# Algorithms for Large Sparse Matrices Based on Finite Element Method

Luigia Puccio, Angela Ricciardello

*Department of Mathematics*
*Messina University, Italy*
*{gina,aricciardello}@unime.it*

## Abstract

Many problems arising from engeneering and scientific computing give rise to large, sparce matrices.

The aim of this work is to describe a set of highly efficient iterative methods for solving linear systems with large sparse matrices, arising from the analysis of seismic body wave propagation.

An "ad hoc" initial boundary value problem is formulated for heterogeneous dissipative media with arbitrary topography. Its numerical implementation is based on Finite Element Method on non structured mesh.

Some results are presented.

*Keywords:* Large Sparse Matrix, Finite Element Method, Gauss-Seidel Method, Incomplete Cholesky Factorization.

## 1. Introduction

In seismology, the physics of continuum domain is described by Partial Differential Equations (PDEs). Their resolution can be realized by means of several numerical techniques, such us *Finite Difference Method* (FDM), *Spectral Element Method* (SEM), *Finite Element Method* (FEM), and so on.

In literature, it is possible to find a wide variety of algorithms for the simulation of earthquake both in 2D [10], [11] and in 3D [1], [8].

In detail, we deal with 3D case. In [1], Fourier and Chebychev methods are used, while in [8], the author introduces a memory optimization procedure that allows large-scale 3D finite-difference problems to be computed on a conventional, single-processor desktop workstation. With this technique, model storage is accommodated using both external (hard-disk) and internal (core) memory.

In this paper, we solve, by means of FEM, an initial boundary value prob-

lem (IBVP) that describes the wave propagation in heterogeneous dissipative media. We have chosen FEM because the traction-free condition at the outer boundary of the medium is already been cast in its formulation. Moreover, its integro-differential equations allow us to consider not regular but integrable functions.

On the contrary, FDM requires much less computer memory and a shorter computational time.

As well, even if SEM has attractive properties regarding accuracy, error tolerance, computational efforts and memory storage, FEM is better suited for handling problems involving non linear behaviors or singularities, and allows a more effective mesh refinement.

The matrix obtained applying FEM is sparse. The entry $(i, j)$ is non zero if and only if the nodes $i$ and $j$ of the mesh are connected, so the pattern of the matrix is obviously symmetric. Also, since we consider a problem in 3 dimensions, the elements of the matrix are $3 \times 3$ blocks. For that, it is memorized with block Compressed Sparse Row (CSR) format.

In the first section of the paper, we describe the wave propagation in a heterogeneous unbounded medium with free-surface. In the second one, we present the FEM and the weak formulation of the problem.

In the last part, the Gauss-Seidel method and a generalization of the incomplete Cholesky factorization have been employed for solving the linear system.

The implementation of our algorithms is tested on a theoretical problem, to compare the numerical solution with the analitical one.

## 2. The model

First of all, we mathematically describe an heterogeneous medium.
The following model is a simplification of the IBVP presented in [2].
Let $\mathbb{R}^3$ be the three dimensional real space and $\underline{x} = (x, y, z) \in \mathbb{R}^3$ a generic vector; let $f(x, y)$ be a regular function in $\mathbb{R}^2$,

$$\Omega = \left\{ (x, y, z) \in \mathbb{R}^3 : z > f(x, y), (x, y) \in \mathbb{R}^2 \right\}$$

whose boundary

$$\partial\Omega = \left\{ (x, y, z) \in \mathbb{R}^3 : z = f(x, y), (x, y) \in \mathbb{R}^2 \right\}$$

describes the so called *free-surface* of $\Omega$.
Its graphically representation is shown in Fig.1. The medium is characterized by the mass density for unit volume $\rho(\underline{x})$ and by the Lamé constants $\lambda(\underline{x})$ and $\mu(\underline{x})$. They, generally, are point functions of the medium, defined for all $\underline{x} \in \Omega$.

Fig. 1. Mathematical scheme of a heterogeneous unbounded medium $\Omega$ with mass density $\rho(\underline{x})$ and the Lamé constants $\lambda(\underline{x})$ and $\mu(\underline{x})$.

The wave propagation is described by the following initial boundary value problem (IBVP), for heterogeneous media,

(1)
$$
\begin{cases}
\rho(\underline{x})\dfrac{\partial^2 \underline{u}}{\partial t^2} = \underline{\nabla}\left((\lambda(\underline{x}) + 2\mu(\underline{x}))\,\underline{\nabla u}\right) - \underline{\nabla}\left(\mu(\underline{x})\underline{\nabla u}\right) + \\
\qquad\qquad + \underline{F}(\underline{x}, t)\ \forall(\underline{x}, t) \in \Omega \times (0, +\infty) \\
\underline{u}(\underline{x}, 0) = 0\ \forall \underline{x} \in \Omega \\
\dfrac{\partial}{\partial t}\underline{u}(\underline{x}, 0) = 0\ \forall \underline{x} \in \Omega \\
\dfrac{\partial}{\partial \underline{n}}\underline{\underline{\sigma}}(\underline{x}, t) = 0\ \forall(\underline{x}, t) \in \partial\Omega \times (0, +\infty)
\end{cases}
$$

The quantity $\underline{\nabla}\left((\lambda(\underline{x}) + 2\mu(\underline{x}))\,\underline{\nabla u}\right) - \underline{\nabla}\left(\mu(\underline{x})\underline{\nabla u}\right)$ represents the inner forces, while $\underline{F}(\underline{x}, t)$ is the input radiation.

## 3. Finite Element Method

FEM can be applied both to engineering and to large-scale seismological problems. It can be especially considered a powerful tool for the numerical resolution of seismic body waves propagation problems.

The integration domain is divided into parts by a non structured mesh, that is able to reconstruct, numerically, the energy shearing processes on the phase directions at the interface. The latter splits regions with different physical properties.

The use of non structured mesh assures the full agreement between interfaces and elements. The interfaces do not split the elements avoiding further approximations of the functions in their nodes.

The mesh is implemented by *LaGriT*, a program prepared by Los Alamos National Security, LLC at Los Alamos National Laboratory (LANL) with the U.S. Department of Energy (DOE). (http://lagrit.lanl.gov/)

The mesh step $dh$ depends on the maximum frequency $f_{max}$ and on the minimum wave velocity $v_{min}$ in $\Omega$ , as follows

$$dh = \sqrt[3]{\frac{v_{min}}{12 f_{max}}}$$

As well FEM is unconditionally stable on non structured mesh.

The equation of waves evolution can be numerically written through the weak formulation and, in particular, using Galerkin method [9], as follows,

$$M^2 \frac{d^2}{dt^2} \underline{U} + M^0 \underline{U} = \underline{f}.$$

$M^2$, $M^0$ are square matrices whose size is three times nodes number. Their entries are

$$M^2_{lm} = \int_\Omega \rho(\underline{x}) N_m(\underline{x}) d\Omega,$$

$$M^0_{lm} = \int_\Omega \left( \lambda(\underline{x}) + \mu(\underline{x}) \right) \underline{\nabla} N_m(\underline{x}) \underline{\nabla} N_l(\underline{x}) d\Omega.$$

$\underline{U}$ and $\underline{f}$ are vectors with entries, respectively,

$$U_m = u_m(t), \quad f_l = \int_\Omega F_l(\underline{x}, t) N_l(\underline{x}) d\Omega.$$

$N_i(\underline{x})$ are the basis function of the weak formulation.

The numerical solution has been greatly enhanced by the use of a scheme for canceling the reflections at the boundaries of the model, the so called *absorbing boundary conditions.*

Because of the half-space $\Omega$ is replaced by a finite domain, in order to avoid artificial reflection introduced by the faces, the artificial boundaries have to be supported by artificial boundary conditions so that the perimeter of computational grid becomes transparent to outward-moving waves .
We adopted the Caserta-Firmani-Ruggiero transparent conditions [10], which are a generalization of the Clayton-Engquist-Stacey conditions [3], [4], [5], [6], [7], [12] because they provide good results both in 2D domain and in our 3D model.

Applying the transparent conditions, our problem can be written in the following form,

$$(2) \qquad M^2 \frac{d2}{dt2}\underline{U} + M^1 \frac{d}{dt}\underline{U} + M^0 \underline{U} = \underline{f}$$

$M^0, M^1, M^2$ have dimension 3 times nodes number.
In $M^0$ and $M^2$ the generic entry $(i, j)$ is not zero, if and only if the nodes $i$ and $j$ of the mesh are connected.
Instead, in $M^1$ the generic entry $(i, j)$ is not zero, if and only if $i$ or $j$ are boundary nodes and, moreover, they are connected.
Then, $M^0$, $M^1$ and $M^2$ are block matrices, mainly zeros. For that, we stored them through compressed sparse row (CSR) format.
The discretization respect to the time is carried out through Newmark method, such us

$$\underline{\dot{U}}^{n+1} = \underline{\dot{U}}^n + \delta t \left( (1 - \beta) \underline{\ddot{U}}^n + \beta \underline{\ddot{U}}^{n+1} \right)$$

$$\underline{U}^{n+1} = \underline{U}^n + \delta t \underline{\dot{U}}^n + \frac{\delta t^2}{2} \left( (1 - \alpha) \underline{\ddot{U}}^n + \alpha \underline{\ddot{U}}^{n+1} \right)$$

So, from the equation (2) , we get

$$(3) \qquad A \underline{\ddot{U}}^{n+1} = \underline{b}^{n+1}$$

where

$$A = M^2 + \beta \delta t M^1 + \alpha \frac{\delta t^2}{2} M^0$$

$$\underline{b}^{n+1} = \underline{f}^{n+1} - \left[ (1 - \beta) M^1 + \frac{\delta t^2}{2} (1 - \alpha) M^0 \right] \underline{\ddot{U}}^n -$$
$$- \left( M^1 + \delta t M^0 \right) \underline{\dot{U}}^n - M^0 \underline{U}^n.$$

The Newmark method has its maximum stability when $\alpha = \frac{1}{4}$ and $\beta = \frac{1}{2}$ . In this case, in fact, it is unconditionally stable and it reaches its maximum accuracy if time step

$$\delta t \leq \frac{dh}{v_{max}}$$

where $dh$ is the mesh step and $v_{max}$ is the maximum velocity in the numerical domain.

## 4. Comparison of Gauss-Seidel Method and Incomplete Cholesky factorization

Matrix sparsity provides an opportunity to conserve storage and reduce computational requirements by storing only significant entries.
There is no single "best" method to represent a sparse matrix.
The selection of the possible storage format is dependent on the algorithm being used, the original sparsity pattern of the matrix, the underlying computer architecture, data format, and so on.
Because of this, typically use of libraries consists of create an internal sparse matrix representation. They return an handle used as parameter in the routine that solves the linear system. If the matrix is no longer needed, the libraries deallocate it to free resources associate with the handle.
In both our developed algorithms, the full matrix is never displayed, but it is constructed piece-by-piece, directly, in CSR format with $3 \times 3$ blocks.
As regards the resolution of a linear system whose coefficients matrix has not a special structure, factorization is usually essential.
The factorization computational cost is proportional to the cube of matrix dimension and the resolution of the system has a computational cost proportional to the square of matrix dimension.
Anyway, the factorization of a sparse matrix involves too many zero elements and the non zero number grows up slowly when matrix dimension increases. To avoid the fill-in of the matrix led by factorization, we made use of incomplete Cholesky factorization. Indeed, it allows us to preserve the number of non zeros of the initial matrix. In general, the factorized matrix is not yet a block-one.
Anyway, if we make only a factorization of the non-zeros blocks, the resolution of the system, at every time step, is anyway slow but the matrix is still sparse and, moreover, the non zeros elements hold the positions taken up before the factorization.
In this way, we entirely keep the structure of the matrix and, thus, the original storage format.
Anyhow, our matrix is not only sparse, but also predominantly diagonally dominant and with large dimension, whereby iterative methods are the most convenient: they converge quickly to the solution of the problem and they avoid the fill-in of the matrix during calculates.
In our case, the fastest serial method for system resolution, is Gauss-Seidel.
On the contrary, the Jacobi method and factorization are parallelizable.

A parallel code is our next step, in order to simulate realistic cases. For this reason, we implemented also a modification of the incomplete Cholesky factorization, in which we require that the $LDU$ has the same pattern of $A$. In our case, indeed, $A$ is factorisable.

Let $A \cong LDU$ be the incomplete Cholesky factorization, and, fixed the time step $n$ , let be

$$\underline{X} = \underline{\ddot{U}}^{n+1}$$

Then, the method looks like

$$LDU\underline{X}^{j+1} = \underline{b}^{n+1} - A\underline{X}^j + LDU\underline{X}^j$$

that allows us a double control

$$\left\|\underline{X}^{j+1} - \underline{X}^j\right\| < \epsilon$$
$$\left\|\underline{b}^{n+1} - A\underline{X}^j\right\| < \epsilon$$

In this case we valuated the analytical solution of the problem 1 and we compared it with the solution obtained using both Gauss-Seidel method and incomplete Cholesky factorization.
Obviously, by the nature of the the matrix $A$, the Gauss-Seidel method provides a solution in less time.
The Gauss-Seidel method has been appropriately validated by comparison with the direct method of MATLAB (`mldivide`).

### 4.1. *Numerical test*

In this section, we validate the numerical model. For this reason, we consider a learning problem, whose solution is known, and we compare it with the numerical solution obtained by means of Gauss-Seidel.
Let $\Omega$ be a cube with edge $L = 2 \ m$, p-velocity $v_p = 8m/s$, s-velocity $v_s = 6m/s$, maximum frequency $f_{max} = 5 \ Hz$, then, by convergence conditions, the smallest spatial step $dh = 0.5m$ and the time step $dt = 1.25E - 04s$.
We suppose that the seismic source lies on the plane $z = 1$.
In our case, the input radiation is represented by the Gabor impulse

$$g(t) = \exp\left[-\frac{\omega_p \left(t - t_s\right)^2}{\gamma}\right] \cos\left[\omega_p \left(t - t_s\right) + \psi\right]$$

where $\omega_p = 2\pi f_p$, $t_s = 0.45\frac{\gamma}{f_p}$ . The dominant frequency of the impulse is $f_p = 0.45$, $\psi = \frac{\pi}{2}$ and $\gamma = 0.0066$. If the mesh has 125 nodes, the matrix $A$ of 1 has the structure shown in Fig.2.

Fig. 2.   Sparsity of the matrix $A$.

It has $375 \times 375 = 140625$ entries, but only 10135 are non zero.
The pattern is symmetric, but if we consider symmetric blocks associated with boundary nodes (in red box of Fig. 2), their entries can be different. For example, let be $i = 37$ and $j = 38$. The block $(i, j)$ is

$$\begin{pmatrix} 0.00104686458333 & -0.00000000390625 & -0.00000000390625 \\ -0.00000000390625 & 0.00105208072917 & -0.00000000390625 \\ -0.00000000390625 & -0.00000000390625 & 0.00108333072917 \end{pmatrix}$$

while the symmetric block $(j, i)$ is

$$\begin{pmatrix} 0.00104686458333 & -0.00000000390625 & -0.00000000000000 \\ -0.00000000390625 & 0.00105208072917 & -0.00000000390625 \\ 0.00000000000000 & -0.00000000390625 & 0.00108333072917 \end{pmatrix}$$

Note that the *spectral radius* of the Gauss-Seidel iteration matrix is $\rho_{GS} = 0.34973375855317$.

In the Table 1, we report the values of numerical and theoretical solutions, the relative error in 2-norm between them obtained varying the number of time integrations (Time Iter.).

If we change nodes number of the mesh (Nodes) or number of time integrations (Time Step), we obtain the solution in seconds (time/s) and iterations (GS Iter) reported in the Table 2.

8

Table 1.  Error between analytical and numerical solution.

| Time Iter. | Error | Theoretical Solution | Numerical solution |
|---|---|---|---|
| 100 | 4.942183521088E-003 | 1.671036533792E-003 | 5.199331317508E-003 |
| 200 | 1.210001712110E-002 | 5.813196916902E-003 | 1.344644672612E-002 |
| 300 | 1.700030576863E-002 | 1.142001371967E-002 | 2.047757582065E-002 |
| 400 | 2.286766208306E-002 | 1.792944314502E-002 | 2.783371121225E-002 |
| 500 | 2.994038160437E-002 | 2.527536467161E-002 | 3.368659346342E-002 |
| 600 | 3.843501644295E-002 | 3.389218363222E-002 | 3.676018980363E-002 |
| 700 | 5.093770791991E-002 | 4.461474660847E-002 | 3.975634523634E-002 |
| 800 | 6.797806578970E-002 | 5.847514713198E-002 | 4.544160832560E-002 |

The Table 2 shows that increasing the number of time step, ie the simulation range, the iteration number of Gauss Seidel method is almost constant, since the matrix iteration does not change over time.

Table 2. Seconds and Iterations taken by Gauss-Seidel

| Nodes | Time Step | time/s | GS Iter |
|---|---|---|---|
| 75 | 10 | 1.692 | 11 |
| 75 | 150 | 26.979 | 12 |
| 75 | 500 | 88.127 | 12 |
| 125 | 10 | 3.195 | 12 |
| 125 | 150 | 50.502 | 13 |
| 125 | 500 | 174.852 | 12 |
| 1000 | 10 | 24.686 | 8 |
| 1000 | 150 | 390. | 8 |
| 1000 | 500 | 1321.83 | 8 |

Summarizing, for this kind of matrices the spectral radius of the Gauss-Seidel iteration matrix is near to zero and the method converges very fast. Literature is rich in algorithms optimized for sparse matrices, like GMRES, GC. They store the sparse matrix in compact format starting from the initial full one.

In our case, we compute only non zero elements and we store them on the fly in CSR format, because the matrix structure is known a priori.

## 5. Conclusion

Our final goal is to realize a 3D code able to describe the dynamics of the interaction between the seismic radiation caused by an earthquake and the near-surface geology. For this reason, in this paper we studied the structure of the matrices deriving from seismological problem like 1 and discretized by FEM.

The Gauss Seidel method results the simplest but finest solver, because it converges in a small number of iterations, as it is shown in Table 2. The good results obtained do not justify the use of other solver like GMRES and GC, but rather, they put forward the idea of an algorithm based on hybrid block Jacobi-Gauss Seidel method.

The use of parallel computing make feasible to simulate seismological IBVP more and more realistic, that requires a wide domain, a high frequency, the introduction of dissipation term, and so on.

## REFERENCES

1. J. Carcione, F. Poletto, D. Gei, 3-D Wave simulation in anelastic media using the Kelvin-Voigt constitutive equation, *Journal if Computational Physics*, **196** (2004), 282–297

2. A. Caserta , A time domain finite-difference technique for oblique incidence of antiplane waves in heterogeneous dissipative media, *Annals of Geophysics*, **41** (1998), pp. 617–631.

3. R. Clayton, B. Engquist, Absorbing boundary conditions for acoustic and elastic wave equations, *Bull. Seismological Soc. America*, **67** (1977), pp. 1529–1540.

4. R. Clayton, B. Engquist, Absorbing boundary conditions for wave-equation migration, *Geophysics*, **45** (1980), pp. 895–904.

5. H. Emmerich, M. Korn, Incorporation of attenuation into time-domain computations of seismic wave fields, *Geophysics*, **52** (1987), pp. 1252–1264.

6. B. Engquist, A. Majda. Absorbing Boundary Conditions for the Numerical Simulation of Waves, *Math. Comp.*, **31** (1977), pp. 629–651.

7. B. Engquist, A. Majda, Radiation Boundary Conditions for Acoustic and Elastic Wave Calculations, *Comm. Pure Appl. Math.* **32** (1979), pp. 313-357.

8. R. W. Graves, Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bulletin of the Seismological*

*Society of America*, **86** (1996), 1091-1106

9. A. Quarteroni, A. Tagliani, E. Zampieri, Generalized Galerkin approximation of elastic waves with absorbing boundary conditions, *Comput. Methods Appl. Mech. Engeg.* **163** (1998), pp. 323-341.

10. V. Ruggiero, P. Lanucara, M. P. Busico, A. Caserta, B. Firmani, Numerical Modelling of the ground motion :a parallel approach for finite element method, *Applied and Industrial Mathematics in Italy Proceedings of 7th Conference, World Scientific*, (2004) pp. 487-495.

11. W. D. Smith, The Application of Finite Element Analysis to Body Wave Propagation Problems, *Geophysical Journal International*, **42**(2) (1975), pp. 747–768

12. R. Stacey, Improved Transparent Boundary Formulations for the Elastic-Wave Equation *Bull. Seismological Soc. America*, **78** (1988) pp. 2089–2097.