DIOMRES (k, m): AN EFFICIENT METHOD BASED ON KRYLOV SUBSPACES TO SOLVE BIG, DISPERSED, UNSYMMETRICAL LINEAR SYSTEMS

Eli de la Torre Vega * & Mario César Suárez Arriaga **

* Instituto de Investigaciones Eléctricas, Int.Ap. Palmira, Cuernavaca, Mor., México
** Universidad Michoacana SNH, 58090 Morelia, Michoacán, México. Fax (43) 14 4735

ABSTRACT

In geothermal simulation processes, MULKOM uses Integrated Finite Differences to solve the corresponding partial differential equations. This method requires to resolve efficiently big linear dispersed systems of non-symmetrical nature on each temporal iteration. The order of the system is usually greater than one thousand and its solution could represent around 80% of CPU total calculation time. If the elapsed time solving this class of linear systems is reduced, the duration of numerical simulation decreases notably. When the matrix is big ($N \geq 500$) and with holes, it is inefficient to handle all the system's elements, because it is perfectly figured out by its elements distinct of zero, quantity greatly minor than N^2. In this area, iteration methods introduce advantages with respect to gaussian elimination methods, because these last replenish matrices not having any special distribution of their non-zero elements and because they do not make use of the available solution estimations. The iterating methods of the Conjugated Gradient family, based on the subspaces of Krylov, possess the advantage of improving the convergence speed by means of preconditioning techniques. The creation of DIOMRES(k,m) method guarantees the continuous descent of the residual norm, without incurring in division by zero. This technique converges at most in N iterations if the system's matrix is symmetrical, it does not employ too much memory to converge and updates immediately the approximation by using incomplete orthogonalization and adequate restarting. A preconditioned version of DIOMRES was applied to problems related to unsymmetrical systems with 1000 unknowns and less than five terms per equation. We found that this technique could reduce notably the time needful to find the solution without requiring memory increment. The coupling of this method to geothermal versions of MULKOM is in process.

INTRODUCTION

The MULKOM family of computer modules for simulating the flow of mass and heat in reservoirs (Pruess, 1988), uses a numerical method that generates big systems of algebraic linear equations. These systems could have some few elements different from zero irregularly distributed in the corresponding matrix. The resolution of such sparse unsymmetrical large systems represents typically 80% of the total CPU time in any simulation, which could last from some minutes to some days, depending on the computer and on the class of problem. Some years ago the linear systems were resolved by gaussian elimination or by factorization. But in this type of dispersed systems, the necessary memory increases a lot, without taking advantage of available knowledge about the possible solution's location. The Jacobi and Seidel iterating methods do not have this inconvenient, but they can only improve one variable simultaneously at each iteration.

The Conjugate Gradient iterating procedures vary all the variables at the same time, seeking the best directions of change and the best longitude of movement. The method we are introducing resolves approximately linear systems of the form: $\mathbf{A}\,\mathbf{x} = \mathbf{b}$, where $\mathbf{A}$ is the matrix of coefficients, $\mathbf{x}$ is the unknown vector and $\mathbf{b}$, the information vector.

Let $\mathbf{x}^*$ be the exact solution; for an approximation $\mathbf{x}_n$ corresponding to the n^{th} iteration, the error is given by the vectorial equation $\epsilon_n = \mathbf{x}^* - \mathbf{x}_n$. The residual is defined by $\mathbf{r}_n = \mathbf{b} - \mathbf{A}\,\mathbf{x}_n$. Both become zero when the approximation attains the exact solution: $\mathbf{x}_n = \mathbf{x}^*$. The methods of the Conjugate Gradient family (Reid, 1971), which contributed with several ideas to the construction of our method, are: ORTHOMIN (Eisenstat, 1983), DIOM (Saad, 1984) and GMRES (Saad, 1986).

THE DIOMRES (k, m) METHOD

DIOMRES generates approximations that take advantage of an initial value ($\mathbf{x}_0$), combining a set of linearly independent vectors ($\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_n$), in a similar way as in the Conjugated Gradient method:

$$x_n = x_0 + \sum_{j=1}^{n} v_j y_j = x_0 + V_n Y_n \quad \ldots\ldots\ldots\ldots \quad (1) \ ,$$

where the columns of matrix $\mathbf{V}_n$ are the searching directions $\mathbf{v}_j$ and the coordinates y_j are chosen by minimizing the residual's squared norm $R_m = r_m^T r_m$ instead of the error's A-norm $\| \varepsilon \|_A^2 = \varepsilon^T A\, \varepsilon$ which requires that $\mathbf{A}$ be symmetrical and definite positive to converge. Including the advantages of two well-known techniques, DIOMRES(k,m)

was developed, by modifying the method GMRES(m) with ideas from DIOM(k). Following the Conjugate Gradient technique, which produces orthogonal residuals with respect to the basis: ($\mathbf{V_n}^T \mathbf{r_n} = \mathbf{0}$), the method DIOM(k) finds the coordinates solving the system:

$$H_n Y_n = V_n^T A V_n Y_n = V_n^T r_0 \quad \ldots\ldots\ldots\ldots \quad (2) \ ,$$

where the elements $h_{i,n} = \mathbf{v_i}^T (\mathbf{A}\ \mathbf{v_n})$ are the required factors to orthogonalize $\mathbf{Av_n}$ with the previous vectorial basis and $h_{n+1,n}$ is needed to normalize:

$$v_{n+1} = (Av_n - \sum v_i h_{i,n}) / h_{n+1,n} \quad \ldots\ldots\ldots\ldots \quad (3)$$

The initial vector $\mathbf{v_1}$ is taken normalizing the initial residual: $v_1 = r_0 / \| r_0 \|$ and arriving to:

$$V_n = V_n H_n^* \qquad V_n^T r_0 = \| r_0 \| \quad \ldots\ldots\ldots\ldots \quad (4) \ ,$$

where $\mathbf{e_1}$ has an element equal to one in the first component and zeros in all the others. $\mathbf{H^*_n}$ is a matrix formed by the first n columns of matrix $\mathbf{H_{n+1}}$. Generating the basis in this way, $\mathbf{H_n}$ has null all its elements below the first inferior diagonal ($h_{i,j}=0$, if $i>j+1$) and if the matrix $\mathbf{A}$ is symmetrical, $\mathbf{H_n}$ also will be; this means that only is necessary to orthogonalize $\mathbf{Av_n}$ with the two previous vectors ($\mathbf{v_n}$ and $\mathbf{v_{n-1}}$) in order to get the same outcomes.

DIOMRES (k,m) also uses this basis but, minimizing the norm of the residual, obtains always better approximations than the foregoing. In this way, this technique could arrive to be as exact as is allowed by the computer's rounding error. Provided that vectors $\mathbf{v_{j+1}}$ are obtained as combinations of $\{\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_j}, \mathbf{Av_j}\}$, the technic originates that the subspace generated by the basis $\{\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_{n+1}}\}$ be the same as the space generated by $\{\mathbf{v_1}, \mathbf{Av_1}, \mathbf{A^2v_1}, \ldots \mathbf{A^nv_1}\}$, which is known as a **Krylov** subspace of order n+1. These last vectorial generators have the property of being linearly independent when the matrix $\mathbf{A}$ is non singular.

GMRES(m) contributed with the factorization: $H_m^* = Q_m U_m^*$ where $\mathbf{Q_m}$ is an adequate orthogonal matrix ($Q_m^T Q_m = I$). $\mathbf{U^*_m}$ is a matrix composed by a superior triangular matrix $\mathbf{U_m}$,; its line $(m+1)^{th}$ contains only zeros, in order to conceive the squared norm of the residual before calculating the new vectorial residual. Substituting the definition of the residual, equations 1 and 4, together with the orthogonality of the basis, plus this factorization and the orthogonality of $\mathbf{Q_n}$, we arrive to:

$$R_n = r_n^T r_n = Z_n^T Z_n + z_n^{*2} - 2 Y_n^T U_n^T Z_n + Y_n^T U_n^T U_n Y_n \quad \ldots (5)$$

whose minimal is obtained from: $\mathbf{Y_m} = \mathbf{U_m}^{-1}\ \mathbf{Z_m}$ where:

$$Q_m^T \| r_0 \| e_1 = \left[Z_m^T \mid z_m^* \right]^T$$

DIOMRES(k,m) also uses this factorization, but utilizing an incomplete orthogonalization; the way to update the solution is also different. In GMRES(m):

$$x_n = x_0 + V_n (U_n^{-1} Z_n) = x_0 + V_n Y_n \quad \ldots\ldots\ldots \quad (6) \ ;$$

this requires to wait for the m^{th} iteration to accomplish the backward substitution and resolve $\mathbf{Y_n}$; and then update the solution; this needs to stock the m vectors of the basis $\mathbf{v_j}$. Whereas, DIOMRES(k,m) performs:

$$x_n = x_0 + (V_n U_n^{-1}) Z_n = x_0 + W_n Z_n = x_{n-1} + w_n z_n \quad \ldots (7);$$

where vectors $\mathbf{w_j}$ are solved by means of a forward substitution when calculating the columns of $\mathbf{U_n}$. These vectors no longer change, allowing to update the solution immediately.

Among the analysed methods to solve several kinds of linear problems, the most rapid was GMRES(m) in some unsymmetrical cases, and ORTHOMIN (k) for the indefinite symmetric and the other non-symmetric cases. Both methods minimize the squared norm of the residual. GMRES(m) solved the memory problem using frequent restarts, and ORTHOMIN(k) using incomplete orthogonalization, which is exact in the symmetric case. However, indefinite matrix systems exist where ORTHOMIN(k) incurs in a division by zero before arriving to the solution and could lose its convergence speed if an inferior value for k is used. DIOMRES(k,m) doesn't have this risk. This technique solves the memory problem choosing a small value for k and the loss of speed is prevented by restarting the process in an opportune moment. If the matrix is symmetric, k =2 and m =N are always used.

DIOMRES(k,m) ALGORITHM

Data needed by DIOMRES are: matrix of coefficients (**A**), vector (**b**), an initial approximation ($\mathbf{x_0}$), maximum number of iterations, an acceptable error's norm (tolerance), number (k) of orthogonalizations to do and a restart period (m). The initial residual and its norm are calculated first:

$$r_0 = b - A x_0 \ ; \quad z_1^* = \| r_0 \| \quad \ldots\ldots\ldots \quad (8) \ .$$

If its value is sufficiently accurate, $\mathbf{x_0}$ is accepted as the solution. Else, the residual is normalized to get the first vector of the basis: $v_1 = r_0 / \| r_0 \|$ ... (9) .

From this moment, at every m^{th} iteration of this cycle, the following steps are carried out:

a) the n^{th} column of $\mathbf{H_m}^*$ and the new basis are calculated:

$$h_{i,n} = \begin{cases} 0, & i \le n-k \qquad \bar{v}_{n+1} = A v_n - \sum_{i=1}^{n} v_i h_{i,n} \\ v_i^T A v_n, & n-k < i < n \\ \| \bar{v}_n \|, & i = n+1 \qquad v_{n+1} = \bar{v}_{n+1} / h_{n+1,n} \end{cases} \quad (10)$$

b) The rotations detected before are applied to the new column of $\mathbf{H}_m^*$; for j=1,2, ... , n-1, and $h^*_{1,n} = h_{1,n}$:

$$\begin{bmatrix} u_{j,n} \\ h^*_{j+1,n} \end{bmatrix} = \begin{bmatrix} q_{1,j} & q_{2,j} \\ -q_{2,j} & q_{1,j} \end{bmatrix} \begin{bmatrix} h^*_{j,n} \\ h_{j+1,n} \end{bmatrix} \quad \text{..... (11)}$$

c) The new rotation data are found:

$$u_{n,n} = \sqrt{h^2_{n,n} + h^2_{n+1,n}}; \quad q_{1,n} = \frac{h_{n,n}}{u_{n,n}}; \quad q_{2,n} = \frac{h_{n+1,n}}{u_{n,n}} \quad \text{... (12)}$$

d) This last rotation is applied to the right side of the system $\mathbf{H}^*$:

$$\begin{bmatrix} z_n \\ z^*_n \end{bmatrix} = \begin{bmatrix} q_{1,n} & q_{2,n} \\ -q_{2,n} & q_{1,n} \end{bmatrix} \begin{bmatrix} z^*_{n-1} \\ 0 \end{bmatrix} \quad \text{... (13)}$$

e) A new vector $\mathbf{w}_n$ is found and the solution is updated:

$$w_n = \frac{(v_n - \sum_{j=1}^{n-1} w_j u_{j,n})}{u_{n,n}}; \quad x_n = x_{n-1} + w_n z_n \quad \text{. . . (14)}$$

f) If available iterations were finished or if desired accuracy ($|z^*_n| \leq Tol$) was reached, $\mathbf{x}_n$ is accepted as the solution.

g) If the cycle of m iterations was completed, the initial value becomes $\mathbf{x}_0 = \mathbf{x}_m$ and the algorithm is restarted.

More details about this algorithm are given in de la Torre (1990). Information on the previous methods and the manner of using preconditioning system techniques to accelerate convergence are also found in the cited document.

PROPERTIES OF DIOMRES(k,m)

The principal properties of this method are:

i) DIOMRES(m,m) produces the same approximations as GMRES(m).
ii) DIOMRES(k, ∞) produces the same approximations as ORTHOMIN(k-1).
iii) The termination in no more than N iterations is guaranteed using k =N when the matrix is symmetric.
iv) Every approximation minimizes the squared norm of the residual into the subspace generated by the last k vectors of the basis and transferred by the latest approximation. This allows to attain the maximum accuracy permitted by computer's rounding.
v) Although, for a same value of k, DIOMRES(k,m) could be in disadvantage in front to another methods (table 1), using an adequate restart period (m), generally allows to reduce the needful value of k achieving a minor requirement of auxiliary memory and time.
vi) DIOMRES(k, ∞) and DIOM(k) have the same time and memory requirements but, as long as DIOMRES(k,m) minimizes the residual, it always generates improved approaches and finishes swifter.
vii) Approximations are updated in each iteration.
viii) It is applicable to both symmetric (and takes advantage of it) and unsymmetric systems.

NUMERICAL EXPERIMENTS

In the following experiments the convergence speed of the aforementioned techniques will be compared and the results illustrated in log-linear graphs. All numerical experiments were performed in a Micro VAX 11/730 employing a rounding unit of 1.39×10^{-17} approximately, in double precision. The vector zero was always taken as the initial approached value.

1.- The first test was carried out with a linear system arisen when resolving numerically the equation of Laplace in two dimensions $\nabla^2 F(x,y) = 0$, using finite differences in a plane region with 30 divisions in the X axis and 50 in the Y axis. The matrix is of order 1500, symmetric and definite positive. Figure 1 shows the graphs obtained during the solution of this system. For this problem, the most rapid method is Conjugate Gradient, (this type of matrix is its specialty). However, DIOMRES(2,N) doesn't stay so far. GMRES(5) resulted to be the worst technique for this type of problems.

2.- The second experiment was taken from a problem proposed by Saad (1984); it corresponds to finite differences solution of the equation: $-\nabla^2 f(x,y) + c\partial f/\partial x + bf(x,y) = g(x,y)$, on a plane with 10 divisions in X and 20 in Y. The resultant system, of order 200, is unsymmetric and indefinite. Figure 2 shows that ORTHOMIN(14) takes a long time to converge and requires much memory. Conjugate Gradient shoots up, then converges in penultimate place. DIOMRES(5,66) results the fastest method, followed very close by DIOM(5) and GMRES(9).

3.- The third test was fulfilled with a linear system arose in a real application, when solving a non linear partial differential equation, from the Electrical Research Institute of Mexico (de la Torre, 1990). This is the type of system that introduces more difficulties and is a good example where Conjugate Gradient diverges and DIOM (k) requires much time and memory to converge. The corresponding matrix is unsymmetrical, indefinite, highly dispersed (5 terms per equation) and of order 961. The methods were proven first without preconditioning. Figure 3 demonstrates that unsymmetrical cases exist where Conjugate Gradient diverges without hope whatever. DIOM(4) fluctuates and is the slowest, while the other techniques differ little in speed.

TABLE 1.- COMPARISON OF TIME AND MEMORY REQUIREMENTS IN DIFFERENT METHODS.

Method	No. Products / Iteration (@)	Memory	Ref.
CONJUGATE GRADIENT	$MT+5N$	4N	Hestenes,1952 Reid, 1971
ORTHOMIN(k)	$MT+(3k+4)N$	(2k+3) N	Stanley, 1983
GMRES (k)	$MT+(k+3+\frac{1}{k})N$	(k+2) N	Saad, 1986
DIOM(k)	$MT+(3k+2)N$	(2k+3) N	Saad, 1984
DIOMRES(k,m)	$\frac{m+1}{m}MT+(3k+2-\frac{k(3k-1)}{2m})N$	(2k+3) N	De la Torre, 1990

(@) N is the number of equations; MT indicates the number of non-zero matrix elements.

4.- The same previous problem was solved in the fourth experiment, but now with preconditioning. All the methods improve their speed, however DIOMRES(2,4) results unquestionably, the fastest and with its superior convergence speed, it didn't require more memory than the rest.

CONCLUSIONS

We introduced a brief description and some numerical results showing the typical behavior of DIOMRES(k, m) when confronted to Conjugate Gradient, GMRES(k), DIOM(k) and ORTHOMIN(k), under different conditions of symmetry, spectrum (set of eigenvalues) and preconditioning. As is appreciated in the graphs, the adequate preconditioning election, the restart period and the number of orthogonalizations, allows us to solve big, unsymmetrical, sparse linear systems fast and with few memory. The method presented herein takes advantage of the available estimates of the solution avoiding the stalemate provoked by insufficient orthogonalization and taking the advantage, with symmetric matrices, of $\mathbf{Av}_n$ orthogonalization with the two more recent vectors. The adequate preconditioning technic generally diminishes the time required to find the solution in all methods. Linear systems exist where DIOMRES is flagrantly faster than the best tested methods. Being a method that minimizes the norm of the residual in each iteration, it is possible to reach all the exactness permitted by the computer's processor. The coupling of this method to MULKOM simulator is in process.

REFERENCES

-**Arnoldi, W.E.**; 1951. "*THE PRINCIPLE OF MINIMIZED ITERATION IN THE SOLUTION OF THE MATRIX EIGENVALUE PROBLEM*".Quart. Appl. Math., 9, (pp.17-29).

-**De la Torre V., E.** 1990. "*SOLUCION DE SISTEMAS LINEALES GRANDES Y HUECOS MEDIANTE MÈTODOS ITERATIVOS BASADOS EN LOS ESPACIOS DE KRYLOV*". Bc.Sc.Thesis on Mathematical Physics, Michoacán University, Morelia, Mich.; México (156 p.). This work was supported by the National Council of Science and Technology.

-**Golub,G.H. & Van Loan,C.**1983. "*MATRIX COMPUTATIONS*". The John Hopkins U.Press Baltimore Maryland, (pp.362-379).

-**Hestenes,M. & Stiefel,E.**; 1952. "*METHODS OF CONJUGATE GRADIENTS FOR SOLVING LINEAR SYSTEMS*". J. Res. Nat. Bur. Standards, 49, (pp. 409-436).

-**Lanczos,C.**1950. "*AN ITERATION METHOD FOR THE SOLUTION OF THE EIGENVALUE PROBLEM OF LINEAR DIFFERENTIAL AND INTEGRAL OPERATORS*". J. Res. Nat. Bur. Standards, 45, (pp. 255-282).

-**Paige,C. & Saunders,M.A.** 1975. "*SOLUTION OF SPARSE INDEFINITE SYSTEMS OF LINEAR EQUATIONS*". SIAM J. Anal., Vol. 12, Num. 4, (pp. 617-629).

-**Pruess, K.** (1988). "*SHAFT, MULKOM, TOUGH: A SET OF NUMERICAL SIMULATORS FOR MULTIPHASE FLUID AND HEAT FLOW* GEOTERMIA - Rev.Mex.de Geo.,Vol. 4, No.1, pp. 185-202.

-**Reid,J.** "*ON THE METHOD OF CONJUGATE GRADIENTS FOR THE SOLUTION OF LARGE SPARSE SYSTEMS OF LINEAR EQUATIONS*". Proc. Conf. on Large Sparse Sets of Linear Equations, Academic Press, New York, 1971, (pp. 231 - 254).

-**Saad,Y.**1981. "*KRYLOV SUBSPACE METHODS FOR SOLVING LARGE UNSYMMETRIC LINEAR SYSTEMS*".Math.Compt.,37,(pp.105-126).

-**Saad.Y.**; 1984. "*PRACTICAL USE OF SOME KRYLOV SUBSPACE METHODS FOR SOLVING INDEFINITE AND NONSYMMETRIC LINEAR SYSTEMS*". SIAM J. Numer. Anal., 5, (pp 203-227).

-**Saad,Y. & Schults,M.**1986. "*GMRES: A GENERALIZED MINIMAL RESIDUAL ALGORITHM FOR SOLVING NONSYMMETRIC LINEAR SYSTEMS*". SIAMJ.Scl.Stat.Com., Vol.7,Num.3,(pp.856-869).

-**Stanley,C.,Eisenstat,H.,ELman,C.,& Shultz, M.** 1983. "*VARIATIONAL ITERATIVE METHODS FOR NONSYMMETRIC SYSTEMS OF LINEAR EQUATIONS*"., SIAM J. Numer Anal. Vol. 20 No.2.

-**Young,D. M.**; 1971. "*ITERATIVE SOLUTION OF LARGE LINEAR SYSTEMS*". Academic Press.

Fig. 1.- Computational Work Vs. Residual For a Symmetric, Definite Positive Linear System with 1500 equations.

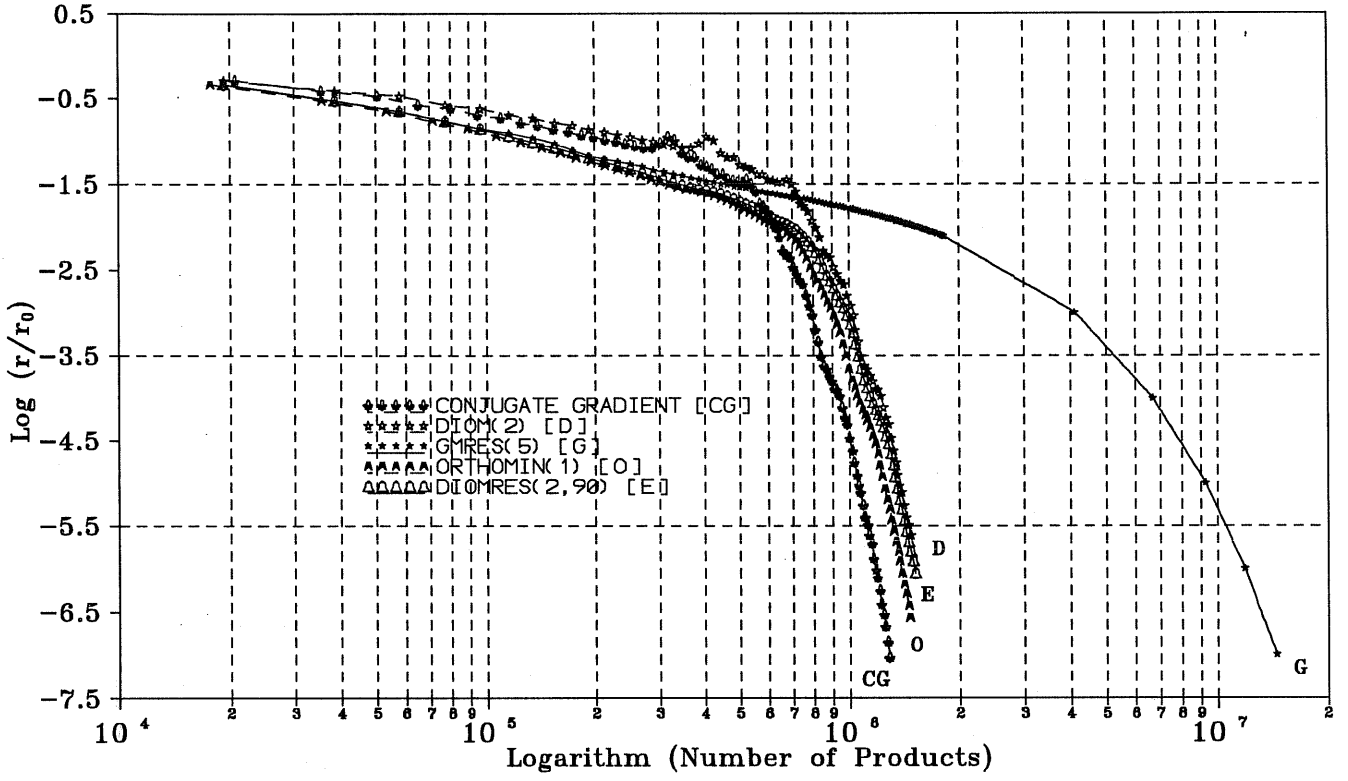

Fig. 2.- Computational Work Vs. Residual For an Unsymmetric, Indefinite Linear System with 200 equations.

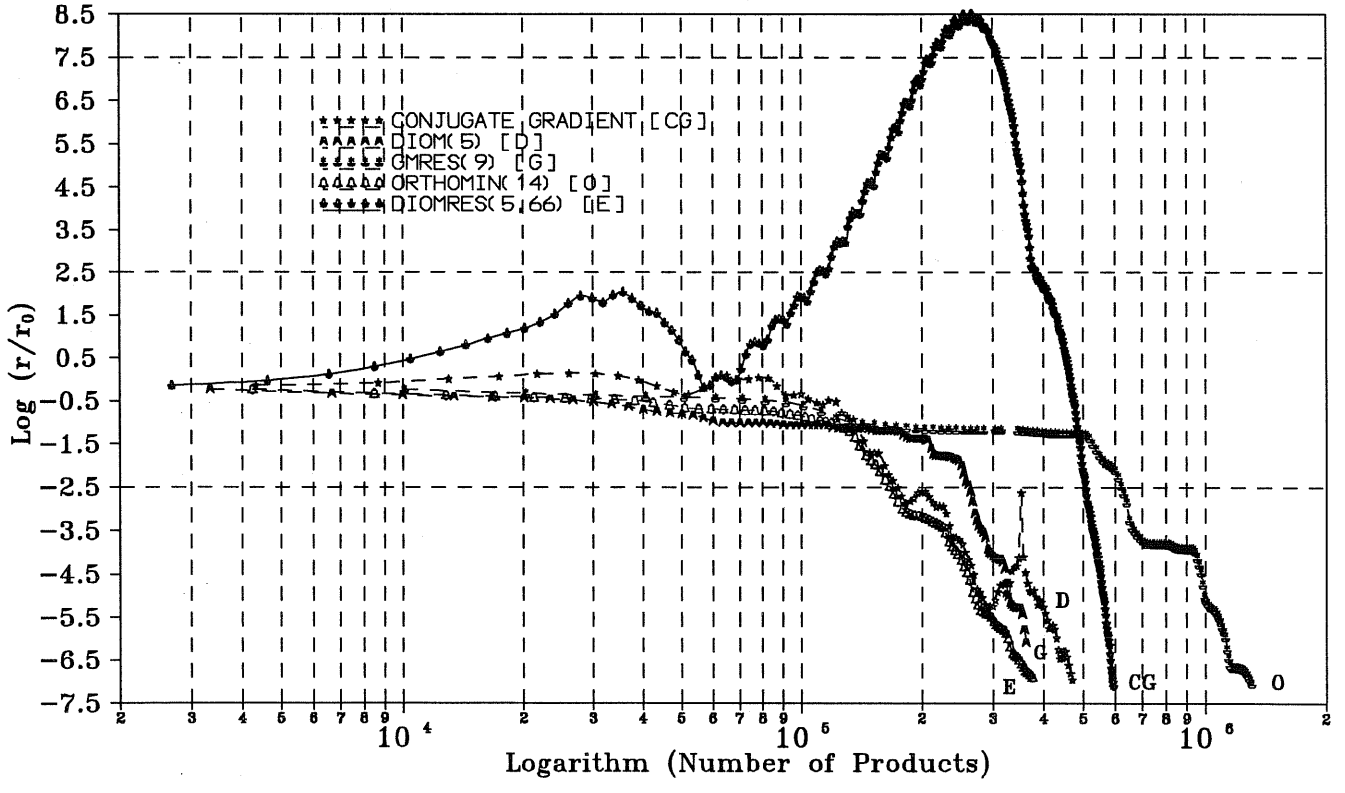

Fig. 3.- Computational Work Vs. Residual For an Unsymmetric, Indefinite Linear System with 961 equations, without Preconditioning.

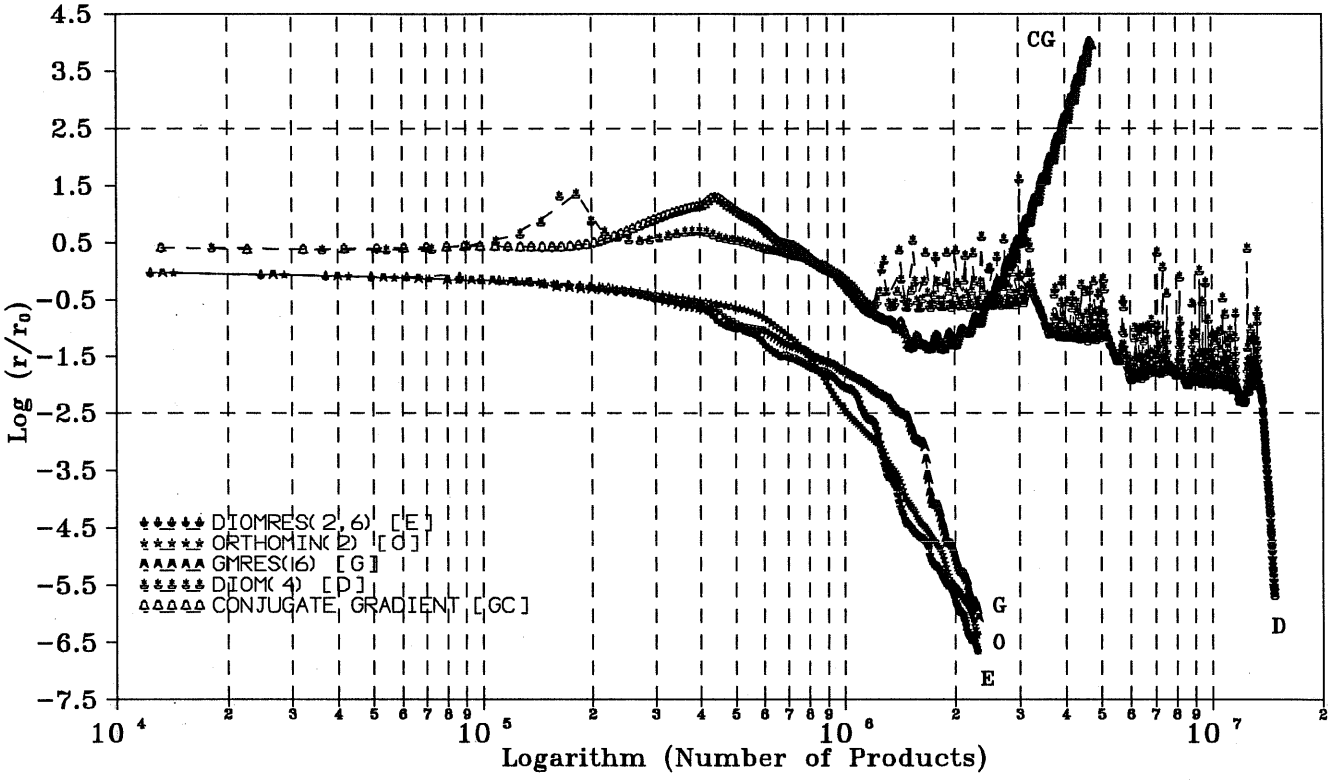

Fig. 4.- Computational Work Vs. Residual For an Unsymmetric, Indefinite Linear System with 961 equations, with Preconditioning.