breakdown eventually occurs.

THEOREM 1 (breakdown in inexact GMRES). *Let us denote by m the step where the breakdown occurs in the inexact GMRES algorithm. The following results hold for the inexact GMRES.*

1. *Suppose that $r_0 \neq 0$ and that $k \leq m$. If H_k is nonsingular, the computed residual $\tilde{r}_k$ is zero iff $h_{k+1,k}$ is zero (i.e., $k = m$).*

2. *Suppose there exists $0 < c < 1$ such that for all $k \leq m$, $\|E_k\| \leq c\frac{\sigma_{\min}(A)}{n}$. Then, the smallest singular value of the Hessenberg matrix $\bar{H}_k$ satisfies $(1-c)\sigma_{\min}(A) \leq \sigma_{\min}(\bar{H}_k)$. This latter inequality implies that H_m is nonsingular and that the breakdown at step m is a* happy *breakdown (i.e., $\tilde{r}_m = 0$).*

Proof. 1. Owing to the inexact Arnoldi relation (2.1), if $h_{k+1,k} = 0$, then we have $\tilde{r}_k = r_0 - \widetilde{A}_k V_k y_k = V_k(\beta e_1 - H_k y_k)$, where y_k minimizes $\|\beta e_1 - H_k y_k\|$. Because H_k is nonsingular, the minimum value is zero, and consequently $\tilde{r}_k = 0$.

On the other hand, suppose, for the purpose of establishing a contradiction that $\tilde{r}_k = 0$, that $h_{k+1,k} \neq 0$. The inexact Arnoldi relation (2.1) implies that $\tilde{r}_k = r_0 - \widetilde{A}_k V_k y_k = V_{k+1}(\beta e_1 - \bar{H}_k y_k)$. Then $\bar{H}_k(2:k+1, 1:k)$ is a square nonsingular upper-triangular matrix. From $\tilde{r}_k = 0$ it follows that $\|\beta e_1 - \bar{H}_k y_k\| = 0$, which in turn implies that $H_k(2:k+1, 1:k) y_k = 0$. Therefore, $y_k = 0$, since $\bar{H}_k(2:k+1, 1:k)$ is nonsingular. Now $\|\tilde{r}_k\| = \|\beta e_1 - \bar{H}_k y_k\| = \beta$, which contradicts $\tilde{r}_k = 0$.

2. The orthogonality of the V_k's and the relaxed Arnoldi relation read $V_{k+1}\bar{H}_k = (A + G_k V_k^T)V_k$, from which it follows, using the variational characterization of singular values, that

$$\sigma_{\min}(A + G_k V_k^T) \leq \sigma_{\min}(\bar{H}_k).$$

Assuming that there exists a constant c, $0 < c < 1$, such that

$$\|E_k\| \leq c\frac{\sigma_{\min}(A)}{n}, \tag{2.6}$$

we get

$$\|G_k V_k^T\| = \left\| \sum_{i=1}^{k} E_i v_i v_i^T \right\| \leq \sum_{i=1}^{k} \|E_i v_i v_i^T\| \leq c\sigma_{\min}(A), \tag{2.7}$$

and inequality $\sigma_{\min}(A) - \|G_k V_k^T\| \leq \sigma_{\min}(A + G_k V_k^T)$ (see [1, Theorem 3.3.16]) implies that $0 < (1-c)\sigma_{\min}(A) \leq \sigma_{\min}(\bar{H}_k)$. This inequality indicates that $\bar{H}_k$ has full rank. In particular, this is true at breakdown, which shows that H_m is nonsingular. The conclusion follows from 1. ∎

The next theorem gives a sufficient condition for the convergence of relaxed GMRES for η_b.

THEOREM 2 (convergence of relaxed GMRES for η_b). *Let us denote by m the step where the breakdown occurs in the inexact GMRES algorithm. Let c be such that $0 < c < 1$ and let ε_c and ε_g be any positive real numbers. Assume for all $k \leq m$ that*

$$\|E_k\| \leq \frac{1}{n}\sigma_{\min}(A) \min\left(c, \frac{(1-c)\|b\|}{\|\tilde{r}_{k-1}\|}\varepsilon_g\right). \tag{2.8}$$

Then there exists ℓ, $0 < \ell \leq m$, such that the stopping criterion

$$\|\tilde{r}_\ell\| \leq \varepsilon_c \|b\| \tag{2.9}$$

is satisfied and

$$\eta_b(x_\ell) \leq \varepsilon_c + \varepsilon_g.$$

Proof. Inequality (2.2) holds with $\epsilon = \varepsilon_g \|b\|$, and it follows that, for all $k \leq m$, $\frac{\|r_k - \tilde{r}_k\|}{\|b\|} \leq \varepsilon_g$. It follows from the second statement of Theorem 1 that, for some $\ell \leq m$, $\|\tilde{r}_\ell\| \leq \varepsilon_c \|b\|$. The bound on $\eta_b(x_\ell)$ then follows from (2.3). The step ℓ can be, for instance, the step where the (happy) breakdown occurs. ∎

Remark 1. The stopping criterion (2.9) is based on $\|\tilde{r}_k\| = \|\bar{H}_k y_k - \beta e_1\|$ which is a byproduct of the algorithm and does not require any additional matrix-vector product [7]. Similarly the control on $\|E_k\|$ involves only some constants of the problem (n, $\sigma_{\min}(A)$, $\|b\|$) and the byproduct $\|\tilde{r}_k\|$. Finally, the constant c can be any value between zero and one.

We now show a similar result for a prescribed accuracy on the backward error $\eta_{A,b}$. In a first step, using the same technique as for Theorem 2, we obtain in the following lemma a control that ensures the convergence with respect to $\eta_{A,b}$.

LEMMA 1. *Let us denote by m the step where the breakdown occurs in the inexact GMRES algorithm. Let ε_c and ε_g be any positive real numbers. Assume that there exists ℓ, $0 < \ell \leq m$, such that, for all $k \leq \ell$, $\tilde{r}_{k-1} \neq 0$ and*

$$\|E_k\| \leq \frac{\sigma_{\min}(\bar{H}_\ell)}{n} \frac{\|A\|\|x_\ell\| + \|b\|}{\|\tilde{r}_{k-1}\|}\varepsilon_g. \tag{2.10}$$

If $\|\tilde{r}_\ell\| \leq \varepsilon_c \|A\|\|x_\ell\|$ then we have

$$\eta_{A,b}(x_\ell) \leq \varepsilon_c + \varepsilon_g.$$

Proof. Substituting $(\|A\|\|x_\ell\| + \|b\|)\,\varepsilon_g$ for ϵ in (2.2) yields $\frac{\|r_\ell - \tilde{r}_\ell\|}{\|A\|\|x_\ell\| + \|b\|} \leq \varepsilon_g$. The conclusion follows from (2.3) and $\|\tilde{r}_\ell\| \leq \varepsilon_c \|A\|\|x_\ell\| \leq \varepsilon_c\,(\|A\|\|x_\ell\| + \|b\|)$. ∎

The above result is not implementable because of the forward reference to the quantities $\|x_\ell\|$ and $\sigma_{\min}(\bar{H}_\ell)$. The next lemma shows that the control ensuring a happy breakdown $\|\tilde{r}_m\| = 0$, i.e., $\|E_k\| \leq c\frac{\sigma_{\min}(A)}{n}$, also enables us to derive a lower-bound on $\|x_\ell\|$ when $\|\tilde{r}_\ell\| \leq \varepsilon_c \|A\|\|x_\ell\|$ is satisfied. This lower bound is expressed in terms of c, ε_c, and some constants of the problem.

LEMMA 2. *Let us denote by m the step where the breakdown occurs in the inexact GMRES algorithm. Let c, x_0, and $x^\star$ be such that $2c\|x_0\| \leq \|x^\star\|$ and $0 < c < 1$. Assume that there exists $\ell > 0$ such that, for all $k \leq \ell$, $\|E_k\| \leq c\frac{\sigma_{\min}(A)}{n}$ and $\|\tilde{r}_\ell\| \leq \varepsilon_c\|A\|\|x_\ell\|$; then*

$$\frac{\|x^\star\|}{4 + 2\varepsilon_c\kappa(A)} \leq \|x_\ell\|. \tag{2.11}$$

Proof. Starting from the definition of $\tilde{r}_\ell$ and from $x_\ell = x_0 + \delta x_\ell$, we get

$$\begin{aligned}\tilde{r}_\ell &= b - Ax_0 - (A + G_\ell V_\ell^T)\,\delta x_\ell \\ &= A(x^\star - x_\ell) - G_\ell V_\ell^T x_\ell + G_\ell V_\ell^T x_0,\end{aligned}$$

which shows that

$$x^\star - x_\ell = A^{-1}\tilde{r}_\ell + A^{-1}G_\ell V_\ell^T x_\ell - A^{-1}G_\ell V_\ell^T x_0. \tag{2.12}$$

From the assumptions $\|E_\ell\| \leq c\frac{\sigma_{\min}(A)}{n}$ and $0 < c < 1$ it follows as in (2.7) that $\|G_\ell V_\ell^T\| \leq c\sigma_{\min}(A)$. Using $\|\tilde{r}_\ell\| \leq \varepsilon_c\|A\|\|x_\ell\|$ and $\|A^{-1}\|\sigma_{\min}(A) = 1$ and taking norms in (2.12) yield

$$\|x^\star - x_\ell\| \leq (c + \varepsilon_c\kappa(A))\|x_\ell\| + c\|x_0\|.$$

Since $\|x^\star\| - \|x_\ell\| \leq \|x^\star - x_\ell\|$, we get

$$\|x^\star\| - c\|x_0\| \leq (1 + c + \varepsilon_c\kappa(A))\|x_\ell\| \leq (2 + \varepsilon_c\kappa(A))\|x_\ell\|,$$

where we have used that $c < 1$. The conclusion follows from the assumption $2c\|x_0\| \leq \|x^\star\|$ that implies $\|x^\star\| - c\|x_0\| \geq \|x^\star\|/2$. $\square$

We note that the assumption $2c\|x_0\| \leq \|x^\star\|$ is not very stringent as it is satisfied for $x_0 = 0$. The next theorem gives a sufficient convergence condition for the stopping criterion $\eta_{A,b}$. Similarly as for Theorem 2 (see Remark 1) the control on $\|E_k\|$ and the stopping criterion involved in relaxed GMRES are based on byproducts of the algorithm and some constants of the problem. Provided that the above constants are available or easily estimated, this result can be used in a practical implementation.

THEOREM 3 (convergence of relaxed GMRES for $\eta_{A,b}$). *Let us denote by m the step where the breakdown occurs in the inexact GMRES algorithm. Let c, x_0, and $x^\star$ be such that $2c\|x_0\| \leq \|x^\star\|$ and $0 < c < 1$. Let ε_c and ε_g be any positive real numbers. Suppose that for all $k \leq m$*

$$\|E_k\| \leq \frac{1}{n}\sigma_{\min}(A)\min\left(c, (1-c)\frac{\gamma^\star}{\|\tilde{r}_{k-1}\|}\varepsilon_g\right), \tag{2.13}$$

where $\gamma^\star = \frac{1}{4+2\varepsilon_c\kappa(A)}\|A\|\|x^\star\| + \|b\|$. There exists ℓ, $\ell \leq m$, such that the stopping criterion

$$\|\tilde{r}_\ell\| \leq \varepsilon_c\|A\|\|x_\ell\| \tag{2.14}$$

is satisfied and

$$\eta_{A,b}(x_\ell) \leq \varepsilon_c + \varepsilon_g.$$

Proof. We first observe that $\|E_k\| \leq \frac{c}{n}\sigma_{\min}(A)$ implies (see Theorem 1) that a *happy* breakdown will necessarily occur; therefore, for some ℓ, $\ell \leq m$, $\frac{\|\tilde{r}_\ell\|}{\|A\|\|x_\ell\|} \leq \varepsilon_c$. Furthermore, from (2.13) it follows that

$$\|E_k\| \leq \left(\frac{(1-c)\sigma_{\min}(A)}{n}\right)\left(\frac{1}{4+2\varepsilon_c\kappa(A)}\|A\|\|x^\star\| + \|b\|\right)\frac{\varepsilon_g}{\|\tilde{r}_{k-1}\|}. \tag{2.15}$$

The first factor in the right-hand side of (2.15) can be bounded above by $\frac{\sigma_{\min}(\bar{H}_\ell)}{n}$, because $\|E_k\| \leq \frac{c}{n}\sigma_{\min}(A)$ and the second statement of Theorem 1 applies. Using the conclusions of Lemma 2, the second factor is bounded above by $\|A\|\|x_\ell\| + \|b\|$ because of (2.11). Therefore, we have $\|E_k\| \leq \frac{\sigma_{\min}(\bar{H}_\ell)}{n}\frac{\|A\|\|x_\ell\|+\|b\|}{\|\tilde{r}_{k-1}\|}\varepsilon_g$. Consequently the conditions of Lemma 1 hold, which concludes the proof. ∎

Remark 2. As illustrated in the forthcoming numerical experiments, it might be noticed that at the drawback of smaller perturbations one can derive a control that does not require the knowledge of $\|x^\star\|$ (or of a lower bound). If we replace $\gamma^\star$ by $\gamma^b = \|b\| < \gamma^\star$ in Theorem 3, the theorem is still valid.

So far, similarly to [15, 19], we have considered that the initial residual involved in the inexact algorithm is computed exactly: $r_0 = b - Ax_0$. We now provide an extension to these results which additionally accounts for inaccuracies in the very first matrix-vector product Ax_0. Such a result is crucial when only approximations of the matrix-vector product are available. This is the case, for instance, when A is a Schur complement matrix in domain decomposition where the local systems are solved iteratively [3].

Theorem 4 (convergence of relaxed GMRES with inexact initial residual). *Let us denote by m the step where the breakdown occurs in the inexact GMRES algorithm. Let ε_A and ε_b be any positive real numbers. Suppose that the initial residual r_0 is approximated with $\tilde{r}_0 = b+\Delta b-(A+\Delta A)x_0$ with $\|\Delta A\| \leq \varepsilon_A\|A\|$ and $\|\Delta b\| \leq \varepsilon_b\|b\|$. Suppose that the assumptions of Theorem 3 hold. There exists ℓ, $\ell \leq m$, such that the stopping criterion (2.14) is satisfied and*

$$\eta_{A,b}(x_\ell) \leq \varepsilon_c + \varepsilon_g + (1+\varepsilon_c+\varepsilon_g)\left(\varepsilon_A\frac{\|x_0\|}{\|x_\ell\|} + \varepsilon_b\right).$$

Proof. We consider the following two algorithms:

- (A_1) is inexact GMRES with the approximated residual $\tilde{r}_0$ as in the theorem's assumption and with initial guess x_0.
- (A_2) is inexact GMRES without any initial perturbations on A or b for the solution of $Ax = b+\Delta b-\Delta Ax_0 = \tilde{b}$ with same initial guess x_0.

If the same perturbations E_k are used for both algorithms (A_1) and (A_2), these two algorithms generate the same matrices V_k and H_k. This follows from writing $\tilde{r}_0 = (b+\Delta b)-(A+\Delta A)x_0 = (b+\Delta b-\Delta Ax_0)- Ax_0$ (which shows that both algorithms share the same initial residual norm β) and from the fact that the perturbations E_k are the same.

Using Theorem 3 on (A_2) leads to $\frac{\|Ax_\ell-\tilde{b}\|}{\|A\|\|x_\ell\|+\|\tilde{b}\|} \le \varepsilon_c+\varepsilon_g$. Furthermore, we have

$$\begin{aligned}
\frac{\|Ax_\ell - b\|}{\|A\|\|x_\ell\|+\|b\|} &\le \frac{\|Ax_\ell-\tilde{b}\|}{\|A\|\|x_\ell\|+\|b\|} + \frac{\|\Delta b - \Delta A x_0\|}{\|A\|\|x_\ell\|+\|b\|} \\
&\le \frac{\|Ax_\ell-\tilde{b}\|}{\|A\|\|x_\ell\|+\|\tilde{b}\|}\frac{\|A\|\|x_\ell\|+\|\tilde{b}\|}{\|A\|\|x_\ell\|+\|b\|} + \frac{\|\Delta b - \Delta A x_0\|}{\|A\|\|x_\ell\|+\|b\|} \\
&\le (\varepsilon_c+\varepsilon_g)\left(1+\frac{\|\Delta b - \Delta A x_0\|}{\|A\|\|x_\ell\|+\|b\|}\right) + \frac{\|\Delta b - \Delta A x_0\|}{\|A\|\|x_\ell\|+\|b\|} \\
&\le (\varepsilon_c+\varepsilon_g) + (1+\varepsilon_c+\varepsilon_g)\frac{\|\Delta b - \Delta A x_0\|}{\|A\|\|x_\ell\|+\|b\|}.
\end{aligned}$$

The fact that

$$\frac{\|\Delta b - \Delta A x_0\|}{\|A\|\|x_\ell\|+\|b\|} \le \frac{\|\Delta b\| + \|\Delta A\|\|x_0\|}{\|A\|\|x_\ell\|+\|b\|} \le \frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|}\frac{\|x_0\|}{\|x_\ell\|}$$

concludes the proof. □

2.3. Numerical experiments. In this section we report on numerical experiments performed with MATLAB and machine precision $\psi \sim 2.10^{-16}$. They illustrate that the theoretical results established in exact arithmetic in the previous section do ensure the convergence of the relaxed GMRES even in the presence of round-off errors. For those experiments we consider only the strategies related to the convergence with respect to $\eta_{A,b}(x_k)$. Consequently the stopping criterion implemented by the algorithm is defined by (2.14). For those experiments we set x_0, ε_A, and ε_b to zero. The matrices E_k are obtained using random matrices (MATLAB function `rand`) multiplied by a relevant scalar to match the upper bound of the corresponding strategies. We set $c = 1/4$, $\varepsilon_c = \varepsilon_g = \varepsilon/2$. We therefore consider the two strategies defined as follows.

1. Strategy $S^\star$:

$$\|E_k\| = \frac{\sigma_{\min}(A)}{4n}\min\left(1, \frac{3\gamma^\star}{2\|\tilde{r}_{k-1}\|}\varepsilon\right).$$

2. Strategy S^b:

$$\|E_k\| = \frac{\sigma_{\min}(A)}{4n}\min\left(1, \frac{3\gamma^b}{2\|\tilde{r}_{k-1}\|}\varepsilon\right).$$

Note that S^b is derived from Remark 2. The strategy S^b is interesting, because it does not require the knowledge of $\|x^\star\|$, but it enables only perturbations smaller than $S^\star$. The two strategies rely on the knowledge of $\sigma_{\min}(A)$, and $S^\star$ relies on the knowledge of the solution norm $x^\star$. In our experiments, we set $b = Ax^\star$, so that $x^\star$ is known, and $\sigma_{\min}(A)$ is computed using the `svd` MATLAB command. Further research is needed to establish similar convergence results while replacing these quantities by others that are simpler to estimate. Finally, for the sake of completeness we depict in Algorithm 1 the details of the implementation of the relaxed GMRES algorithm.

Typical behaviors are presented for the linear system $Ax = b$ in Figure 1, where A is the matrix `PDE225` preconditioned using the incomplete LU factorization (i.e., ILU(t) [14]) with threshold $t = 10^{-1}$ (i.e., we run the algorithms on

Algorithm 1. Relaxed GMRES with strategy S.

Choose a convergence threshold $\varepsilon = \varepsilon_c + \varepsilon_g$
Choose an initial guess x_0
$r_0 = b - Ax_0$; $\beta = \|r_0\|$
$v_1 = r_0/\|r_0\|$;
for $k = 1, 2, \ldots$ **do**
  $z = (A + E_k)v_k$, E_k being such that strategy S holds
  for $i = 1$ `to` k **do**
    $h_{i,k} = v_i^T z$
    $z = z - h_{i,k}v_i$
  end for
  $h_{k+1,k} = \|z\|$
  $v_{k+1} = z/h_{k+1,k}$
  Solve the least-squares problem $\min \|\beta e_1 - \bar{H}_k y\|$ for y_k
  if $\|\tilde{r}_k\| = \|\beta e_1 - \bar{H}_k y_k\| \leq \varepsilon_c \|A\| \|x_k\|$ **then**
    Set $x_k = x_0 + V_k y_k$
    Exit
  end if
end for

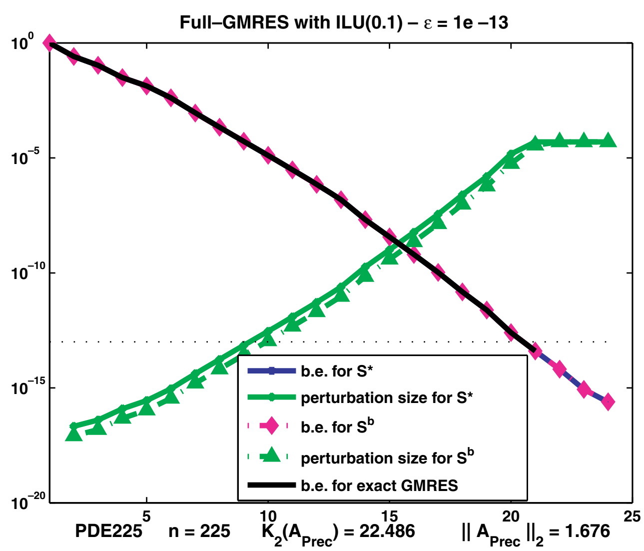

FIG. 1. *Relaxed GMRES with strategy $S^\star$ and S^b*, PDE225, $\varepsilon = 10^{-13}$.

$A = U^{-1}L^{-1}A_{PDE225}$). The right-hand side b is such that $x^\star = (1, \ldots, 1)^T$ is the solution of $Ax = b$. In this figure, we plot the convergence history, that is, the backward errors $\eta_{A,b}(x_k)$ along the iterations. The line without any tip is the convergence curve with the exact full GMRES; the line with $\diamond$ (resp., $+$) is the convergence of relaxed GMRES with strategy $S^\star$ (resp., S^b). Notice that the three curves perfectly overlap. The dashed horizontal line corresponds to the targeted backward error $\varepsilon = 10^{-13}$, and the line with $\triangle$ (resp., $*$) is the relative norm of the perturbation $\frac{\|E_k\|}{\|A\|}$ associated

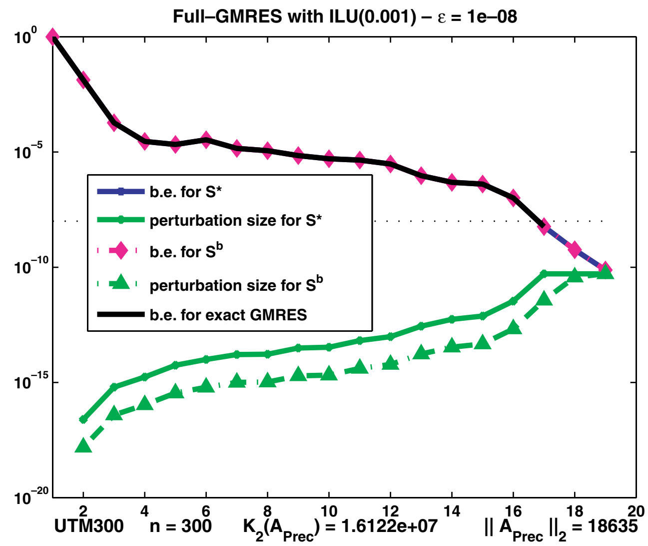

FIG. 2. *Relaxed GMRES with strategy* $S^\star$ *and* S^b, `UTM300`, $\varepsilon = 10^{-8}$.

with strategy $S^\star$ (resp., S^b). It can be seen that as relaxed GMRES converges the accuracy of the matrix-vector product is significantly relaxed without preventing relaxed GMRES from converging to the targeted backward error. We also see that one extra relaxed GMRES iteration is performed before the convergence is detected while $\eta_{A,b}(x_k)$ is already below ε. This delay is due to the stopping criterion implemented in the inexact algorithm that relies on an upper bound for $\eta_{A,b}$ based on (2.3).

In Figure 2, we display the same results for the matrix `UTM 300` preconditioned by ILU(10^{-3}). Again it can be seen that the accuracy of the matrix-vector product can be relaxed without preventing the convergence to occur. However, in that sort of extreme case, the perturbation size is smaller than ψ in the first iterations. This does not make much sense in finite precision calculation because the perturbation below ψ cannot be represented in finite precision. Finally, the example with matrix `UTM 300` better illustrates that S^b is more conservative than $S^\star$.

2.4. Design and behavior of some relaxation heuristics. It is established in [6] that the classical (unperturbed) GMRES algorithm implemented using reliable orthogonal transformation is a backward stable method in finite precision. This means that the quantity $\eta_{A,b}(x_k)$ is of the order of machine precision for some step ℓ, $\ell \leq n$. Let us use this result for a machine precision $\psi = \varepsilon$. Provided that all the operations occurring in the GMRES algorithm are performed with machine precision ε, a backward error $\eta_{A,b}(x_k) \sim C\,\varepsilon$ can be reached, where C depends on the problem size and on the details of the arithmetic. In that context we can define three heuristics that are closely related to the strategies studied in the previous section. They are simply derived by thresholding the norm of the perturbations and preventing them from becoming smaller than ε.

This leads to the definition of the following three heuristics.

1. Heuristic $S^\star(\varepsilon)$:

$$\|E_k\| = \max\left(\varepsilon\|A\|, \frac{\sigma_{\min}(A)}{4n}\min\left(1, \frac{3\gamma^\star}{\|\tilde{r}_{k-1}\|}\varepsilon_g\right)\right).$$

TABLE 2.1
Number of iterations of GMRES with various strategies.

					Heuristics		
Matrix	n	t	ε	$N_{\rm ex}$	N_ε	$N_\varepsilon^\star$	N_ε^b
e05r0400	236	10^{-3}	10^{-14}	21	21	21	21
e05r0000	236	10^{-2}	10^{-06}	25	25	26	26
GRE115	115	10^{-1}	10^{-10}	15	15	15	15
GRE185	185	10^{-2}	10^{-14}	21	21	21	21
GRE343	343	10^{-1}	10^{-10}	29	29	30	30
CAVITY03	317	10^{-3}	10^{-10}	18	18	18	18
PDE225	225	10^{-1}	10^{-13}	21	21	22	22
SAYLR1	238	10^{-1}	10^{-11}	29	29	30	30
UTM300	300	10^{-3}	10^{-08}	17	17	18	18
WEST0381	381	10^{-2}	10^{-06}	12	12	13	13
BFW398A	398	10^{-1}	10^{-08}	40	40	41	41

2. Heuristic $S^b(\varepsilon)$:

$$\|E_k\| = \max\left(\varepsilon\|A\|, \frac{\sigma_{\min}(A)}{4n}\min\left(1, \frac{3\gamma^b}{\|\tilde{r}_{k-1}\|}\varepsilon_g\right)\right).$$

3. Heuristic $S(\varepsilon)$:

$$\|E_k\| = \varepsilon\|A\|.$$

The latter heuristic is related to exact GMRES run in a floating point arithmetic with machine precision ε.

In Table 2.1 we report on the number of iterations of exact GMRES and relaxed GMRES with the three heuristics on a set of matrices from Matrix Market. In that table we display the size of the matrix n, the threshold used for the $ILU(t)$ preconditioner, and the targeted accuracy ε for $\eta_{A,b}$; the number of exact GMRES iterations is $N_{\rm ex}$, and the number of relaxed GMRES iterations with heuristic $S(\varepsilon)$ ($S^\star(\varepsilon)$ and $S^b(\varepsilon)$) is N_ε (resp., $N_\varepsilon^\star$ and N_ε^b). For the inexact GMRES the stopping criterion is defined by (2.14). The right-hand side b is such that $x^\star = (1, \ldots, 1)^T$ is the solution of $Ax = b$, where A stands again for the left-preconditioned matrix. We first see that the three variants of inexact GMRES always converge to the targeted accuracy. Even though not displayed for those examples the convergence histories with the three heuristics perfectly overlap the one of exact GMRES along the iterations for all the matrices. We observed this behavior for all the right-hand sides we have considered—not only for $b = A(1, \ldots, 1)^T$. Thus the reason the number of iterations for $S(\varepsilon)$ is sometimes smaller than for $S^\star(\varepsilon)$ and $S^b(\varepsilon)$ solely resides in the stopping criterion which induces a short delay in the convergence detection.

The only example we have encountered that behaves differently is the GRCAR matrix with $b = e_1$ and $\varepsilon \geq 10^{-8}$ (also considered in [15]); in Figure 3 we display for that example the convergence history of $\eta_{A,b}(x_k)$. We see in this peculiar case that exact GMRES converges quickly; that is, the solution lies in a low dimensional Krylov subspace. If inexact matrix-vector products are used, this low dimensional invariant space is not captured as quickly and the convergence is significantly delayed. We mention that this behavior disappears if a smaller value of ε (a larger Krylov space is required) is selected but still exists for larger ε (even smaller "invariant" space). It also disappears for other right-hand sides such as $b = A(1, \ldots, 1)^T$.

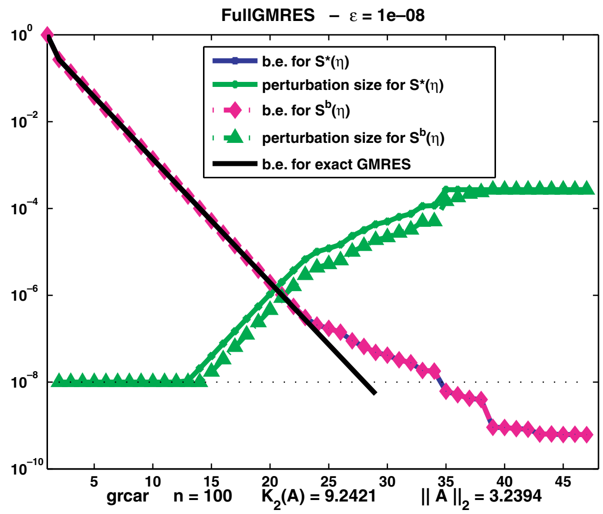

FIG. 3. *Convergence history of the inexact GMRES on* GRCAR *with* $b = e_1$.

3. GMRES with relaxed right-preconditioning. In practice GMRES is very often used with a preconditioner. When the preconditioner is used on the right, we solve the linear system $AM^{-1}u = b$ and we recover the solution by $x^\star = M^{-1}u^\star$. An important class of preconditioners, often referred to as implicit preconditioners, tries to approximate the matrix A and then solves the linear system associated with M (M is chosen such that this operation is much easier with M than with A). We are interested in the case where the application of the preconditioner is inexact: the system $Mz = v$ involved in the preconditioned GMRES is not solved exactly. This happens, for instance, when block preconditioners are used. Examples are Schwarz preconditioners in domain decomposition [17] or, more simply, block Jacobi preconditioners where each block is solved approximately using an iterative scheme. In this case, we assume that the matrix-vector product by A is performed exactly, but we have $Mz = v + p$, where p is the residual associated with the preconditioning operation. Our aim is to find a strategy to monitor $\|p\|$ in such a way that the convergence of GMRES reaches a backward error ε. A related study is developed in [15], where the focus is on the residual norm and the issue related to the recovery of x_ℓ from u_ℓ by the preconditioning $Mx_\ell = u_\ell + p_\ell$ is not addressed.

The inexact Arnoldi relation now reads

$$AM^{-1}V_k + AM^{-1}P_k = (AM^{-1} + AM^{-1}P_kV_k^T)V_k = V_{k+1}\bar{H}_k, \tag{3.1}$$

where $P_k = [p_1, \ldots, p_k]$, p_k being the residual associated with the kth preconditioning operation $Mz_k = v_k + p_k$. We present here results obtained for an unperturbed initial residual r_0. These results extend easily to cover the case where r_0 is perturbed, at the cost of much heavier notation as in Theorem 4. The following theorem, which is a direct application of Theorem 3 on the linear system $AM^{-1}u = b$, shows a possible control of $\|p_k\|$ so that the convergence of the relaxed right-preconditioned GMRES occurs.

THEOREM 5 (convergence of right-preconditioned GMRES for $\eta_{AM^{-1},b}$). *Let us denote by m the step where the breakdown occurs in the inexact right-preconditioned GMRES algorithm. Let c, u_0, and $u^\star$ be such that $2c\|u_0\| \le \|u^\star\|$ and $0 < c < 1$. Let*

ε_c and ε_g be any positive real numbers. Suppose that for all $k \leq m$

$$\|p_k\| \leq \frac{1}{n\kappa(AM^{-1})} \min\left(c, \frac{(1-c)\gamma^p}{\|\tilde{r}_{k-1}\|}\varepsilon_g\right), \tag{3.2}$$

where $\gamma^p = \frac{1}{4+2\varepsilon_c\kappa(AM^{-1})}\|AM^{-1}\|\|u^\star\| + \|b\|$.

There exists ℓ, $\ell \leq m$, such that the following stopping criterion is satisfied:

$$\frac{\|\tilde{r}_\ell\|}{\|AM^{-1}\|\|u_\ell\|} \leq \varepsilon_c \quad \textit{and} \quad \eta_{AM^{-1},b}(u_\ell) = \frac{\|b - AM^{-1}u_l\|}{\|AM^{-1}\|\|u_\ell\| + \|b\|} \leq \varepsilon_u = \varepsilon_c + \varepsilon_g.$$

Proof. Setting $E_k = AM^{-1}p_k v_k^T$, the inexact Arnoldi relation (3.1) is exactly (2.1). The relaxation strategy of Theorem 3 writes

$$\|AM^{-1}p_k\| \leq \min\left(c\frac{\sigma_{\min}(AM^{-1})}{n}, \frac{(1-c)\sigma_{\min}(AM^{-1})}{n}\frac{\gamma^p}{\|\tilde{r}_{k-1}\|}\varepsilon_g\right).$$

The conclusion follows from $\|AM^{-1}p_k\| \leq \|AM^{-1}\|\|p_k\|$. □

The result of Theorem 5 gives only a sufficient condition such that backward error associated with the approximated solution u_ℓ of the *preconditioned* system $AM^{-1}u = b$ is smaller than ε. To get the solution to the original system $Ax = b$ an additional preconditioning operation $Mx_\ell = u_\ell + p$ has to be performed, where p denotes the associated residual and $\rho = \|p\|$. Let us compute an upper bound for the backward error on the original system. We set $r_\ell = b - AM^{-1}u_\ell$. From $r_\ell = AM^{-1}(u^\star - u_\ell)$ it follows that

$$\begin{aligned} \|u_\ell\| - \|u^\star\| \leq \|u_\ell - u^\star\| &\leq \|MA^{-1}\|\|r_\ell\| \\ &\leq \varepsilon_u\|MA^{-1}\|(\|AM^{-1}\|\|u_\ell\| + \|b\|) \\ &\leq \varepsilon_u\kappa(AM^{-1})(\|u_\ell\| + \|u^\star\|), \end{aligned}$$

where we have used $\|b\| = \|AM^{-1}u^\star\| \leq \|AM^{-1}\|\|u^\star\|$. This shows that

$$\|u_\ell\| \leq \frac{1 + \varepsilon_u\kappa(AM^{-1})}{1 - \varepsilon_u\kappa(AM^{-1})}\|u^\star\|,$$

provided that $\kappa(AM^{-1})\varepsilon_u < 1$. Using this bound leads to

$$\begin{aligned} \|b - Ax_\ell\| = \|r_\ell + AM^{-1}p\| &\leq \varepsilon_u\|AM^{-1}\|\frac{1 + \varepsilon_u\kappa(AM^{-1})}{1 - \varepsilon_u\kappa(AM^{-1})}\|u^\star\| + \varepsilon_u\|b\| + \|AM^{-1}\|\rho \\ &\leq 2\varepsilon_u\|AM^{-1}\|\frac{1 + \varepsilon_u\kappa(AM^{-1})}{1 - \varepsilon_u\kappa(AM^{-1})}\|u^\star\| + \|AM^{-1}\|\rho, \end{aligned} \tag{3.3}$$

where we used $\|b\| \leq \|AM^{-1}\|\|u^\star\|$ and $\frac{1+\varepsilon_u\kappa(AM^{-1})}{1-\varepsilon_u\kappa(AM^{-1})} \geq 1$. We summarize this result in the next theorem.

THEOREM 6. *Suppose that the relaxed right-preconditioned GMRES is run on $AM^{-1}u = b$ under the assumptions of Theorem 5 and that u_ℓ is the corresponding estimate of $u^\star$. Suppose in addition that $Mx_\ell = u_\ell + p$, with $\rho = \|p\|$, and that $\kappa(AM^{-1})\varepsilon_u < 1$. The backward error $\eta_{A,b}(x_\ell)$ of x_ℓ considered as a solution of $Ax = b$ satisfies*

$$\eta_{A,b}(x_\ell) \leq \frac{1}{\|A\|\|x_\ell\| + \|b\|}\left(2\varepsilon_u\|AM^{-1}\|\frac{1 + \varepsilon_u\kappa(AM^{-1})}{1 - \varepsilon_u\kappa(AM^{-1})}\|u^\star\| + \|AM^{-1}\|\rho\right).$$

TABLE 3.1
GMRES with relaxed right preconditioner.

ρ	$\eta_{A,b}(x_\ell)$	Upper bound
10^{-14}	$4 \cdot 10^{-14}$	$6 \cdot 10^{-14}$
10^{-11}	$3 \cdot 10^{-12}$	$1 \cdot 10^{-11}$
10^{-08}	$3 \cdot 10^{-09}$	$1 \cdot 10^{-08}$
10^{-05}	$3 \cdot 10^{-06}$	$1 \cdot 10^{-05}$
10^{-01}	$7 \cdot 10^{-04}$	$3 \cdot 10^{-03}$

We present a numerical illustration for Theorem 6 in order to demonstrate the effect of the last preconditioning operation on the backward error of the computed solution x_ℓ. We consider the matrix E05R0000 with an $ILU(10^{-2})$ as right-preconditioner and a targeted backward error on the preconditioned system $\eta_{AM^{-1},b}(u_\ell) \leq 10^{-10}$. To recover the unpreconditioned solution we perform a final preconditioning step $Mx_\ell = u_\ell + p$, where a random vector p is chosen such that $\|p\|$ is equal to a prescribed quantity ρ. We report in Table 3.1 the backward errors $\eta_{A,b}(x_\ell)$ obtained at the end of the process (inexact right-preconditioned GMRES and the final preconditioned step $Mx_\ell = u_\ell + p$) and the associated upper bound presented in Theorem 6. We see that the upper bound is tight on this example and that the backward error on x_ℓ strongly depends on the accuracy of the final $Mx_\ell = u_\ell + p$ as monitored by the quantity $\rho = \|p\|$.

Remark 3. We mention that similar results can be derived for inexact left-preconditioning. In that context, the quantity ε_b from Theorem 4 enables us to account for the inaccuracy in the first preconditioning step to build the right-hand side of the preconditioned system.

Remark 4. We consider FGMRES with a varying preconditioner satisfying $Mz_k = v+p_k$, where p_k is the residual vector whose norm is used to control the convergence of the algorithm. FGMRES is very close to the inexact right-preconditioned GMRES in the sense that the same Arnoldi relation (3.1) holds, which can be rewritten $AZ_k = V_{k+1}\bar{H}_k$. Contrary to the inexact GMRES framework, [14, Proposition 9.3] shows that when a breakdown occurs at step m in FGMRES and if the associated H_m is nonsingular, not only is $\tilde{r}_m$ zero but also $x_m = Z_\ell y_m = A^{-1}b$. In other words, the nonsingularity of H_m yields that, at the breakdown, x_m is the solution to $Ax = b$. Theorem 1 and (3.2) show that the nonsingularity of H_m is guaranteed if $\|p_k\| \leq \frac{c}{n\kappa(AM^{-1})}$.

Theorem 6 mainly emphasizes the role played by the last preconditioning step. Its use in real applications would deserve additional efforts to find reasonable estimates of the quantities involved in the control such that $\kappa(AM^{-1})$ or $\|AM^{-1}\|$.

4. Some remarks on inexact restarted GMRES. In this paper we are interested in relaxation techniques for which theoretical convergence of GMRES can be established. This is the main reason why we have considered only strategies in the context of full GMRES. There is a clear interest in practice to use the restarted GMRES(m) algorithm [12]. Unfortunately, no general theoretical result exists on the convergence of restarted GMRES. Establishing such a result is probably the first step before studying an inexact restarted algorithm. Nevertheless, we mention that some heuristic strategies [4, 8, 16] as well as those described in this paper enable the convergence of the inexact restarted algorithm in practice.

As previously in this paper we focus on provable convergence theory. In that respect, we consider a variant of restarted GMRES where the restart is not governed by the dimension of the Krylov space affordable in term of memory but rather by a targeted backward error η. We study this variant of restarted GMRES as a fixed point iteration scheme, where at each restart (i.e., each fixed point iteration) the residual can be inaccurately computed. This fixed point iteration is summarized in Algorithm 2, where the function *solve* involved in step 5 is any function ensuring that $\eta_{A,r_k}(s_k) \leq \eta$. If *solve* is exact full GMRES we end up with what we denote by GMRES(η). Notice, that Algorithm 2 differs from classical GMRES(m) in that the calculation of s_k is stopped when the associated backward error is less than η rather than on a memory limit. We define the normwise condition number of the linear system [10, p. 121] by $\kappa(A,b) = \frac{\|A^{-1}\|}{\|x^\star\|}(\|A\|\|x^\star\| + \|b\|)$.

Algorithm 2. Fixed point scheme.

1: Let ε_A, ε_b, and η be positive.
2: Choose an initial guess x_0
3: **for** $k = 1, 2, \ldots$ **do**
4: Compute $r_k = (b + \Delta b_k^{res}) - (A + \Delta A_k^{res})x_k$
5: $s_k = solve(A, r_k, \eta)$
6: $x_{k+1} = x_k + s_k$
7: **end for**

where s_k is such that $\eta_{A,r_k} \leq \eta$ with $\|\Delta A_k^{res}\| \leq \varepsilon_A\|A\|$ and $\|\Delta b_k^{res}\| \leq \varepsilon_b\|b\|$.

THEOREM 7. *We consider a sequence $\{x_k\}$ generated using Algorithm 2 and assume that $\varepsilon_A\kappa(A) < 1/7$ and that $\eta\kappa(A) < 1/7$. If $\varepsilon_A = \varepsilon_b = 0$, $\lim_{k\to+\infty} x_k = x^\star$. If $\max(\varepsilon_A, \varepsilon_b) > 0$, we have*

$$\frac{\|x_k - x^\star\|}{\|x^\star\|} \leq 5\kappa(A,b)\max(\varepsilon_A, \varepsilon_b)$$

for k large enough.

In particular, these results hold for the update s_k computed with relaxed GMRES of section 2.2.

Proof. Let us consider the fixed point iteration

$$x_{k+1} = x_k + s_k.$$

Since $s_k = solve(A, r_k, \eta)$, there exists q_k such that

$$As_k = r_k + q_k, \text{ where } \|q_k\| \leq \eta\left(\|A\|\|s_k\| + \|r_k\|\right). \tag{4.1}$$

We set $e_k = x_k - x^\star$, which yields $e_{k+1} = e_k + s_k$. Taking norms in $r_k = b + \Delta b_k^{res} - (A + \Delta A_k^{res})x_k$ leads to

$$\begin{aligned}
\|r_k\| &= \|b + \Delta b_k^{res} - (A + \Delta A_k^{res})(x^\star + e_k)\| \\
&= \|\Delta b_k^{res} - \Delta A_k^{res}x^\star - (A + \Delta A_k^{res})e_k\| \\
&\leq \max(\varepsilon_b, \varepsilon_A)(\|b\| + \|A\|\|x^\star\|) + (1 + \varepsilon_A)\|A\|\|e_k\|
\end{aligned}$$

and to

$$\|q_k\| \leq \eta \max(\varepsilon_b, \varepsilon_A)(\|b\| + \|A\|\|x^\star\|) + \eta(2+\varepsilon_A)\|A\|\|e_k\| + \eta\|A\|\|e_{k+1}\|. \tag{4.2}$$

Using $e_{k+1} = e_k + s_k$, multiplying (4.1) by A^{-1}, and using the definition of r_k yield

$$\begin{aligned}\|e_{k+1}\| &\leq \|e_k + A^{-1}b + A^{-1}\Delta b_k^{res} - A^{-1}(A + \Delta A_k^{res})(x^\star + e_k)\| + \|A^{-1}\|\|q_k\| \\ &= \|A^{-1}\Delta b_k^{res} - A^{-1}\Delta A_k^{res}(x^\star + e_k)\| + \|A^{-1}\|\|q_k\| \\ &\leq \max(\varepsilon_b, \varepsilon_A)\|A^{-1}\|(\|b\| + \|A\|\|x^\star\|) + \varepsilon_A\kappa(A)\|e_k\| + \|A^{-1}\|\|q_k\|.\end{aligned}$$

Under the assumption $\eta\kappa(A) < 1/7$, we have $(1 - \eta\kappa(A)) > 0$ and inequality (4.2) yields

$$\|e_{k+1}\| \leq \frac{(1+\eta)\max(\varepsilon_b, \varepsilon_A)}{1 - \eta\kappa(A)}\|A^{-1}\|(\|b\| + \|A\|\|x^\star\|) + \frac{2\eta + \varepsilon_A + \eta\varepsilon_A}{1 - \eta\kappa(A)}\kappa(A)\|e_k\|.$$

Suppose that $\eta\kappa(A) < 1/7$ and $\varepsilon_A\kappa(A) < 1/7$; then

$$\|e_{k+1}\| \leq 4/3 \max(\varepsilon_b, \varepsilon_A)\kappa(A,b)\|x^\star\| + 2/3\|e_k\|.$$

If $\varepsilon_A = \varepsilon_b = 0$, $\lim_{k\to+\infty} x_k = x^\star$; else we have

$$\limsup_{k\to\infty} \|e_k\| \leq 4\max(\varepsilon_b, \varepsilon_A)\kappa(A,b)\|x^\star\|,$$

which concludes the proof. □

The above theorem shows that for a large enough k the forward error is bounded by five times the product of a backward error by the normwise condition number $\kappa(A,b)$. This is reasonable from a perturbation theory point of view.

The calculation of s_k by the function *solve* in Algorithm 2 can be implemented by relaxed GMRES. This latter variant is referred to as relaxed GMRES(η). In Figure 4 we display the convergence history of relaxed GMRES(η) for $\eta = 10^{-4}$ and for two different values of ε_A and ε_b. The curves with $\times$ represent the norm of the perturbations involved in the relaxed GMRES. The curves with $+$ represent the backward error $\eta_{A,b}(x_k)$ at each iteration of relaxed GMRES that implements *solve*. In the experiments we did not implement any stopping criterion; we display the first seven outer iterations that result in 45 iterations of relaxed GMRES. It can be observed that the final backward error reached by the algorithm is related to the perturbation size $\max\{\varepsilon_A, \varepsilon_b\}$ imposed on the residual r_k of step 4 of Algorithm 2. Remarkably the perturbations involved in relaxed GMRES of step 5 are significantly larger than $\max\{\varepsilon_A, \varepsilon_b\}$. The spikes observed in the perturbation size curves correspond to the restarts in relaxed GMRES(η).

5. Conclusion. In this paper we are interested either in relaxation techniques for which theoretical convergence of relaxed GMRES can be established or in heuristics that are closely related to the backward stability of the GMRES algorithm with reliable orthogonalization schemes. The proposed strategies ensure the convergence in backward error $\eta_{A,b}$ or η_b down to a prescribed accuracy ε. Finally, our strategies rely on the knowledge of $\sigma_{\min}(A)$, the solution norm that might be difficult to estimate. Further research is needed to establish similar convergence results while replacing these quantities by others that are simpler to estimate.

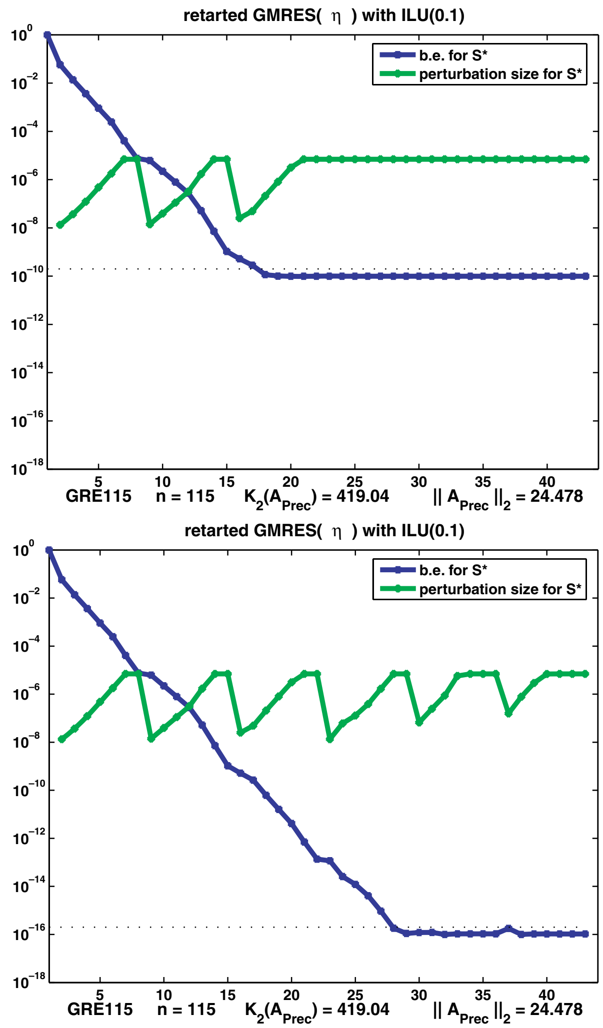

FIG. 4. *Restarted GMRES(ε) with inexact matrix-vector products,* `GRE115`*, with $\varepsilon_A = \varepsilon_b = 10^{-10}$ (top), and $\varepsilon_A = \varepsilon_b = 10^{-16}$ (bottom).*

Acknowledgments. We would like to thanks the anonymous referees for their valuable remarks, questions, and comments that enabled us to substantially improve this paper. We also express our gratitude to Anne Greenbaum for her continuous encouragement.

REFERENCES

[1] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1995.

[2] R. BARRETT, M. BERRY, T. F. CHAN, J. DEMMEL, J. DONATO, J. DONGARRA, V. EIJKHOUT, R. POZO, C. ROMINE, AND H. VAN DER VORST, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, Philadelphia, 1993.

[3] A. BOURAS, V. FRAYSSÉ, AND L. GIRAUD, *A Relaxation Strategy for Inner-Outer Linear Solvers in Domain Decomposition Methods*, Technical report TR/PA/00/17, CERFACS, Toulouse, France, 2000.

[4] A. Bouras and V. Frayssé, *Inexact matrix-vector products in Krylov methods for solving linear systems: A relaxation strategy*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 660–678.

[5] N. Cundy, J. van den Eshof, A. Frommer, S. Krieg, Th. Lippert, and K. Schafer, *Numerical methods for the QCD overlap operator* III: *Nested iterations*, Comput. Phys. Comm., 165 (2005), pp. 221–242.

[6] J. Drkošová, M. Rozložník, Z. Strakoš, and A. Greenbaum, *Numerical stability of the GMRES method*, BIT, 35 (1995), pp. 309–330.

[7] V. Frayssé, L. Giraud, S. Gratton, and J. Langou, *A set of GMRES routines for real and complex arithmetics on high performance computers*, ACM Trans. Math. Software, 31 (2005), pp. 228–238.

[8] L. Giraud, S. Gratton, and J. Langou, *A Note on Relaxed and Flexible GMRES*, Technical report TR/PA/04/41, CERFACS, Toulouse, France, 2004. Available online from http://www.cerfacs.fr/algor/reports.

[9] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, Frontiers Appl. Math. 17, SIAM, Philadelphia, 1997.

[10] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, 2nd ed., SIAM, Philadelphia, 2002.

[11] K. Mer-Nkonga and F. Collino, *The fast multipole method applied to a mixed integral system for time-harmonic Maxwell's equations*, in Proceedings of the European Symposium on Numerical Methods in Electromagnetics, B. Michielsen and F. Decavèle, eds., ONERA, Toulouse, France, 2002, pp. 121–126.

[12] Y. Saad and M. H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems.*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.

[13] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469.

[14] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.

[15] V. Simoncini and D. B. Szyld, *Theory of inexact Krylov subspace methods and applications to scientific computing*, SIAM J. Sci. Comput., 25 (2003), pp. 454–477.

[16] G. L. G. Sleijpen, J. van den Eshof, and M. B. van Gijzen, *Restarted GMRES with Inexact Matrix–Vector Products*, Technical report TR/PA/04/75, CERFACS, Toulouse, France, 2004.

[17] B. F. Smith, P. E. Bjørstad, and W. Gropp, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, Cambridge, UK, 1996.

[18] J. van den Eshof, G. L. G. Sleijpen, and M. B. van Gijzen, *Relaxation strategies for nested Krylov methods*, J. Comput. Appl. Math., 177 (2005), pp. 347–365.

[19] J. van den Eshof and G. L. G. Sleijpen, *Inexact Krylov subspace methods for linear systems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 125–153.

[20] J. S. Warsa, M. Benzi, T. A. Warein, and J. E. Morel, *Preconditioning a mixed discontinuous finite element method for radiation diffusion*, Numer. Linear Algebra Appl., 11 (2004), pp. 795–811.