THE FACTORIZED SPARSE APPROXIMATE INVERSE PRECONDITIONED CONJUGATE GRADIENT ALGORITHM FOR FINITE ELEMENT ANALYSIS OF SCATTERING PROBLEMS

X. W. Ping and T. J. Cui

Institute of Target Characteristics and Identification
State Key Laboratory of Millimeter Waves
School of Information Science and Engineering
Southeast University
Nanjing 210096, China

Abstract—The edge-based finite element method is used for the solution of scattering problems. The factorized sparse inverse preconditioner is considered for the conjugate gradient iterative solution of the large sparse linear systems generated from the finite element method. The efficiency of the proposed preconditioner is illustrated on a set of model problems in the final of the paper. The results suggest that the sparse inverse preconditioner is very efficient for the solution of large-scale electromagnetic scattering problems.

1. INTRODUCTION

In the past several decades, full wave simulation methods such as FDTD [1–3], MOM [4, 5], FEM [6–9] has gained great success in theory and in practical applications. Of these methods, the finite element method (FEM) is most popular for its ability in simulating arbitrary geometries, nonlinear materials and eddy current effects. Despite of its ability in modeling highly irregular geometries as well as penetrable and inhomogeneous materials, an important drawback is the huge computational expenses. When applied to three-dimensional problems in electromagnetics, the number of unknowns escalates rapidly as the size of the problem increases. Therefore, the limiting factor in dealing with three-dimensional problems is the associated demands on computer storage and solution time. For most wave propagation problems, such as the waveguide discontinuities or the

Corresponding author: X. W. Ping (xwping@emfield.org).
microwave integration circuits, the generated FEM linear system can usually be solved efficiently with a craftily written direct method, such as the multifrontal method [10]. However, direct methods scale poorly with problem size in terms of operation counts and memory requirements. When scattering problems are considered, the number of unknowns is usually very large. This is because the absorbing boundary conditions (ABCs) [11, 12] should be adopted to truncate the computational domain. To obtain accurate results, ABCs should be placed at least $0.2\lambda$ away in order to obtain accurate results. As a result, the FEM unknowns generated are usually larger than 30,000 even a very small problem is considered, and systems with several millions of unknowns are routinely encountered in many applications. The large size of the problems usually precludes the consideration of Gaussian elimination-based direct solution methods due to the prohibitive memory requirement. Further more, FEM linear systems are usually rather ill-conditioned [13] which is very challenging to solve with classical iterative methods [14–16]. Optimal or quasi-optimal multigrid (multilevel) methods [17], which work well for positive definite systems, often run into serious difficulties when applied to highly indefinite systems. The p-version multigrid method [18] needs much memory to solve the inverse of the lowest order FEM matrix and is unsuitable for large problems. The preconditioners based on shifted Helmholtz operators [19], which is very efficient for wave propagation problems, is very inefficient for scattering problems as we tested. Further more, these methods are all highly sequential. Thus, the solution of FEM linear systems is the main difficulty in FEM simulation of scattering problems.

In the past years, most of research contributions on the FEM algorithm are focused on wave propagation problems. The consideration of scattering problems with FEM is comparatively less. Though for perfectly conducting scatterers, the combination of the boundary integral method (BEM) with the multilevel fast multipole algorithm (MLFMA) [20] is very efficient. However, there are many difficulties in dealing with inhomogeneous problems with BEM-MLFMA. Till now, FEM is still the most powerful method in dealing with inhomogeneous materials. Therefore, the development of efficient iterative solvers, especially those with high parallelism for solving large-scale problems, is of vital important for the FEM solution of scattering problems. In this paper, the preconditioned Krylov subspace iterative method [14] is considered. In the next section, a survey on the preconditioned Krylov subspace iterative methods is given. Then the approximate inverse preconditioner [21] is introduced for the FEM simulation of scattering problems.
2. THE FINITE ELEMENT FORMULATION

The FEM computational domain for a typical scattering problem is shown in Fig. 1. The scattering target is formed with arbitrary media, such as inhomogeneous dielectrics, metals, etc. The permittivity and permeability of the media are denoted as $\varepsilon_r$ and $\mu_r$ respectively. In order to truncate the FEM computational domain, the outer boundary of the scattering target is lined with absorbing perfectly matched layers (PMLs). The distance between the PML interfaces with the scatter is denoted with $d_1$, and the thickness of the PML layer is denoted as $d_2$. For simplicity, the computational domain is denoted as $\Omega$, the region occupied by the scatterer is denoted as $\Omega_{sc}$, the bounding surface of the scatterer is denoted as $\Gamma_{sc}$, the normal vector of the bounding surface is $\hat{n}$.

The vector Helmholtz equation based on electric field is written:

$$\nabla \times \mu_r^{-1} \cdot \nabla \times E - k_0^2 \varepsilon_r E = 0$$  \hspace{1cm} (1)

where $k_0 = \omega \sqrt{\mu_0 \varepsilon_0}$. In FEM simulation, it is more efficient to work with the scattered electric field than the total electric field. By substituting $E = E^{sc} + E^{inc}$ into Eq. (1), one obtain:

$$\nabla \times \mu_r^{-1} \cdot \nabla \times E^{sc} - k_0^2 \varepsilon_r E^{sc} = -\nabla \times \mu_r^{-1} \cdot \nabla \times E^{inc} + k_0^2 \varepsilon_r E^{inc}$$  \hspace{1cm} (2)

According to the generalized variational principle, the functional pertinent to the scattering field is written:

$$F(E^{sc}) = \frac{1}{2} \int_{\Omega} \left[ \mu_r^{-1} (\nabla \times E^{sc}) \cdot (\nabla \times E^{sc}) - k_0^2 \varepsilon_r E^{sc} \cdot E^{sc} \right] dV$$

$$+ \int_{\Omega_{sc}} \left[ \mu_r^{-1} (\nabla \times E^{sc}) \cdot (\nabla \times E^{inc}) - k_0^2 \varepsilon_r E^{sc} \cdot E^{inc} \right] dV$$

$$+ \int_{\Gamma_{sc}} \left[ E^{sc} \cdot \hat{n} \times \nabla \times E^{inc} \right] dS$$  \hspace{1cm} (3)

Figure 1. Illustration of the FEM computational domain of scattering problems.
Note that in the derivation of Eq. (3), the following formulation is used:
\[ \nabla \times \nabla \times E^{inc} - k_0^2 E^{inc} = 0 \] (4)

In this paper, tetrahedron elements are used for FEM mesh discretization. The Whitney basis functions for tetrahedrons are of the following form:
\[ N = \varepsilon_{i1} \nabla \varepsilon_{i2} - \varepsilon_{i2} \nabla \varepsilon_{i1} \] (5)

Substituting Eq. (5) into Eq. (3), and forcing \( \delta F(E^{sc}) = 0 \), the FEM linear system can be obtained. In the next section, iterative solution of the FEM linear system is discussed.

3. PRECONDITIONED KRYLOV SUBSPACE ITERATIVE SOLVERS

For simplicity, the FEM linear system is written in the following form:
\[ Ax = b \] (6)

where the coefficient matrix \( A = [a_{ij}] \in \mathbb{C}^{n \times n} \) is a sparse symmetric matrix. When problem is large, iterative methods are believed to be the only viable solution means. Especially attractive are the Krylov subspace iterative methods [14, 15] which involve the coefficient matrix only in terms of matrix-vector multiplications and a few vector operations (dot products, vector updates) and can be efficiently parallelized on high-performance computers. Of these methods, the general minimal residual solver (GMRES) and the conjugate gradient (CG) solver is most popular. As FEM matrix is ill-conditioned, GMRES is usually not so efficient as CG due to the loss of iteration history by truncation. Further more, GMRES needs more memory than CG. As a result, CG iteration is used in this paper. Standard CG is only suit for self-adjoint, positive definite systems. For solving the non-Hermitan FEM matrix, Eq. (6) should be changed into the following form:
\[ A^H Ax = A^H b. \] (7)

Here \( A^H \) is the conjugate transpose of \( A \). The convergence rate of CG is mainly determined by the condition number of the coefficient matrix. For the \( i \)th iteration, the error is bound by the following relation [22]:
\[ \| x - x^{(i)} \|_A \leq 2 \| x - x^{(0)} \|_A \left( \frac{\sqrt{k} - 1}{\sqrt{k} + 1} \right)^i \] (8)
Here $x$ denotes the exact solution and $x^{(i)}$ the approximate solution at the $i$th iteration step. $K$ is the condition number defined as:

$$K = \frac{\lambda_{\text{max}}^* \lambda_{\text{max}}}{\lambda_{\text{min}}^* \lambda_{\text{min}}} \quad (9)$$

with $\lambda_{\text{max}}$ the largest and $\lambda_{\text{min}}$ the smallest eigenvalue of $A$. Therefore, it is possible to obtain information about the convergence of CG by interpreting the spectrum distribution of the system matrix [13]. As the FEM matrices are usually ill-conditioned, standard CG converges very slowly. Fortunately, the efficiency and robustness of such iterative method can be improved dramatically by applying appropriate preconditioners to the linear system:

$$M^{-1}AN^{-1}y = M^{-1}b, \quad x = N^{-1}y \quad (10)$$

where $M, N$ are the preconditioning matrices. If $M, N$ are skillfully constructed, the modified linear system is better conditioned and the convergence of CG may be substantially enhanced.

The most popular preconditioners are those based on incomplete LU factorizations of $A$ [23], which are fairly robust and exhibit good convergence for positive definite matrices, yet may fail on strongly indefinite matrices. Though the addition of a perturbed diagonal matrix can stabilize the ILU algorithm, however, there are still two difficulties to applying ILU preconditioners in FEM. Firstly, the best perturbed diagonal matrix is hard to determine. Secondly, when the fill-in elements increased, the efficiency of the ILU preconditioner doesn’t necessarily improve, as is often the case in the FEM analysis [24]. Thus the efficiency of the ILU preconditioner cannot be improved through adding more fill-in elements. As have been demonstrated in [25], the SSOR preconditioner is more stable and efficient than the ILU preconditioner. Further more, as the ILU factorization is based on Gaussian elimination, and the forward and backward triangular solves that form the preconditioning operations are highly sequential in nature, it is difficult to be implemented efficiently on parallel computers, especially for unstructured problems.

In this paper, preconditioning techniques based on the sparse approximate inverses (AINV) and factorized sparse approximate inverses (FSAI) are investigated [21]. The basic idea underlying this class of algorithms is that a sparse matrix $M \approx A^{-1}$ or $MN \approx A^{-1}$ is explicitly computed. Approximate inverse techniques rely on the assumption that for a given sparse matrix $A$, it is possible to find a sparse matrix $M$ which is a good approximation, in some sense, of $A^{-1}$. It can be proved that the inverse of an irreducible sparse matrix is structurally full. Nevertheless, it is often the case that many of the
entries in the inverse of a sparse matrix are small in absolute value, thus making the approximation of \( A^{-1} \) with a sparse matrix possible. There are many methods to construct the approximate inverse preconditioner, such as the incomplete biconjugation algorithm [26], the Frobenius norm minimization algorithm [27], etc. Compared with the biconjugation algorithm and the ILU preconditioners, the approximate inverse preconditioner based on Frobenius norm minimization has many advantages: The preconditioning operation is highly parallel, the construction and application of the preconditioner is immune from such numerical difficulties as pivot breakdowns and instability. As a result, the Frobenius norm minimization algorithm is considered in this paper. The construction process of FSAI and AINV is quite similar. However, the FSAI preconditioner can save about 50% memory compared with AINV preconditioner. As the FEM matrix is symmetric, the FSAI preconditioner is mainly proposed in this paper. In the following, the construction of the FSAI preconditioner is mainly discussed.

4. SPARSE APPROXIMATE INVERSE PRECONDITIONING

Let \( A = (a_{ij}) \) be a \( n \times n \) complex sparse symmetric matrix which can be factorized into \( A = U^T U \). The FSAI preconditioning is searched which is a good approximation to \( U^T \):

\[
AZ \approx U^T
\]

where \( Z = (z_{ij}) \) is an upper triangular matrix that is completely determined by its sparsity pattern \( S \):

\[
z_{ij} = 0 \quad \text{if} \quad (i, j) \in S
\]

The sparsity pattern \( S \) is assumed to satisfy the conditions:

\[
\{(i, j) : i > j\} \subseteq S \subseteq \{(i, j) : i \neq j\}
\]

In this paper, \( S \) is chosen to be the same as the pattern of the upper triangular part of the original matrix \( A \). The Cholesky factor \( U \) is unknown. However, \( Z \) can be computed to make the upper triangular of \( AZ \) as minimal as possible. This can be realized by two steps. Firstly, defining \( \tilde{Z} = (\tilde{z}_{ij}) \) an auxiliary matrix with the same zero pattern as \( Z \), the non-zero entries of \( \tilde{Z} \) is searched by solving the following least square problems [14]:

\[
F(\tilde{Z}) = \sum_{k=1}^{n} \min \| A_k \tilde{z}_k - e_k \|_F^2
\]
where $A_k = (a_{ij}), 1 \leq i, j \leq k, \bar{z}_k = [\bar{z}_{k1}, \bar{z}_{k2}, \ldots, \bar{z}_{kk}], e_k = \{0, 0, \ldots, 1\}$ are vectors of order $k$. Denoting $J$ the set of indices with non-zero entries in $\bar{Z}_k$, and $I$ the so-called shadow of $J$ in $A_k$. $I$ is defined as the set of the indices of the non-zero rows in the submatrix $A_k(:, J)$. Hence (14) can be reduced to the solution of the following least square problems:

$$F(\bar{Z}) = \min \|A'_k \bar{z}'_k - e'_k\|, \quad k = 1, \ldots, n$$

with

$$A'_k = A_k (I, J), \quad \bar{z}'_k = \bar{z}_k (J), \quad e'_k = e_k (I)$$

This minimization problem (15) is embarrassingly parallel. To solve the least squares problems, the classical method is to perform QR-decompositions to $A'_k$. However, to reduce the costs of computing $\bar{Z}$, the conjugate gradient (CG) iterative techniques is applied in this paper to solve the least square problems in Eq. (15). To apply CG iteration, the least square problems in Eq. (15) is firstly transformed into the following linear equations:

$$A'^H_k A'_k \bar{z}'_k = A'^H_k e'_k, \quad k = 1, \ldots, n$$

The matrix-vector multiplications during CG iterations can be done in sparse mode in two steps with $q = A'_k p, r = A'^H_k q$, where $p, q$, and $r$ are small dense vectors. The sparse multiplication $A'_k p$ is carried out by $|J|$ SAXPYs (i.e., Scalar Alpha X Plus Y) of sparse column vectors, while $A'^H_k q$ is computed as $|J|$ inner products of a sparse column of $A'_k$ with the full vector $q$. Through this scheme, the construction of the preconditioner can take full advantage of the sparsity of $A'_k$. What’s more, as the number of nonzeros in $\bar{z}'_k$ is usually very small, CG iteration converges very rapidly. Thus the preconditioner can be constructed efficiently.

After $\bar{Z}$ is constructed, the FSAI preconditioning matrix $Z$ can be obtained through the following operation:

$$Z = \bar{Z} D$$

where:

$$D = \left[\text{diag} (\bar{Z})\right]^{-1/2}$$

Once $Z$ is obtained, the preconditioned matrix is written:

$$A \rightarrow Z^T A Z$$

and the preconditioned equations is written:

$$Z^T A Z y = Z^T b, \quad x = Z y$$

In Eq. (15), if the sparsity pattern of $\bar{z}_k$ is chosen the same as the $k$th column of $A$, the AINV preconditioner will be constructed.
The construction process is the same with that of FSAI, except the operation in Eq. (17) is not needed, i.e., the AINV preconditioner is $\bar{Z}$. Note that to guarantee accuracy, right preconditioning is used for AINV in this paper.

As the solution of the normal equations is needed, version 3 of the CG algorithm in [6] is used. In this algorithm, the residual is the same as that of the original system each iteration. Thus no digits of accuracy are lost. To show the validity and capability of the FEM technique and the efficiency of FSAI preconditioning, a set of electromagnetic problems are tested in the next section.

5. NUMERICAL RESULTS

In the experiments, unless otherwise stated, the initial approximate solution is taken to be zero, and the iterative procedures terminated when the normalized residual norm reaches $-60$ dB.

The first example considered is a perfectly conducting (PEC) sphere with diameter $1.0\lambda$. The distance from the inner interface of PML to the outer surface of the sphere is $0.5\lambda$, and the thickness of the PML layer is $0.25\lambda$. In order to apply FEM simulation, the problem is divided into 674972 tetrahedrons. The number of FEM unknown edges generated is 771156, and the number of nonzero elements in the FEM matrix is 3910031. As the FEM matrix is symmetric, only half of the matrix is needed to be stored, the memory needed to store the FEM matrix is 399.2 Mb.

![Figure 2. Bi-static radar cross sections of a PEC sphere with diameter $1\lambda$.](image)
Firstly, the bistatic radar cross sections of the PEC sphere is computed. The incident elevating angle is 90°. The comparison of FEM results with the exact Mie series solutions is depicted in Fig. 2. In this example, a maximum error is about 0.16 dB in the VV-polarized case, and 0.06 dB in the HH-polarized case, which is good agreement considering the dynamic range of over 20 dB in the RCS. This showed that FEM can give accurate results when PML is placed enough far away from the targets.

Secondly, the efficiency of the FSAI and AINV preconditioner is tested. The SSOR and standard IC0 preconditioner is adopted in this paper for comparison. The convergence histories of SSOR-CG, IC0-CG, FSAI-CG, and AINV-CG are plotted in Fig. 3. From the figure, we can see that both AINV-CG and FSAI-CG converges much faster than SSOR-CG and IC0-CG. The construction time of FSAI-CG, and AINV-CG are respectively 63 s and 228 s, and the iteration time for SSOR-CG, IC0-CG, FSAI-CG, and AINV-CG are respectively 28440 s, 38031 s, 12820 s, and 11463 s. It can be seen that the efficiency of standard IC0 preconditioner is poor in this example. A perturbation to the diagonal would improve the efficiency of the IC preconditioner. However, the best perturbation matrix is hard to determine. Further more, IC is not suit for parellel computation, thus it is not further discussed in this paper. It can be seen that the efficiency of AINV-CG and FSAL-CG is comparable. However, the memory used by the AINV preconditioner is 650.3 Mb, while the memory used by the FSAI preconditioner is 399.2 Mb. Note that in [24], the comparison of FSAI and SSOR is made. However, [24] is mainly aimed at wave propagation problems. Further more, the number of unknowns tested in [24] is much small compared with scattering problems. By comparison, it can be seen that the FSAI preconditioner is more efficient for scattering problems.
Next, a dielectric sphere is simulated. The diameter of the sphere is $1\lambda$, and the dielectric constants are $\varepsilon_r = 4.0 - 1.0 j$, $\mu_r = 1.0$. The distance from the inner interface of PML to the outer surface of the sphere is $0.2\lambda$, and the thickness of the PML layer is $0.3\lambda$. The computational region is divided into 197428 tetrahedrons. The number of FEM unknown edges generated is 220604. The memory to store the FEM matrix is 111.6 Mb. The bi-static RCS parameters computed with FEM and the Mie series solutions are plotted in Fig. 4. From the figure, it can be seen that both curves are totally coincident.

![Figure 4](image1)

**Figure 4.** Bi-static radar cross sections of a dielectric sphere with diameter $1\lambda$.

![Figure 5](image2)

**Figure 5.** Convergence history of different PCG iterative solvers in the solution of scattering by the dielectric sphere with diameter $1\lambda$. 

In this example, the CPU time needed for construction the FSAI and AINV preconditioner are respectively 18 s and 63 s. The CPU time needed by SSOR-CG, IC0-CG, FSAI-CG, and AINV-CG are respectively 2823 s, 7184 s, 1592 s and 1444 s. The convergence histories of SSOR-CG, IC0-CG, FSAI-CG, and AINV-CG for the dielectric sphere are depicted in Fig. 5. For this example, it can be seen that FSAI-CG, and AINV-CG still exhibit a much better convergence behavior than SSOR-CG. To examine the scaling of the proposed method with \( \lambda \), the diameter \( d \) of the dielectric sphere is varied from 0.5 to 2.5, while other parameters remain unchanged. The variation of iteration numbers with \( d \) is plotted in Fig. 6. From the figure, it can be seen that the FSAI and AINV preconditioner scale well with \( \lambda \).

To examine the ability and efficiency of FEM in simulating inhomogeneous and anisotropic targets, the third example examined is an anisotropic sphere with a diameter of 1\( \lambda \). The FEM model and meshes are the same as the above example except that the relative permittivity in the \( r, \theta, \phi \) direction is:

\[
\varepsilon_r = \begin{bmatrix}
2.0 - 4.0j \\
\varepsilon_r
\end{bmatrix}
\]

Note that in the \( x, y, z \) direction, the permittivity is a 3 \( \times \) 3 full tensor, and is varied with the value of \( \theta \) and \( \phi \). The bi-static RCS parameters computed with FEM and the hybrid FEM/BEM method (FEBI) [28] are plotted in Fig. 7. As can be seen, both results agreed quite well. This proved that our code is very accurate in simulating anisotropic targets. In this problem, SSOR-CG used about 8000 iterations to converge, while FSAI-CG used 2944 iterations, AINV-
CG used 2623 iterations. Fig. 8 showed the convergence curves of different iterative solvers. Compared with Fig. 5, the iteration number of SSOR-CG increased greatly, while the iteration number of AINV-CG and FSAI-CG is nearly invariant.

The last example is a conducting-dielectric compound cylinder [29]. The dielectric part is formed by plexiglass ($\varepsilon_r = 2.6$), the metal part is aluminum. The configuration and dimensions of the structure are illustrated in Fig. 9. In our FEM simulation, aluminum is treated as PEC for simplicity. This example is divided into 232893 tetrahedrons, and the number of unknown edges is 262007. The computed and measured backscatter cross sections at 3.0 GHz are depicted in Fig. 10. In the figure, the simulated results agreed very well with the measured data.

![Figure 7](image1.png)

**Figure 7.** Bistatic RCS of the anisotropic sphere, the circle represents data given by the FEBI method.

![Figure 8](image2.png)

**Figure 8.** Convergence history of different PCG iterative solvers in the solution of scattering by the anisotropic sphere with diameter $1\lambda$. 
Figure 9. Configuration of the conducting-dielectric compound cylinder, \( a = 5.08 \, \text{cm}, \, b = 10.16 \, \text{cm}, \, d = 7.62 \, \text{cm} \).

Figure 10. Computed and measured backscatter cross sections at 3.0 GHz for the inhomogeneous conducting-dielectric cylinder.

For comparison, the problem is computed repeatedly using SSOR-CG, FSAL-CG and AINV-CG with incident angles from \( \theta = 0^\circ \) to \( \theta = 180^\circ \). The iteration numbers at different angles are recorded and depicted in Fig. 11. From the figure, FSAL-CG and AINV-CG showed a much better convergence behavior than SSOR-CG at all angles. Once again, the efficiency of FSAL-CG in solving scattering problems of inhomogeneous media was demonstrated.
Figure 11. Number of iterations of AINV-CG and FSAI-CG for the inhomogeneous conducting-dielectric.

6. CONCLUSION

In this paper, the FSAI preconditioner is proposed to solve the linear systems obtained from FEM simulation of scattering problems. The comparison of FSAI-CG, AINV-CG and SSOR-CG are made. From the illustrated results, the efficiency of FSAI and AINV preconditioners is comparable. However, FSAI needs much less memories. Compared with the SSOR preconditioner, the FSAI preconditioner is very robust and can greatly reduce iterations and CPU time for the tested problems. These results demonstrate that the FSAI preconditioning strategy is especially effective for solving large-scale FEM linear systems generated from scattering problems even in sequential algorithms.

ACKNOWLEDGMENT

This work is supported in part by the China Postdoctoral Science Foundation (No. 20080441007, 200902497), in part by the Jiangsu Planned Projects for Postdoctoral Research Funds (No. 0801001C), in part by the National Science Foundation of China (No. 60802001).

REFERENCES

Progress In Electromagnetics Research, PIER 98, 2009


15. Benzi, M., “Preconditioning techniques for large linear systems:


