



## Vectorizing finite element methods

Jocelyne Erhel, A. Lichnewsky, F. Thomasset

► **To cite this version:**

Jocelyne Erhel, A. Lichnewsky, F. Thomasset. Vectorizing finite element methods. RR-0383, INRIA. 1985. <inria-00076173>

**HAL Id: inria-00076173**

**<https://hal.inria.fr/inria-00076173>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**IRIA**

**CENTRE DE ROCQUENCOURT**

Institut National  
de Recherche  
en Informatique  
et en Automatique

Domaine de Voluceau  
Rocquencourt  
BP 105  
78153 Le Chesnay Cedex  
France  
Tel (3) 954 90 20

# Rapports de Recherche

N° 383

## VECTORIZING FINITE ELEMENT METHODS

Jocelyne ERHEL  
Alain LICHNEWSKY  
François THOMASSET

Mars 1985

## VECTORIZING FINITE ELEMENTS METHODS

Jocelyne ERHEL  
INRIA, Rocquencourt, France

Alain LICHNEWSKY  
Université Paris Sud, Orsay, France  
and INRIA, Rocquencourt, France

François THOMASSET  
INRIA, Rocquencourt, France

### RESUME

Dans ce rapport, nous décrivons les techniques développées pour mettre en oeuvre des algorithmes d'algèbre linéaire sur des calculateurs vectoriels tels que CRAY-1S, Fujitsu VP200, Hitachi S810, ... Les algorithmes étudiés traitent des matrices creuses issues de méthodes d'éléments finis. En particulier, nous nous intéressons ici à l'algorithme du gradient conjugué préconditionné. Nous introduisons une renumérotation des inconnues pour supprimer les dépendances de données et augmenter le parallélisme. Le cas particulier des structures "régulières" de matrices correspond aux méthodes de différences finies. Dans le cas général, il faut exhiber des vecteurs assez longs pour compenser le coût des indirections.

Afin de systématiser la recherche de rangements parallèles de nos matrices et compléter les techniques heuristiques que nous venons de mentionner, nous introduisons une formulation en termes d'optimisation. Pour la résolution de ces problèmes nous mettons en oeuvre une méthode de Monté-Carlo basée sur le principe de la "trempe" en mécanique statistique.

### ABSTRACT

This paper deals with some techniques we have developed to implement - on vector processors such as CRAY-1S, Fujitsu VP200, Hitachi S810, ... - linear algebra algorithms for sparse matrices associated with finite element methods ; the particular algorithm we consider here is preconditioned conjugate gradient. We introduce a renumbering of unknowns in order to remove data dependencies and to increase parallelism. We consider first the case of "regular" matrix structures, encountered in the context of finite difference methods. In the general case of finite element methods, we must find vectors long enough to mask the overhead inherent to the indirect indexation.

In order to deal systematically with the search of parallel matrix structures and to complete heuristic techniques mentioned above, we introduce a formulation in terms of optimisation. We solve this problem by a simulated annealing method, derived by analogy with statistical physics of random systems.

## VECTORIZING FINITE ELEMENTS METHODS

Jocelyne ERHEL  
INRIA, Rocquencourt, France

Alain LICHNEWSKY  
Universite Paris Sud, Orsay, France,  
and INRIA, Rocquencourt, France

François THOMASSET  
INRIA, Rocquencourt, France

Paper presented to the International Seminar on Computer Networking and Performance Evaluation, September 18-20, 1985, Tokyo, Japan.

### ABSTRACT

This paper deals with some techniques we have developed to implement - on vector processors such as CRAY-1S, FUJITSU/VP200, HITACHI/S810, ... - linear algebra algorithms for sparse matrices associated with finite element methods; the particular algorithm we consider here is preconditioned conjugate gradient. First of all we consider the case of "regular" matrix structures which we encounter in the context of finite difference methods, and we introduce a renumbering of unknowns in order to remove data dependencies. Then we extend this renumbering to the general case of finite element methods. Finally we present an optimization procedure based on Monte-Carlo techniques to improve the renumbering.

### 0. INTRODUCTION

Preconditioned conjugate gradient methods are very efficient algorithms for linear systems of equations, whose matrix is symmetric and positive definite (see /MeGo 1983/ and the bibliography therein); it involves iterated solution of linear systems, with a fixed matrix  $M$ , the so-called "preconditionner". Although solving a system with matrix  $M$  is much simpler than the original problem, it is generally lots of recurrences that inhibit parallelisation; this can be overcome by applying proper renumberings of the unknowns. Unfortunately we generally observe that, the more implicit is the preconditionner, the faster the convergence of conjugate gradient iterates.

The efficiency of the Preconditioned Conjugate techniques and its adaptation to vector processes has been the result of many studies, including Meijerink and Van der Vorst /MVV 1977, 1981/; Schreiber and Wei-Pei Tang /SWT1983/; Rogrigue and Wolitzer /RoWol1982/; Concus, Golub and Meurant /CGM1982/; Johnson, Micchelli and Paul /JMP1982/; Meurant /Meu1985/.

In the case of finite difference methods a family of renumberings is easily found by applying coloring considerations; similar structures can be obtained within the context of finite element methods using an algorithm derived from Cuthill-McKee's reordering technique for reducing the bandwidth of a finite element matrix /CMK1969/.

We are then faced with two sorts of difficulties: one is related with the structure of sparse matrices in finite element methods, which may be dealt with by calls to SCATTER/GATHER routines. The other difficulty concerns the quality of the renumberings: finding an "optimal" or even a "good" renumbering appears to be a hard problem of a combinatorial nature. We propose in the last section to apply random perturbations to the ordering of the unknowns, in a Monte-Carlo style, so as to decrease some criterium on the diagonals of the matrix.

### 1. THE PRECONDITIONED CONJUGATE GRADIENT

Suppose we need to solve the linear system:

$$(1) \quad A \cdot x = b,$$

where  $x$  and  $b$  are vectors,  $A$  is an order- $n$  symmetric, positive definite matrix. Let  $M$  be another symmetric positive definite matrix (the "preconditionner"). Then the solution  $x$  is found as the limit of the following iterative process ( $r, z, p, s$  are auxiliary vectors;  $\alpha, \beta, \text{newbeta}$  are scalars):

---

(STEP-0)(Initialization):

$r \leftarrow b - A \cdot x$   
let  $z$  be the solution of :  $M \cdot z = r$   
 $p \leftarrow z$   
 $\beta \leftarrow z \cdot r, \quad s \leftarrow A \cdot p$   
 $\alpha \leftarrow \beta / p \cdot s$

(STEP-1):

$x \leftarrow x + \alpha * p$   
 $r \leftarrow r - \alpha * s$

(STEP-2):

let  $z$  be the solution of :  $M \cdot z = r$

(STEP-3):

$\text{newbeta} \leftarrow z \cdot r$   
 $p \leftarrow z + (\text{newbeta}/\beta) * p$   
 $\beta \leftarrow \text{newbeta}$

(STEP-4):

$s \leftarrow A \cdot p$   
 $\alpha \leftarrow \beta / p \cdot s$

(STEP-5):

IF NOT CONVERGED THEN GOTO STEP-1.

---

The nicest property of this algorithm is its potentially fast convergence; this depends on spectral properties of matrices  $A$  and  $M$ : in short the closer the matrix  $M$  is from the inverse of  $A$ , the faster the convergence.

So we have some flexibility in the choice of  $M$ : the two

extreme choices are: i/  $M=A$ , in which case each iteration is equivalent to the original problem; ii/  $M=Identity$ : we have the classical conjugate gradient of Hestenes and Stiefel /HeSt1952/, which suffers from slow convergence.

We have chosen the technique introduced by Meijerink and Van der Vorst /MVV1977/ known as "Incomplete Cholesky factorization", which is described in the following section.

## 2. CHOICE OF THE PRECONDITIONNER: INCOMPLETE CHOLESKY FACTORIZATION

First let us recall that Cholesky factorization of matrix  $A$  consists in computing a lower triangular matrix  $N$  such that:

$$(2) \quad A = N \cdot N^t$$

Unfortunately the sparsity pattern of  $A$  is not conserved in matrix  $N$ :  $N$  is much more dense than  $A$  ("fill-in" phenomenon), leading to excessive storage requirements. (Typically in finite difference methods,  $A$  would be a 5- or 7-diagonal matrix, while a whole bandwidth would be filled in  $N$ ).

Thus instead of (2), we write:

$$(3) \quad A = L \cdot L^t + R,$$

where the residual matrix  $R$  and the lower matrix  $L$  are such that the structure of  $L$  is simpler than with (2). Then we shall take as a preconditionner :

$$M = L \cdot L^t.$$

In order to describe the structural constraints on  $L$ , it is useful to associate to a symmetric matrix  $A$  its labelled undirected graph:  $G(A)=(V(A),E(A))$ . The vertex set  $V(A)$  consists of  $n$  elements (one vertex per line of  $A$ ), and the edge set  $E(A)$  describes the sparsity of  $A$  in the following way:

$$(4) \quad \{i, j\} \in E(A) \text{ iff } A(i, j) \neq 0$$

Let there be given a priori the graph  $G(M)$ , which we shall note  $G(IC)=(V(A),E(IC))$ . Then the set of constraints on  $L$  and  $R$  are given by:

a/ the only nonzeros in  $L$  are those compatible with  $G(IC)$ , that is:  $L(i, j)=0$  if  $\{i, j\} \notin E(IC)$

b/ impose  $R(i, j)=0$  if  $\{i, j\} \in E(IC)$ .

This construction is possible for  $M$ -matrices or diagonally dominant matrices; for more general positive definite systems, the shifting method of /Mant1978/ can be used.

Now the graph  $G(IC)$  can be quite arbitrary, but  $G(IC)$  will be usually much smaller than the perfect elimination graph  $G(M)$ , so as to limit the amount of fill-in. A typical choice,







- 
- Initially, all nodes are unmarked
- /step 0/ Pick up a set of nodes  $F_0$  (initial front)  
Mark the nodes in  $F_0$
- /step 1/ Set  $k := k+1$   
 $F_k$  = set of unmarked nodes adjacent to nodes in  $F_{k-1}$   
Mark the nodes in  $F_k$  (kth front)
- /step 2/ If some nodes remain unmarked then go to step 1
- 

Now, the idea is to renumber the nodes within a front in order to give regular structures to submatrices  $A_i$ .

Consider for simplicity the case of triangular elements of degree 1 ( $P_1$  in 2-D domains).

If the fronts were paths, renumbering inside them would lead to tridiagonal submatrices  $A_i$ . Unfortunately, with general Finite Element meshes, this is not always true : therefore, we remove the spurious nodes from the fronts in order to obtain pathlike fronts. These nodes are numbered after all the nodes in the fronts.

Then, a zebra-like numbering of the fronts (themselves renumbered), yields a structure of the matrix similar to that of the Finite Difference case, with blocks marked E, F and G corresponding to the spurious nodes. The structure of the matrix A is given below :

$$\begin{bmatrix}
 A_1 & & & & C_2 & & & & & & F_1 \\
 & A_3 & & & B_3 & C_4 & & & & & F_2 \\
 & & A_5 & & & B_5 & C_6 & & & & F_3 \\
 & & & A_7 & & & B_7 & C_8 & & & F_4 \\
 B_2 & C_3 & & & A_2 & & & & & & F_5 \\
 & B_4 & C_5 & & & A_4 & & & & & F_6 \\
 & & B_6 & C_7 & & & A_6 & & & & F_7 \\
 & & & B_8 & & & & A_8 & & & F_8 \\
 E_1 & E_2 & E_3 & E_4 & E_5 & E_6 & E_7 & E_8 & & & G
 \end{bmatrix}$$

where  $C_i = tB_i$  ( $i=2,8$ )  
 $F_i = tE_i$  ( $i=1,8$ )

The diagonal blocks  $A_i$  are themselves tridiagonal, and this corresponds to 2-diagonal blocks in the factorized triangular matrix L. The solution of the system  $Lx = y$  could be parallelized in the same fashion as in the regular case. Now, because the submatrices  $A_i$  have not the same size, it is not possible to vectorize accross the blocks. Therefore, we consider all odd-numbered (resp. even-numbered) blocks as one tridiagonal block  $X_1$  (resp.  $X_2$ ), and we are lead to solve first-order linear

recurrences, that can be vectorized by cyclic reduction for instance.

The blocks  $B_i$  and  $C_i$ , coupling the nodes of odd and even fronts, are sparse banded submatrices, with bandwidth less than the maximum front size. We treat the submatrices  $B_i$  (resp.  $C_i$ ) as one block  $Y$  (resp.  $Z$ ), which is stored according to a sparse-diagonal-oriented scheme (storing only non-zero coefficients in diagonal order and using pointers). It is still possible to vectorize the product by the block  $Y$  or the block  $Z$ , but with the use of indirect addressing (GATHER/SCATTER operations). The performances are therefore degraded, compared to the regular case with filled diagonals and constant-increment vectors.

It should be noted that the product of the matrix  $A$  by a vector needs also to compute the product of  $Y$  and  $Z$  by a subvector.

In the next section, we introduce heuristics to improve the structure of the submatrix  $Y$  and to optimize the vectorization.

The blocks  $E_i$  and  $F_i$  containing the spurious nodes are sparse, with no regular pattern : we did not try to vectorize the operations on them (to treat them as dense blocks is slower by experience). The overhead due to those blocks remains acceptable, because, by experience, the number of spurious nodes is much smaller than the total number of unknowns (there is about a factor of 50). In the new structure of  $A$ , the blocks corresponding to the spurious nodes are  $S_i$  ( $i=1,2$ ) and  $T_i$  ( $i=1,2$ ) (coupling spurious nodes with fronts) and  $S$  (coupling spurious nodes between them).

The structure of the matrix finally used is then :

$$\begin{bmatrix} X1 & Z & T1 \\ Y & X2 & T2 \\ S1 & S2 & S \end{bmatrix}$$

Where  $A_i$  ( $i=1,2$ ) are tridiagonal  
 $Y$  is "sparse-diagonal" -  $Z = tY$   
 $S_i$  ( $i=1,2$ ) and  $S$  are sparse (by rows or columns)

Remark :

In the case of elements of higher degree, we can use multicoloring as in /SWT 1983/ to obtain a similar pattern.

Results :

- Two meshes have been tested :
- RING : a ring shaped domain with 649 unknowns.
  - BINACA : a discretization of an industrial interest, with 1702 unknowns.

We have measured execution time and speed of convergence for both Dirichlet and Neumann boundary conditions. We have compared the

present method with renumbering and restructured matrix (FRONT), with a sparse method without renumbering (SPARSE). The algorithm considered in both cases is Preconditioned Conjugate Gradient by Incomplete Cholesky. We give times in seconds on the CRAY-1S.

DOMAIN	B.C.	SPARSE			FRONT		
		# ITER	CPU( $\Delta$ )	MFLOPS/S	# ITER	CPU( $\Delta$ )	MFLOPS/S
RING(649)	D	14	0.29	1.5	21	0.07	5.8
	N	34	0.88	1.5	54	0.25	5.9
BINACA(1702)	D	30	1.92	1.5	55	0.64	5.5
	N	57	3.86	1.5	88	1.13	5.6

Table 2

We have also tried a slightly different method with another storage scheme for submatrix Y and therefore another algorithm to compute the product of Y and tY by a subvector.

We give in the table below the obtained results, both on CRAY-1S and on FUJITSU VP200. It should be noted that the program was not tuned for VP200 ; in particular, the solution of first order linear recurrences was not vectorized. (1)

DOMAIN	CRAY-1S		VP200	
	CPU( $\Delta$ )	MFLOPS/S	CPU( $\Delta$ )	MFLOPS/S
RING(649)	0.059	6.9	0.051	8.1
BINACA(1702)	0.41	7.7	0.31	10.2

Table 3

### 5. THE RATIONALE FOR AN OPTIMIZATION PHASE.

Our motivations to improve the above given ordering by an optimisation technique are twofold. First, we feel that this may ultimately lead to a systematic treatment of the subject, and that we will be able to use adequate sets of constraints to impose our structural requirements on the matrices. Our second reason is that we want to be able to construct similar methods in the 3 Dimensional case, and that our current "front" heuristics will then have to be iterated, giving first surfaces and then lines. In this process, many "spurious" points are created departing from the appropriate matrix structure. The purpose of the optimization phase is then to regain control over these irregularities.

The optimization problem can then be stated in the following way:

-----  
 (1) Up to date results will be presented at the conference

Given a matrix structure and an initial numbering of the unknowns U, which satisfies a set of constraints  $U \in U_{ad}$ . (1)

Find a numbering U which satisfies the same constraints  $U \in U_{ad}$ , while minimizing a given objective functional  $F(U)$ .

The functionals F that we have currently investigated are the following:

a/ Number of diagonals in the matrix.

b/ Weighted number of diagonals, where the weight depends on the diagonal location.

c/ -  $(\sum_{i=-n, \dots, n} (\text{number of elements on diagonal}(i))^2)$

d/ -  $\sqrt{(\sum_{i=-n, \dots, n} (\text{number of elements on diagonal}(i))^2)}$

All these functional express that we want to minimize the number of points that do not enter the by-diagonals organization of the matrix. This is quite obvious for the last two criteria also, if one keeps in mind that the total number of coefficients is invariant through all renumberings :

$\sum_{i=-n, \dots, n} \text{number of elements on diagonal}(i) = \text{constant.}$

The last one seems to express our requirement in a less direct way than the two first ones, however, our first experiments have shown its practical value. The two "obvious" first ones have the drawback that the optimization method has little or no way to sense transformations that do not result in the creation or disparition of a diagonal. The third tends to favour diagonals with a large number of elements, which is certainly one of our qualitative objectives.

This choice leaves us with a hard problem. We have not worked out the details concerning the actual complexity of our constrained problems, but there are available results pertaining to related problems :

- there seems to be no efficient way to test for bandwidth 3. Which means that, even if one of the permutations of the matrix was tri-diagonal, U may be too hard to construct for practical matrix sizes /Tarj1976/.

- The minimum bandwidth problem is known to be NP-complete.

6.A MONTE-CARLO / "SIMULATED ANNEALING" MINIMIZATION ALGORITHM.

(1) This constraints express the fact that the renumbered matrix has the structure given by figure 4.

It now seems quite natural to try to combine a heuristic construction method, which we had as a starting point, to a heuristic minimization technique of Monte-Carlo type. This approach, which is fully justified in statistical physics, has shown to be of large practical value for several hard problems /VeKil1983/. It then has mostly a heuristic justification, and we have, so far, made several experiments on simple configurations of our problem. We will now describe the method as well as typical results on our simple problems. Further work is being done in order to apply the method to full scale problems, to combine with the former heuristics, and to determine the most appropriate functionals. (2)

Our minimization algorithm exploits the following analogies with statistical physics:

atomic positions	configuration given by admissible permutations U
internal energy	objective functional F(.)
cool into low energy state	approximately minimize the functional F(.)
absolute temperature	parameter T in following algorithm

It proceeds along the following lines:

---

#### ANNEALING ALGORITHM:

Starting at step n with configuration  $U(n)$ , we select an arbitrary random transposition of indices in the matrix which transforms  $U(n)$  into an admissible configuration U.

We take  $U(n+1) = U$  (3)

if  $F(U) = F(U(n))$

or else with probability  $\exp((F(U(n)) - F(U))/T)$ .

Otherwise  $U(n+1)$  is set equal to  $U(n)$  (4).

This is repeated letting T go slowly to 0 so as to make the configurations corresponding to low values of the objective functional most probable.

---

We shall now describe the typical behaviour of this algorithm on two very simple cases. Although we plan to combine this algorithm with adaptations of Cuthil-McKee ordering in the long run, this preliminary study tests the intrinsic behaviour of the optimization process. In these two experiments, we started with the same arbitrary permutation of a 100 by 100 matrix, derived from a finite difference 5 point star on a square mesh. The selected criterion was functional (c).

---

- (2) Up to date results covering these topics will be presented at the Conference.
- (3) i.e. "we accept the new configuration"
- (4) i.e. "we reject the new candidate".

Initial state

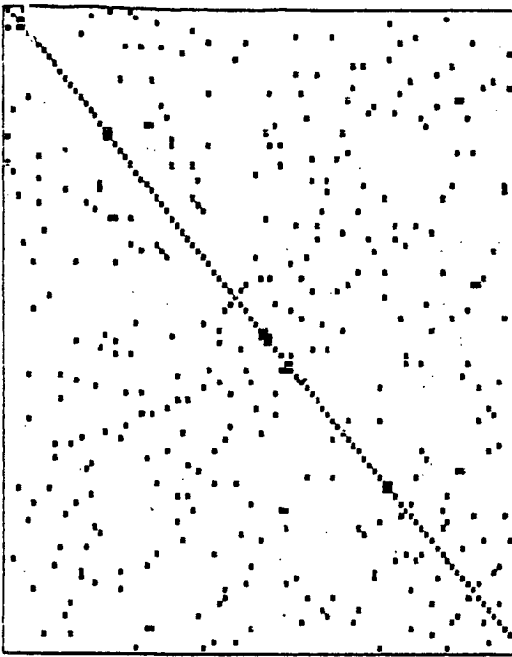


Figure 1

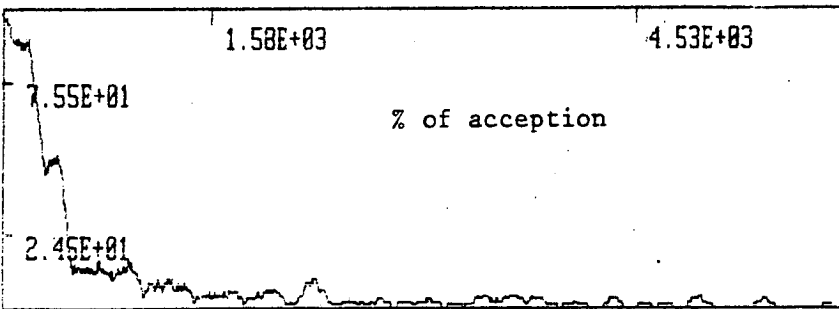
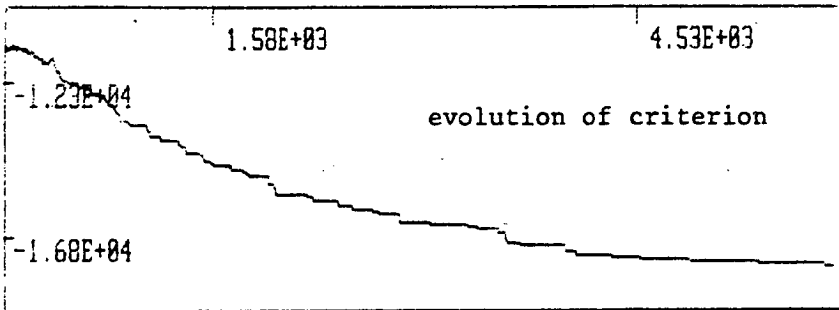


Figure 2

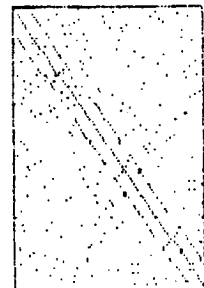
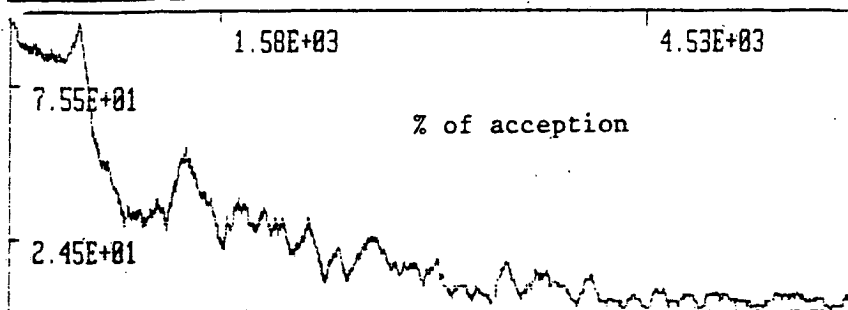
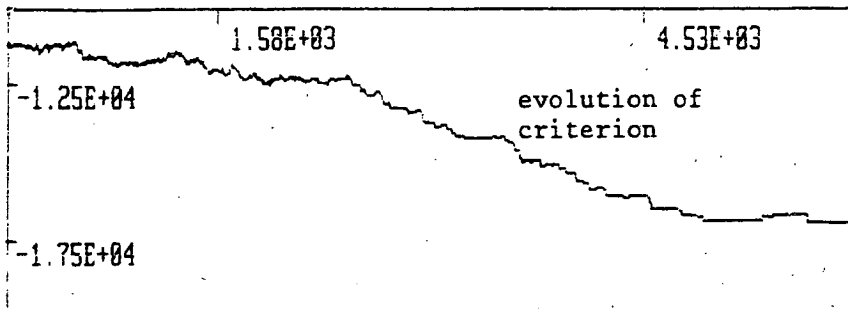


Figure 3

The minimization in these two examples were conducted as follows:

Ex1: (Figure 2 ): 200 iterations with T=100  
200 iterations with T= 25  
8500 iterations with T= 10.

Ex2: (Figure 3 ): 500 iterations with T=100  
12000 iterations with T= 25

Two aspects of the results seem interesting. First, the algorithm succeeds in reducing the optimization criterion, even if we remain quite far from the absolute minimum which was expected due to the nature of the problem. The other aspect is that the matrices show quickly a lot of structure, especially as two strongly populated diagonals are created. This structure was not apparent in the initial state (Figure 1 ) .

#### REFERENCES

/Axel 1976/

AXELSSON (O.)

"Solution of linear systems of equations: iterative methods", in Sparse Matrix Techniques, V.A. Barker Ed., Springer.

/CGM1982/

CONCUS (P.), GOLUB (G.H.), MEURANT (G.) /CGM 1982/

"Block preconditioning for the conjugate method",  
Report LBL-14856, Stanford University, Stanford CA.

/CMK1969/

CUTHILL (E.), MAC KEE (J.)

"Reducing the bandwidth of sparse symmetric matrices",  
Proc. ACM National Conference, 1969.

/Erhel1983/

ERHEL (J.)

"Parallelisation d'un algorithme de gradient conjugué  
preconditionné",  
Rapport de Recherche INRIA n0. 189, Fevrier 1983.

/HeSt1952/

HESTENES (M.R.), STIEFEL (E.)

"Methods of conjugate gradient for solving linear systems",  
NBS J. Res. 49 (1952), pp 409-436.

/JoMP1982/

JOHNSON (O.G.), C.A. MICHELLI & G. PAUL /JMP 1982/

"Polynomial preconditioning for conjugate gradient  
calculations", IBM Thomas J. Watson Research Center,  
Yorktown Heights, NY, 1982

/Lich1983/

LICHNEWSKY (A.)

"Some vector and parallel implementations for preconditioned conjugate gradient algorithms", NATO Advanced Research Workshop on High-Speed Computations, 20-22 June 1983, Julich, F.R.G.

/Mant1978/

MANTEUFFEL (T.A.)

"The shifted incomplete Cholesky factorization", SAND/78/8226, Sandia Lab., Albuquerque, NM.

/MVV 1977/

MEIJERINK (J.A.), VAN DER VORST (H.A.)

"An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix", Math. Comp., Vol. 31, pp 148-162.

/MVV 1981/

MEIJERINK (J.A.), VAN DER VORST (H.A.)

"Guidelines for the usage of incomplete decomposition in solving sets of linear equations as they occur in practical problems", J. Comp. Ph., Vol. 44, pp 134-155.

/MeGol1983/

MEURANT (G.), GOLUB (G.H.)

"Resolution numerique des grands systemes lineaires", Eyrolles, Paris, June 1983.

/Meur1985/

MEURANT (G.)

"The block preconditioned conjugate gradient method on vector computers", to appear in BIT.

/Reid1971/

REID (J.K.)

"On the method of conjugate gradients for the solution of large sparse systems of equations", pp 231-254 in "Large sparse sets of linear equations", ed. J.K. Reid, Academic Press, 1971.

/Rowol1982/

RODRIGUE (G.), WOLITZER (D.)

"Preconditioning by incomplete block cyclic reduction", Research Report UCID-19502, LLNL, Livermore, CA.

/SWT 1983/

SCHREIBER (R.), WEI-PEI-TANG

"Vectorizing the conjugate gradient method", to appear.

/Tarj1976/

TARJAN (R.E.)

"Graph Theory and Gaussian Elimination", in Sparse Matrix Computations, Bunch & Rose eds.,



Academic Press, New York.

/VDV 1983/

VAN DER VORST (H.A.)

"On the vectorization of simple ICG methods",  
First Int. Coll. on Vector & Parallel Computation,  
AFCET-GAMNI-ISINA, Paris.

/VeKi1983/

VECCHI (M.P.), KIRKPATRICK (S.),

"Global wiring by simulated annealing",  
IEEE Trans CAD, Vol CAD-2, Oct. 1983, pp 215-222

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique

