Comparing FEM-BEM and FEM-DBCI for open-boundary electrostatic field problems

G. Aiello¹, S. Alfonzetti¹, a, G. Borzi², E. Dilettoso¹, and N. Salerno¹
¹ Dipartimento di Ingegneria Elettrica, Elettronica e dei Sistemi (DIEES), Universita’ di Catania, Viale A. Doria, 6 95125 Catania, Italy
² Dipartimento di Ingegneria Civile, Universita’ di Messina, Salita Sperone, 31 98166 Messina, Italy

Received: 18 January 2007 / Accepted: 12 April 2007
Published online: 30 May 2007 – © EDP Sciences

Abstract. This paper compares the hybrid FEM-BEM and FEM-DBCI methods for the solution of open-boundary electrostatic field problems. Both methods couple a differential equation for the interior problem with an integral equation for the exterior one. The comparison shows that FEM-BEM is more accurate than FEM-DBCI but requires more computing time.

PACS. 41.20.Cv Electrostatics; Poisson and Laplace equations, boundary-value problems – 02.70.Dh Finite-element and Galerkin methods – 02.70.Pt Boundary-integral methods

1 Introduction

In the literature several methods have been devised to enable the Finite Element Method (FEM) [1] to solve static and quasi-static electromagnetic field problems in unbounded domains, such as infinite elements [2], coordinate transformations [3], the hybrid FEM/BEM (Boundary Element Method) method [4,5], and the hybrid FEM-DBCI (Dirichlet Boundary Condition Iteration) method proposed by the authors to solve electrostatic [6,7], time-harmonic skin effect [8,9] and eddy current [10] problems.

In the FEM-DBCI method a differential equation, which governs the interior problem, is coupled with an integral one which expresses the unknown Dirichlet condition on the fictitious truncation boundary. The integral equation makes use of the free-space Green function, the integration being performed on a surface strictly included by the truncation boundary. The resulting global algebraic system is partly sparse and partly dense and it is efficiently solved in an iterative way: assuming an initial guess for the Dirichlet condition on the truncation boundary, the sparse FEM equation is solved by means of the conjugate gradient (CG) solver; the dense equation is then used to improve the Dirichlet condition [6]; the procedure is iterated until convergence is reached. This solution strategy is efficient because the CG is applied to the sparse equation only, and the dense equation is used only a few times. An improvement on this solving method is obtained by means of the Generalized Minimal Residual (GMRES) method, as described in [7]. Recently the authors have shown that a similar iterative solution strategy can also be used for the solution of hybrid linear FEM-BEM systems of open-boundary electrostatic problems [11]. Moreover, the GMRES solver can also be applied to a reduced system of linear equations in which the unknowns are the nodal values of the normal derivative of the electric potential on the fictitious truncation boundary [12].

This paper compares the hybrid FEM-BEM and FEM-DBCI methods for the solution of open-boundary electrostatic field problems. The paper is organized as follows. In Section 2 the FEM-BEM formulation is described in detail for an open-boundary electrostatic field problem and the iterative solution of the resulting global system is outlined. In Section 3 the hybrid FEM-DBCI method is briefly recalled. In Section 4 the two hybrid methods are compared by means of a set of numerical examples. Finally the authors’ conclusions are given in Section 5.

2 FEM-BEM formulation and iterative solution

Consider a system of $N_C$ conductors, arbitrarily voltaged at $V_k, k = 1, \ldots, N_C$, embedded in an unbounded free space. Finite-extension non-homogeneous dielectric bodies and/or continuous charge distributions may be present in the proximity of the conductors, as depicted in Figure 1. In order to compute the solution for the electrical potential $\psi$ near the conductors and inside the dielectric bodies and distributed charges, a fictitious truncation boundary $\Gamma_F$ enclosing the whole system is introduced so that a bounded domain $D$ is obtained, whose boundary is given by $\Gamma_F$ and by the conductor surfaces $\Gamma_C$. In $D$ the Poisson

---

a e-mail: alfo@diees.unict.it
An electrostatic system (voltaged conductors, non-homogeneities and charges) enclosed by a truncation boundary.

The electric field \( E \) satisfies the following integral equation:

\[
\nabla \cdot (\varepsilon \nabla v) = -\rho
\]

where \( \varepsilon \) is the electric permittivity and \( \rho \) the charge density.

Equation (1) is subject to the Dirichlet conditions on the conductor surfaces, whereas an unknown Neumann condition is assumed to hold on \( \Gamma_F \).

Discretizing the domain \( D \) by means of finite elements the following FEM system is obtained:

\[
\begin{bmatrix}
A & A_F \\
A_F^T & A_{FF}
\end{bmatrix}
\begin{bmatrix}
v \\
v_F
\end{bmatrix}
= \begin{bmatrix}
b_0 \\
0
\end{bmatrix}
- \begin{bmatrix}
0 \\
C
\end{bmatrix}
\begin{bmatrix}
q_F
\end{bmatrix}
\tag{2}
\]

where: \( v \) and \( v_F \) are the vectors of the unknown values of the potential \( v \) in the nodes inside the domain and on the fictitious boundary \( \Gamma_F \), respectively, \( A, A_F \) and \( A_{FF} \) are sparse matrices of geometrical coefficients, \( b_0 \) and \( q_F \) is the vector of the unknown values of the inward normal derivative \( q = \partial v / \partial n \) of the potential evaluated in boundary nodes other than those of \( v \), as shown in Figure 2 for first- and second-order triangular finite elements, having a side lying on \( \Gamma_F \). On these sides, \( v \) and \( q \) are approximated by:

\[
v = \sum_n v_n \alpha_n
\]

\[
q = \sum_m q_m \beta_m
\]

where \( v_n \) and \( q_m \) are the nodal values of \( v \) and \( q \), and \( \alpha_n \) and \( \beta_m \) are the associated shape functions. Note that, if \( \nu \) is the order of the \( \alpha_n \) functions, \( \nu - 1 \) is the order of the \( \beta_m \) ones, in congruence with the fact that \( q \) is the derivative of \( v \).

In order to solve the unbounded field problem, it is necessary to derive another equation relating the unknown vectors. This equation is obtained by evaluating the BEM integral equation at the nodes \( Q_j \) of the \( q \) unknowns:

\[
\frac{1}{2} v(Q_j) + \int_{\Gamma_F} v(P) \frac{\partial G(P, Q_j)}{\partial n_P} d\Gamma = \int_{\Gamma_F} Q(P) G(P, Q_j) d\Gamma
\]

where \( G \) is the free space Green function, given by:

\[
G(P, Q) = \frac{1}{2\pi} \ln \frac{1}{r}
\]

Equation (5) is a rectangular sparse matrix of unknowns as in (4), these integral equations are rewritten in matrix form as

\[
\mathbf{H}v_F = \mathbf{G} \mathbf{q}_F
\]

where \( \mathbf{H} \) and \( \mathbf{G} \) are dense matrices. Note that the matrix \( \mathbf{G} \) is square by construction.

A common approach to solve the global system (2)--(8) is by means of an iterative CG-like solver for non-symmetric matrices. A similar approach is also used in which the array \( q_F \) is derived from the integral equation and substituted in the differential one, to obtain the reduced system:

\[
\begin{bmatrix}
A & A_F \\
A_F^t & A_{FF} + CG^{-1}H
\end{bmatrix}
\begin{bmatrix}
v \\
v_F
\end{bmatrix}
= \begin{bmatrix}
b_0 \\
0
\end{bmatrix}
\tag{10}
\]

Both these approaches suffer from the fact that in each step of the iterative solver the matrix-vector multiplication is very costly, due to the presence of dense parts in the system matrices. To alleviate the heaviness of the integral equation, whose complexity is \( N^2 \), various FMM (Fast Multipole Method) techniques have been used to reach an \( N \log N \) complexity [13]. However, the integral equation still remains the most time-consuming part of the solving algorithm. Irrespective of whether FMM is used for the BEM equation, it is of great importance to minimize the number of times that this equation is solved in the context of an iterative solution.
In [11] a simple iterative solving strategy was proposed which takes into account the very different nature of equations (2) and (8), as described below:

1. a first guess is arbitrarily selected, for example \( q_{\text{F}}^{(0)} = 0 \);
2. equation (2) is solved for \( v \) and \( v_{\text{F}} \) by means of CG;
3. the square matrix \( G \) is first decomposed into \( L \) and \( U \) matrices, and then equation (8) is solved for \( q_{\text{F}} \);
4. at the generic \( n \)-th step, a convergence indicator is computed, measuring the distance between the new solution for \( q_{\text{F}} \) and the old one:

\[
\eta_n = 100 \frac{\|q_{\text{F}}^{\text{new}} - q_{\text{F}}^{(n-1)}\|_2}{\|q_{\text{F}}^{\text{new}}\|_2} \tag{11}
\]

5. if convergence is not reached go to step 2, assuming a relaxed new guess for \( q_{\text{F}} \) such as:

\[
q_{\text{F}}^{(n)} = \gamma q_{\text{F}}^{\text{new}} + (1 - \gamma) q_{\text{F}}^{(n-1)} \tag{12}
\]

where \( \gamma \) is the relaxation coefficient.

This iterative scheme exhibits the following characteristics:

(a) since the first guess for the CG solver in each step is the solution obtained in the previous iteration step, the various solutions of system (2) get faster as the iteration proceeds;

(b) the LU decomposition is performed only once at the beginning of the iterative procedure; round-off errors in the LU decomposition are treated by selecting an appropriate (double precision) accuracy in the computing and storing of the matrices \( H \) and \( G \) [14];

(c) the whole iterative procedure is convergent if an appropriate relaxation coefficient \( \gamma \) is selected; this, however, is not known a priori; if an inappropriate coefficient \( \gamma \) is used, divergence may occur;

(d) consequently, the integral equation (8) is used only a few times, if compared to its use in an iterative CG-like solver for the whole non-symmetric system (9) or the reduced one (10).

A similar iterative technique was proposed in [15], in which the first guess for the unknown condition on the truncation boundary is a Dirichlet one. This makes the whole solution strategy more complex.

Looking at the above iterative solution more deeply, one realizes that the iterative algorithm can be interpreted as a stationary iterative method applied to the reduced system:

\[
M q_{\text{F}}^{(n)} = k \tag{13}
\]

where:

\[
M = G - H(A_{\text{F}} A_{\text{F}}^{-1} A_F - A_{\text{FF}})^{-1} C \tag{14}
\]

\[
k = H(A_{\text{F}} A_{\text{F}}^{-1} A_F - A_{\text{FF}})^{-1} A_F A_{\text{F}}^{-1} b_0. \tag{15}
\]

Since it is well-known that stationary methods are very weak and possibly non-converging, one can think of using the more robust GMRES solver [16]. The matrix \( M \) and vector \( k \) are not directly available. However, the vector \( k \) is simply built as follows:

1. assume a zero initial guess \( q_{\text{F}}^{(0)} = 0 \);
2. solve the FEM equation (2) by means of the CG solver to obtain \( v^{(0)} \) and \( v_{\text{F}}^{(0)} \);
3. compute \( k = H v_{\text{F}}^{(0)} \), which coincides with the initial residual vector.

Similarly, matrix \( M \) can be used to perform matrix-vector multiplication \( M q_{\text{F}}^{(n)} \), as follows:

1. given the vector \( q_{\text{F}}^{(n)} \);
2. solve the FEM equation (2) by means of the CG solver to obtain \( v^{(n)} \) and \( v_{\text{F}}^{(n)} \);
3. compute \( M q_{\text{F}}^{(n)} = G q_{\text{F}}^{(n)} - H v_{\text{F}}^{(n)} \).

The reason for the choice of GMRES among various non-stationary CG-like solvers for non-symmetric matrices, such as BiCG (BiConjugate Gradient), QMR (Quasi Minimal Residual), CGS (Conjugate Gradient Squared), and BiCGStab (BiCG stabilized), relies on the fact that GMRES performs a true minimization of the residual and is thus the optimal method for accelerating the iterative solution of (13) as it minimizes the number of matrix-vector multiplications (neglecting the other operations required).

The residual can be computed directly with the approximate solution, thus requiring a further matrix-vector multiplication, or by using the orthonormal basis of the Krylov subspace, as explained in [16]. The latter option is definitely preferable since matrix-vector multiplications are much more expensive than in a case where the coefficient matrix is directly available.

The major drawbacks of GMRES are the computing time and memory required to compute and store the orthonormal basis, which increases linearly with the number of iterations, so restarting procedures are often used. In this case the computing time and memory required for the orthonormal basis are only a small fraction of the total, because GMRES works on a reduced system. Most of the computing time and memory is spent on solving (8), i.e. performing matrix-vector multiplications. It is therefore convenient to use long restarts which generally result in a full GMRES due to the quick convergence characteristic of the simple iterative procedure. The fact that the relaxed iterative procedure converges with a suitable choice of a positive relaxation parameter \( \gamma \) indicates that the eigenvalues of the matrix \( M \) have positive real parts, and this assures that GMRES converges to the true solution even with a very short restarting parameter \( m \) [16].

In comparing the GMRES solving algorithm with the simple iterative one, it is to be noted that the GMRES solution does not require the LU decomposition of matrix \( G \). This is a great advantage, especially for problems with a large number of unknowns. Conversely, the various solutions of the FEM equations by means of the CG-solver are not related to each other, so the number of CG steps does not decrease as the solution proceeds.

### 3 FEM-DBCI formulation

In the FEM-DBCI method the differential equation is the same as the FEM-BEM one, but an unknown Dirichlet
condition is assumed on \( \Gamma_F \). Therefore:

\[
A v = b_0 - A_F v_F. \tag{16}
\]

The Dirichlet condition on \( \Gamma_F \) is expressed by means of the integral equation:

\[
v(P_j) = \int_{\Gamma_M} \left[ \frac{\partial v(P)}{\partial n_P} G(P, P_j) - v(P) \frac{\partial G(P, P_j)}{\partial n_P} \right] d\Gamma
\]

where the integration is performed on a surface \( \Gamma_M \) strictly enclosed by \( \Gamma_F \), but enclosing the whole electrostatic system \([6]\), and \( P_j \) is a node on \( \Gamma_F \). Therefore:

\[
v_F = N v \tag{18}
\]

where \( N \) is a dense matrix, in which null columns appear for the internal nodes not involved in the computation \([6]\).

The whole system (16)–(18) is well suited to be solved by means of an iterative algorithm similar to that of the previous section, in which the initial guess is made on \( v_F \). Analogously, a more robust solving strategy was proposed in \([7]\) based on the use of the GMRES to solve the reduced system of (16)–(18) in which the unknown is \( v_F \).

### 4 Comparing FEM-BEM and FEM-DBCI

Comparing the two methods the following considerations can be made. First, in the FEM-DBCI integral equation (18) only one dense matrix \( N \) is present, whereas in the FEM-BEM integral equation (8) two dense matrices \( H \) and \( G \) appear. Since the number of entries of \( N \) is (in general) approximately equal to the sum of those of \( H \) and \( G \), the two methods exhibit comparable memory requirements for the integral equations. The same conclusion can be drawn for the FEM equations, in which the small sparse matrices \( A^t_F \), \( A_{FF} \) and \( C \) are not present in the FEM-DBCI global system. However, the key point which determines the different behaviour of the two methods is that one method adopts a Dirichlet condition on the truncation boundary, whereas the other adopts a Neumann one. This difference implies different convergence characteristics: the FEM-DBCI converges notably faster than FEM-BEM. This behaviour has always been observed experimentally by us, and will be verified in the next section by means of a set of examples. Note that similar behaviour was also observed and studied by comparing the FEM-DBCI method with the FEM-RBCI (Robin Boundary Condition Iteration) one, in which a Robin (mixed) boundary condition is assumed on the truncation boundary \([17]\).

From the point of view of accuracy, it can be noted that in FEM-DBCI a numerical derivative of the potential is performed on the integration surface \( \Gamma_M \), whereas this is not necessary in FEM-BEM. FEM-BEM can therefore be expected to give more accurate results than FEM-DBCI. These considerations can be verified by means of a set of simple examples, in which the computations were performed by means of ELFIN, a large FEM code developed by the authors for electromagnetic CAD research \([18]\).

**4.1 Two-wire transmission line**

Consider a two-wire transmission line constituted by two parallel conducting circular cylinders of radius \( R \) whose centres are separated by a distance of \( d = 2.4R \), voltaged with opposite potentials \( V_0/2 \) and \(-V_0/2\). This electrostatic problem exhibits the well-known analytical solution \([19]\):

\[
v^*(x, y) = \frac{V_0}{4 \cosh^{-1}(d/2R)} \ln \left( \frac{(x + a)^2 + y^2}{(x - a)^2 + y^2} \right) \tag{19}
\]

where:

\[
a = \sqrt{\frac{d^2}{4} - R^2}. \tag{20}
\]

The fictitious boundary was selected as constituted by two circumferences of radius 1.14\(R \) centred at the cylinder centres, so the gap between the conductors and the fictitious boundary is 0.14\(R \). The analysis can be restricted to the first quadrant only, by imposing homogeneous Neumann and Dirichlet boundary conditions on the \(x\)- and \(y\)-axis, respectively. The resulting bounded domain is regularly discretized with second-order triangles. The integration curve \( \Gamma_M \) of the FEM-DBCI was selected as constituted by the triangle sides of the finite-element layer closer to the conductor surface. The global systems were solved with the GMRES schemes (with restarting parameter \( m = 10 \)). To test the accuracy of the various numerical solutions obtained, an accuracy indicator is defined as:

\[
\zeta = 100 \sqrt{\frac{\int_D (v^* - v)^2 dxdy}{\int_D (v^*)^2 dxdy}} \tag{21}
\]

where \( D \) is the domain of FEM analysis (lying in the first quadrant). The computing times for the solution of the linear systems (on a Pentium IV, 3.2 GHz, 4 Gb RAM, with an end-iteration tolerance of 0.01% for the GMRES and 0.0001% for the CG) and the accuracy indicators \( \zeta \) are plotted in Figures 3 and 4, respectively, versus the number of nodes in the various meshes used. As expected, we can see that FEM-BEM is more accurate than FEM-DBCI, but it requires more computing time.
Fig. 4. Accuracy indicator for the first example.

Fig. 5. Triangular discretization on the conductor surface of the second example.

4.2 Conductor cube

Consider a conductor cube of edge size $s$, voltaged at $V_0$ and embedded in free space. For this classical electrostatic problem a very accurate estimate of the normalized capacitance is given by $0.6606780 \pm 2.7 \times 10^{-7}$ [20], the normalizing value being $4\pi\varepsilon_0 s$. In order to apply the FEM-DBCI method a cube fictitious boundary $\Gamma_F$ is selected homologously to the conductor, placed at a distance of $s/2$. For symmetry reasons, the analysis is restricted to one octant only, by imposing homogeneous Neumann conditions on the three symmetry planes. The gap between the two cubes is filled with 70,371 first-order tetrahedra with 13,988 nodes. The triangular discretization on the conductor surfaces is depicted in Figure 5. To avoid the sharp corners on the conductor surface, the integration surface $\Gamma_M$ is selected at a distance $s/10$ from the conductor.

The number of iterations performed (with an end-iteration tolerance of 0.01% for the GMRES and 0.001% for the CG), the computing times for the solution of the linear systems and the normalized capacitances $c$ are given in Table 1. As before, FEM-BEM is more accurate than FEM-DBCI, but more time-consuming.

<table>
<thead>
<tr>
<th>Method</th>
<th>Iterations</th>
<th>CPU time (s)</th>
<th>Capacitance $c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEM-BEM</td>
<td>9</td>
<td>2.724</td>
<td>0.65263</td>
</tr>
<tr>
<td>FEM-DBCI</td>
<td>3</td>
<td>0.821</td>
<td>0.67305</td>
</tr>
</tbody>
</table>

Table 1. Comparison between FEM-BEM and FEM-DBCI for the second example.

Fig. 6. Contours of the potential on a symmetry plane of the third example.

4.3 Square plate condenser

This problem considers a condenser constituted by two parallel square plates of negligible thickness. The size of the plate edges is $s$ and the distance between them is $d = s$. The region between the plates is filled with a dielectric of relative permittivity $\varepsilon_r = 9$, whereas the whole condenser is embedded in air. The two plates are oppositely voltaged at $V_0/2$ and $-V_0/2$. A Cartesian reference frame is selected having the origin in the centre of the condenser and the $x$ and $y$ axes parallel to the plate edges. The unbounded space around the condenser is truncated by means of a cube surface homologous to the condenser, whose edge size is $2s$. For symmetry reasons, the analysis can be restricted to the first octant only, by imposing homogeneous Neumann boundary conditions on the $xz$ and $yz$ planes and a homogeneous Dirichlet condition on the $xy$ plane. The mesh was formed by 93,816 first-order tetrahedra with 17,148 nodes. To avoid the field singularities near the edges of the condenser plates, the FEM-DBCI integration surface $\Gamma_M$ is placed at a distance of $s/10$ from the condenser. Contours of the potential on a symmetry plane are shown in Figure 6.

The number of iterations performed (with an end-iteration tolerance of 0.01% for the GMRES and 0.001% for the CG), the computing times for the solution of the linear systems and the normalized capacitances $c$ (the normalizing value being $\varepsilon_0 s^2/d$) are given in Table 2. Once
Table 2. Comparison between FEM-BEM and FEM-DBCI for the third example.

<table>
<thead>
<tr>
<th>Method</th>
<th>Iterations</th>
<th>CPU time (s)</th>
<th>Capacitance c</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEM-BEM</td>
<td>11</td>
<td>4.397</td>
<td>11.590</td>
</tr>
<tr>
<td>FEM-DBCI</td>
<td>3</td>
<td>1.182</td>
<td>11.631</td>
</tr>
</tbody>
</table>

again we can see that FEM-BEM requires more computing time than FEM-DBCI.

5 Conclusions

In this paper we have compared the hybrid FEM-BEM and FEM-DBCI methods for the solution of open-boundary electrostatic field problems. The comparison has shown that FEM-BEM is more accurate than FEM-DBCI, but requires more computing time. For this reason we conclude that FEM-DBCI is more appropriate for applications which require a shorter computing time, as for example in the stochastic optimisation of electrical devices, where some thousands of analyses should typically be performed to obtain a satisfactory result. Conversely, FEM-BEM is more appropriate in cases in which a high level of precision is required in a single computation.

This work was supported in part by the MIUR (The Italian Ministry for University and Research).

References


To access this journal online:
www.edpsciences.org