Comparison of Quasi Minimal Residual and Bi-Conjugate Gradient Iterative Methods to solve Complex Symmetric Systems arising from Time-Harmonic Magnetic Simulations

H. De Gersem\textsuperscript{c}, D. Lahaye\textsuperscript{d}, S. Vandewalle\textsuperscripts{c} and K. Hameyer\textsuperscript{c}

\textsuperscript{c}Katholieke Universiteit Leuven, Dep. EE (ESAT), Div. ELEN, Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium
\textsuperscript{d}Katholieke Universiteit Leuven, Dep. Computer Science, Celestijnenlaan 200A, B-3001 Leuven, Belgium

Abstract - Finite element discretizations of low-frequency time-harmonic magnetic problems lead to sparse, complex symmetric systems of linear equations. The question arises which Krylov subspace methods are appropriate to solve such systems. The Quasi Minimal Residual method combines a constant amount of work and storage per iteration step with a smooth convergence history. These advantages are obtained by building a quasi minimal residual approach on top of a Lanczos process to construct the search space. Solving the complex systems by transforming them to equivalent real ones of double dimension has to be avoided as such real systems have spectra that are less favourable for the convergence of Krylov based methods. Numerical experiments are performed on electromagnetic engineering problems to compare the Quasi Minimal Residual method to the Bi-Conjugate Gradient method and the Generalized Minimal Residual method.

I. INTRODUCTION

Sinusoidally excited eddy current problems arise in e.g. the design and optimization of induction furnaces, transformers and alternating current machines. They are commonly treated as time-harmonic magnetic field problems [1]. Their finite element discretizations result in sparse, complex symmetric systems of equations. Solving those systems often absorbs more than 90% of the overall CPU-time of the numerical simulation [2]. Hence, a study of iterative methods for this specific kind of matrices is appropriate.

To solve large sparse systems, Krylov subspace iterative solvers have become competitive and often outperform direct solving techniques [3], [4], [5], [6], [7]. This investigation deals with the efficiency and robustness of different Krylov subspace solvers as solvers for complex symmetric systems. Such systems were also considered in [8] and [9]. In the electromagnetic computing community, one typically uses the Bi-Conjugate Gradient Method (BiCG) or the Generalized Minimal Residual Method (GMRES). In this paper, the Quasi Minimal Residual (QMR) method is studied as an alternative for the latter two.

II. TIME-HARMONIC MAGNETIC FORMULATIONS

The relevant equations are derived from the Maxwell equations by taking the following consideration into account. The dimensions of the models are small compared to the electromagnetic wave length corresponding to the frequency range of 0-1000 Hz. Therefore, the displacement-current term $\partial\mathbf{D}/\partial t$ in Maxwell's equations is neglected [1]. The divergence-free condition for the magnetic flux density $\mathbf{B}$ is ensured by expressing $\mathbf{B}$ in terms of the magnetic vector potential $\mathbf{A}$ as $\mathbf{B} = \nabla \times \mathbf{A}$. Faraday's law is reduced to

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t},$$

with $\mathbf{E}$ the electric field strength and $V$ the electric scalar potential. The current density $\mathbf{J}$ is related to $\mathbf{E}$ by

$$\mathbf{J} = \sigma \mathbf{E} = -\sigma \nabla V - \sigma \frac{\partial \mathbf{A}}{\partial t} = \mathbf{J}_s + \mathbf{J}_e,$$

where $\sigma$ is the electric conductivity, $\mathbf{J}_s$ the source current density and $\mathbf{J}_e$ the eddy current density. Ampère's law in terms of $\mathbf{A}$ and $V$ is

$$\nabla \times (\nu \nabla \times \mathbf{A}) + \sigma \frac{\partial \mathbf{A}}{\partial t} = -\sigma \nabla V,$$

where $\nu$ is the magnetic reluctivity. In a lot of devices, the magnetic field varies sinusoidally in time. In that case $\mathbf{A}$ and $V$ are represented by the complex phasors $\hat{\mathbf{A}}$ and $\hat{V}$ as

$$\mathbf{A} = \text{Re}\{\hat{\mathbf{A}} e^{j\omega t}\}$$

and

Manuscript received September 22, 1998.
H. De Gersem, phone 00 32 16 321020, fax 0032 16 321985, e-mail herbert.degersem@esat.kuleuven.ac.be.
D. Lahaye, phone 0032 16 327082, fax 0032 16 327996, e-mail domenico.lahaye@cs.kuleuven.ac.be.
This research is supported by the Fund for Scientific Research-Flanders (project G.0427.98), The Research Fund of K.U. Leuven (OT /94/16) and the "Interuniversity Poles of Attraction Programme - Belgian State, Prime Minister's Office -Federal Office for Science, Technical and Cultural Affairs" (IUAP No. P4/20).
Given an initial guess $x_0$, Krylov subspace methods solve the linear system (11) iteratively by computing iterands

$$x_m \in x_0 + K^m(A,r_0),$$
with $r_0 = b - Ax_0$ the initial residual, converging to the exact solution for increasing iteration step $m$. Let the columns of $V_m = [v_1, ..., v_m]$ form a basis for $K^m(A,r_0)$. Vector $x_m$ is then expressed as $x_m = x_0 + V_m y_m$ where $y_m$ is a vector of coefficients. Krylov subspace methods differ in the construction of the basis $V_m$ and the criterium employed for determining $y_m$.

A. Krylov Subspace Basis Construction

Two methods are commonly used to construct a basis for $K^m(A,r_0)$: Arnoldi’s method and Lanczos’s method.

A.1 Arnoldi’s method

Assume that a set of $m$ orthonormal vectors $\{v_1, ..., v_m\}$ has already been computed. The Arnoldi method expands this set to a orthonormal basis for $K^{m+1}(A,r_0)$ by orthogonalizing $t = Av_m$ with respect to this set and normalizing the resulting vector to give $v_{m+1}$. The matrix $V_m$ and the vector $v_{m+1}$ then satisfy

$$AV_m = V_{m+1}\tilde{H}_{m+1,m} \text{ where } V_{m+1}^HV_{m+1} = I_{m+1}. \quad (13)$$

Here, $\tilde{H}_{m+1,m}$ is an upper $(m+1) \times m$ Hessenberg matrix with elements defined by the Arnoldi algorithm. The $(m+1)$-th basis vector $v_{m+1}$ is constructed using all previous vectors $\{v_1, ..., v_m\}$, resulting in long recurrences. These long recurrences are expensive: both the memory requirements (storing all vectors $v_i$) and the CPU-time (orthogonalizing with respect to all $v_i$) grow linearly with $m$.

A.2 Lanczos method

The cost of long term recurrences can be avoided by resorting to a bi-orthogonalization procedure. Then the basis $V_m$ is orthogonalized with respect to a set of basis vectors $\{w_1, ..., w_m\}$ of a second Krylov subspace $L_m$. In this case, we have that

$$W_m^HV_m = D_m T_{m,m} \text{ where } W_m^HV_m = D_m \quad (14)$$

and where $D_m$ and $T_{m,m}$ are diagonal and tridiagonal respectively. For complex symmetric matrices, a convenient choice for the space $L_m$ is

$$L_m = \overline{K^m(A,r_0)} = \{v \in K^m(A,r_0)\}. \quad (15)$$

Then $W_m$ can be taken as $\overline{F}_m$ and the $V_m$-sequence is

$$AV_m = V_{m+1}\tilde{F}_{m+1,m} \text{ where } V_m^HV_m = D_m. \quad (16)$$
$\bar{T}_{m+1,m}$ is the $(m+1) \times m$ tridiagonal matrix formed as an expansion of $T_{m,m}$. In this way the need of constructing the space $L_m$ explicitly is avoided. As the matrix $T_{m,m}$ in (14) is tridiagonal, the basis $V_m$ can be computed by short recurrences: computing $v_{m+1}$ requires as basis vectors only $v_m$ and $v_{m-1}$. The Lanczos algorithm (14) corresponds to the Arnoldi algorithm where the bilinear form $<v,w>_\tau = w^Tv$ is used instead of the inner product $<v,w>_H = w^Hv$. This Lanczos algorithm may break down in the event that a generated basis vector is a quasi-null vector, i.e. a vector for which $<v,v>_\tau = 0$ even if $v \neq 0$.

B. Projection and Minimal Residual Criteria

Once the basis $V_m$ has been constructed, the coefficient vector $y_m$ can be computed by either a Ritz-Galerkin, Petrov-Galerkin or a (quasi-) minimal residual approach.

B.1 Ritz-Galerkin approach

The Ritz-Galerkin approach requires that the residual vector $r_m = b - A \bar{x}_m$ is orthogonal to the Krylov subspace with respect to the proper inner product:

$$V_m^H r_m = 0.$$ (17)

B.2 Petrov-Galerkin approach

In the Petrov-Galerkin approach, the residual $r_m$ is made orthogonal to the space $L_m$. With the choice of $L_m$ in (15), this requirement is equivalent to

$$V_m^H r_m = 0.$$ (18)

B.3 (Quasi) Minimal Residual approach

In the minimal residual approach, the Euclidean norm $\|r_m\|_2$ is minimized over the space $K^m(A,r_0)$ at each iteration step.

The basis vectors generated by the Lanczos algorithm are $<,>_\tau$-orthogonal, rather then $<,>_H$-orthogonal. Hence, the approach determines $y_m$ such that

$$\|r_m\|_2 = \|V_{m+1}(e_0 - \bar{T}_{m+1,m}y_m)\|_2$$

$$= \min_{y \in \mathbb{C}^m} \|V_{m+1}(e_0 - \bar{T}_{m+1,m}y)\|_2$$ (19)

requires solving a least-squares problem that leads to an algorithm for which work and storage per iteration step grow linearly with $m$. Therefore, Freund [8] suggested to replace the true minimization by the following quasi-minimization:

$$\|r_m\|_2 = \|\omega_0 e_1 - \Omega_{m+1} \bar{T}_{m+1,m}y_m\|_2$$

$$= \min_{y \in \mathbb{C}^m} \|\omega_0 e_1 - \Omega_{m+1} \bar{T}_{m+1,m}y\|_2$$ (20)

where $\Omega_{m+1} = \text{diag}(\omega_0, \ldots, \omega_{m+1})$ is a scaling matrix with weights $\omega_k = \|r_k\|_2$.

The Lanczos and Arnoldi algorithms are combined with one of the above criteria to obtain different Krylov subspace methods for solving (11) (Table I). For real matrices, the Conjugate Orthogonal Conjugate Gradient (COCG) and Quasi Minimal Residual (QMR) algorithms reduce to the Conjugate Gradient (CG) and Minimal Residual (MINRES) algorithms respectively. Both the Full Orthogonalization Method (FOM) and the Generalised Minimal Residual method (GMRES) use long recurrences. GMRES minimizes the residual at each iteration. FOM minimizes the error.

Thanks to the structure of the matrix $\bar{T}_{m+1,m}$, the least-squares problem on the right-hand side of (20) can be solved appropriately using a Givens QR factorization. Numerically stable implementations of Givens rotations for complex variables are available in recent BLAS I implementations. Numerical stability of the algorithm further requires the three-terms recurrences for generating the basis to be implemented as coupled two-terms recurrences [12], [13].

Both the COCG and coupled two-terms QMR algorithms are susceptible to breakdown at two stages: in the Lanczos algorithm due to quasi-null basis vectors (breakdown of the first kind) and the linear system that defines the next iterate becoming singular (breakdown of the second kind). Both breakdowns have been observed in our numerical experiments. To overcome these breakdowns, a look-ahead strategy [14], [15], [16] needs to be incorporated into the algorithm.

IV. PRECONDITIONING QMR

To derive a preconditioned version of the QMR algorithm, we apply the QMR algorithm to the transformed system $A\bar{x} = \bar{b}$ where $A = C^{-T}AC^{-1}$, $\bar{x} = Cx$, $\bar{b} = C^{-T}b$ and $C$ is such that the preconditioning matrix $M$ can be written as

$$M = C^TC.$$ (22)
The preconditioner $M$ is thus complex and symmetric. We proceed in a way similar to Section 10.3 in [17]. Due to (21), the resulting algorithm requires in the $m$-th iteration step the normalization of the vector $Cv_m^{-1}$, obtained from the Lanczos process by $<,>_T$-orthogonalizing $\hat{A}C^{-1}v_{m-1}$ with respect to $\{C^{-1}v_{m-2},C^{-1}v_{m-1}\}$ and thus the computation of $\|C^{-1}v_m\|_2$. (23)

For real matrices (23) can be rewritten in such a way that the matrix $C$ disappears, as follows

$$\|C^{-1}v_m\|_2 = <v_m,C^{-T}C^{-1}v_m>_T = <v_m,M^{-1}v_m>_T$$

(24)

with $M$ defined in (22). For complex matrices however, we have that

$$\|C^{-1}v_m\|_2 = <v_m,C^{-H}C^{-1}v_m>_H \neq <v_m,M^{-1}v_m>_H.$$  

(25)

For the Symmetric Successive Over-Relaxation (SSOR) preconditioner, the factor $C$ can be explicitly constructed. For other preconditioners, the need for explicitly constructing the factor $C$, can be avoided by defining other normalization weights in (21).

V. COMPARISON OF QMR WITH COCG

As COCG does not have a minimal residual property, its convergence behaviour can be irregular (Fig. 2 and Fig. 3). This behaviour prevents COCG to be numerically robust. One can prove the following estimate for the difference at step $m$ between the iteratively updated residual $\|r_m\|_2$ and the true residual $\|p - Ax_m\|_2$ (Section 5.3 in [6]):

$$\|r_m\|_2 - \|p - Ax_m\|_2 \leq m\bar{\xi} f(A) \max_{j=1,...,m} \|f_j\|_2$$

(26)

where $\bar{\xi}$ is the relative machine precision and $f(A)$ is a function of matrix $A$. QMR on the other hand has a smoother convergence history.

The overall number of iterations needed by the COCG and QMR algorithms to reach a prescribed tolerance is about the same (Fig. 2). This can be explained by taking a closer look to the linear systems that the COCG and QMR algorithms solve at each step $m$ to determine $y_m$. For the COCG and QMR algorithms, these linear systems only differ by an additional row in the QMR case. Details can be found in e.g. in Section 6.5.5 in [3].

VI. COMPARISON OF QMR WITH GMRES

A natural question is by how much the quasi-minimization computed by QMR differs from the true minimization computed by GMRES. In Fig. 4 the convergence histories of SSOR-QMR and SSOR-GMRES are plotted for a representative test case, small enough to run a non-restarted...
version of GMRES. From this figure it is concluded that, for the examples considered here, the quasi-minimization approximates the true minimization rather well. Here, QMR behaves like the optimal method. This means that efforts in speeding up computations should be directed towards finding more efficient (possibly multilevel type) preconditioners [18].

VII. COMPLEX VERSUS EQUIVALENT REAL SYSTEMS

Splitting the real and imaginary parts of the complex system matrix leads to two possible equivalent real systems of double dimension [8] (Fig. 5). For the spectra of the matrices $A_*$ and $A_{**}$ the following properties hold:

\[
\sigma(A_*) = \sigma(A) \cup \sigma(A')
\]

and

\[
\sigma(A_{**}) = \left\{ \lambda \in R | \lambda^2 \in \sigma(A A') \right\}
\]

respectively (Fig. 5). Both the spectra for $A_*$ and $A_{**}$ are less favourable for the convergence of Krylov subspace iterative methods than the spectrum of the original complex matrix $A$.
converges much smoother compared to COCG. The harmonic magnetic finite element discretizations. QMR symmetric matrices arising from low-frequency time-subspace projection method to solve the sparse and complex variants for the studied models.

of iteration steps of COCG, QMR and their preconditioned discretizations of both models. Table III presents the number of equations and non-zero matrix elements for different shown in Fig. 10.

windings are voltage-driven. A magnetic flux line plot is modelled by external electric impedances (Fig. 9). The stator end-windings and rotor end-rings are stator slots and 34 rotor slots is computed under loaded field at two instants of time.

resistive load. Fig. 7 shows the flux line plots of the magnetic primary is voltage driven. The secondary is connected to a copper conductors in which skin effect is not negligible. The voltage windings. The secondary windings consist of hollow (secondary) Eddy current effects are neglected in the high voltage windings (primary) and low voltage windings

VIII. PRACTICAL EXAMPLES

A three-phase power transformer (Fig. 6) consists of high voltage windings (primary) and low voltage windings (secondary) Eddy current effects are neglected in the high voltage windings. The secondary windings consist of hollow copper conductors in which skin effect is not negligible. The primary is voltage driven. The secondary is connected to a resistive load. Fig. 7 shows the flux line plots of the magnetic field at two instants of time.

A three-phase induction motor (Fig. 8) with 4 poles, 36 stator slots and 34 rotor slots is computed under loaded condition. The stator end-windings and rotor end-rings are modelled by external electric impedances (Fig. 9). The stator windings are voltage-driven. A magnetic flux line plot is shown in Fig. 10.

Table II compares the numbers of nodes, extra circuit equations and non-zero matrix elements for different discretizations of both models. Table III presents the number of iteration steps of COCG, QMR and their preconditioned variants for the studied models.

IX. CONCLUSIONS

A Quasi Minimal Residual approach is used as Krylov subspace projection method to solve the sparse and complex symmetric matrices arising from low-frequency time-harmonic magnetic finite element discretizations. QMR converges much smoother compared to COCG. The convergence rate of QMR approximates the one of GMRES, indicating that QMR is a valuable low-cost iterative method to solve the complex symmetric matrices of these discretizations. Further improvement should concentrate on the problem of preconditioning and on incorporating look-ahead strategies to cope with breakdown. The methods are applied to electromagnetic problems of technical importance.

REFERENCES