PERGAMON

Computers and Structures 73 (1999) 33–43

Computers
& Structures

www.elsevier.com/locate/compstruc

Problem-dependent preconditioners for iterative solvers in FE elastostatics

Pascal Saint-Georges*, Guy Warzee, Yvan Notay, Robert Beauwens

Service des Milieux Continus, CP 194/5 Université Libre de Bruxelles, 50 av. F.D. Roosevelt, B-1050, Bruxelles, Belgium

Received 8 November 1996; accepted 20 July 1998

Abstract

Approximate factorizations are probably the most powerful preconditioners at the present time in the context of iterative solution methods for FE structural analysis. In this contribution we focus on some aspects of the reduction method proposed previously, which allow the use of perturbed approximate factorizations. In particular, we show that it is not suitable for systems arising from discretizations with plate or shell elements. In contrast, corrected incomplete Cholesky preconditioners are shown to exhibit a much better convergence for such systems. © 1999 Civil-Comp Ltd and Elsevier Science Ltd. All rights reserved.

Keywords: iterative solution of linear systems; preconditioning; finite element method

1. Introduction

Many iterative solvers for FE structural analyses have been experienced during the last few years. They are all based on the preconditioned conjugate gradient method because of its optimal speed of convergence. The preconditioner has, however, to be carefully chosen to avoid unpredictable behaviours or even non-convergence. Nevertheless, iterative solution procedures are gaining popularity and interest amongst FE software users due to their small memory requirements: some of the popular direct solvers used by now need gigabytes of disk space to deal with problems with hardly more than 100,000 degrees of freedom. A key feature to make iterative solvers decisively more attractive than direct ones is then *robustness*.

Incomplete Cholesky (IC) factorizations can be used in many fields and often permit fast convergence provided the system matrix is a *Stieltjes* matrix, a more restrictive condition than simple positive definiteness. In previous works [1], the present authors have proposed a so-called DC-reduction extracting a Stieltjes matrix **S** from an initial positive definite stiffness matrix **K**, in order to apply an incomplete factorization to **S** instead of **K**. This procedure works as long as **S** and **K** are spectrally equivalent, which was theoretically and numerically proven in [1] for stiffness matrices arising from the discretization of plane stress/strain and solid structures. Thanks to this technique, perturbed modified IC (XIC) methods could be used instead of the basic IC preconditioner, which considerably speeds up the convergence.

In this contribution, it is shown that DC-reduction cannot be used for discretizations including rotational degrees of freedom, like those in which plate and shell elements appear. Since efficient reduction schemes have still not been found, no equivalent 'pre-processor' enabling the use of XIC methods is available. Some authors [2,3] have proposed *corrected* IC (CIC)

* Corresponding author.
E-mail address: pstgeorg@smc.ulb.ac.be (P. Saint-Georges)

0045-7949/99/$ - see front matter © 1999 Civil-Comp Ltd and Elsevier Science Ltd. All rights reserved.
PII: S0045-7949(98)00277-6

methods that are robust in the sense that they always yield a positive definite preconditioner if the factored matrix is positive definite. With this approach, a reduction is no longer necessary, but the obtained preconditioner does not take advantage of the advances in perturbed modified IC methods. It is shown here that, despite the latter remark, CIC methods are well adapted for plate and shell analysis, although XIC-DC preconditioners remain much more effective for problems including only translational degrees of freedom.

This confirms that there is for the present time no universal preconditioner: a choice has to be made to get the 'best' one depending on the nature of the considered problem. A general purpose FE software should switch between preconditioning methods, based on the type of element used.

Two classes of problems are therefore considered in the remainder of the paper:

- type T, for structures involving only translational degrees of freedom like solid, plane stress/strain and rod structures;
- type R, for discretizations where rotational degrees of freedom appear, like plate, shell and beam structures.

Efficient preconditioning could also in principle be obtained by multigrid techniques. Grid coarsening is however difficult to implement in FE structural analyses because the meshes used in practice are usually close to the coarsest ones compatible with the actual geometry. It is therefore necessary to resort to algebraic multigrid or multilevel methods presently under development. We expect that multilevel approximate factorizations will bring further significant improvement with respect to the methods presented here. However the methods presented here are fully operational while the multilevel methods still need further developments to be applicable to irregular three-dimensional grids.

In Section 2, preconditioning of linear systems is introduced by focusing on a first trend in approximate factorization preconditioning, based on the introduction of small perturbations on the diagonal of the factored matrix in order to get a better conditioning. The applicability of methods of the first trend, namely XIC, has already been discussed in a previous work [1] where they have been shown to be very effective for T-problems. The poor performance of XIC on R-problems is highlighted. The CIC preconditioners are described in Section 3, where they are considered as a second trend in IC-based preconditionings. Numerical results show that CIC preconditioners behave much better than XIC for R-problems. Advantages and drawbacks of both methods are discussed in Section 4 which collects numerical results comparing their efficiency on academic and industrial problems. Tests with various more commonly accepted preconditioners like Jacobi (diagonal scaling) or element-by-element methods (EBE) [4] are presented. A comparison is also made with the frontal solver of the industrial FE software SAMCEF, and it is shown that switching between XIC and CIC preconditioners permits lower CPU times and memory requirements than with this direct solution method.

2. A first trend: approximate factorizations aiming at efficiency

2.1. The method(s)

Iterative solution methods are highly affected by the conditioning of the matrix of the system $\mathbf{Kq}=\mathbf{f}$ to be solved. A rough theoretical measure of conditioning is given by the condition number $\boldsymbol{\kappa}$, i.e. the ratio of the largest to the lowest eigenvalue of the considered matrix. In elastostatics FE analyses, stiffness matrices $\mathbf{K}$ are symmetric positive definite and the conjugate gradient method is generally chosen because it yields the most rapid convergence. In this case, the condition number directly affects the convergence through the number of iterations i_ϵ which is bounded by

$$i_\epsilon \leq \frac{1}{2}\sqrt{\kappa}\ \log\frac{2}{\epsilon}+1 \tag{1}$$

if ϵ is the precision required on the solution according to

$$\frac{\|\mathbf{q}_i-\mathbf{q}\|}{\|\mathbf{q}\|}\leq\epsilon$$

$\mathbf{q}_i$ is the approximation of $\mathbf{q}$ obtained at iteration i and the norm is the Euclidean norm.

Remark. All the numerical results presented in this paper have been produced with $\epsilon=10^{-8}$, except for the industrial FE analyses for which $\epsilon=10^{-6}$ was used.

Preconditioning may be viewed as transforming the given system $\mathbf{Kq}=\mathbf{f}$ into

$$[\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2}]\{\mathbf{B}^{1/2}\mathbf{q}\}=\{\mathbf{B}^{-1/2}\mathbf{f}\}$$

where $\mathbf{B}$, the preconditioner, is a symmetric positive definite matrix that needs be inverted at each iteration (in the sense that a system of the form $\mathbf{Bs}=\mathbf{r}$ is to be solved at each iteration). The hope is that the new system matrix $\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2}$ is better conditioned than $\mathbf{K}$. Note that the conditioning of $\mathbf{K}$ depends on a series of parameters such as the mesh size h, the aspect ratio r of the elements, the possible presence of discontinuities

Table 1
The IC factorization

```
$\mathbf{U} \leftarrow \mathrm{up}(\mathbf{S})$
$\mathbf{P} \leftarrow \mathrm{diag}(\mathbf{S})$
For $r = 1 \ldots n$
  For $i=r+1 \ldots n$ such that $(\mathbf{S})_{ri} \neq 0$
  $\chi \leftarrow (\mathbf{U})_{ri}(\mathbf{P})_{rr}^{-1}$
  $(\mathbf{P})_{ii} \leftarrow (\mathbf{P})_{ii} - \chi(\mathbf{U})_{ri}$
  For $j=i+1 \ldots n$ such that $(\mathbf{S})_{rj} \neq 0$
    If $(\psi)_{ij}=1$
      then $(\mathbf{U})_{ij}=(\mathbf{U})_{ij} - \chi(\mathbf{U})_{rj}$
```

Table 2
The DRIC factorization

```
$\mathbf{U} \leftarrow \mathrm{up}(\mathbf{S})$
$\mathbf{P} \leftarrow \mathrm{diag}(\mathbf{S})$
For $r = 1 \ldots n$
  $\tau_0 \leftarrow -(\mathbf{P}_{rr}^{-1} \sum_{r<i} (\mathbf{U})_{ri}$
  If $\tau_0 > \tau$ then $\omega \leftarrow 2\tau/\tau_0 - 1$ else $\omega \leftarrow 1$
  For $i=r+1 \ldots n$ such that $(\mathbf{S})_{ri} \neq 0$
    $\chi \leftarrow (\mathbf{U})_{ri}(\mathbf{P})_{rr}^{-1}$
    $(\mathbf{P})_{ii} \leftarrow (\mathbf{P})_{ii} - \chi(\mathbf{U})_{ri}$
    For $j=i+1 \ldots n$ such that $(\mathbf{S})_{rj} \neq 0$
      If $(\psi)_{ij}=1$
        then $(\mathbf{U})_{ij}=(\mathbf{U})_{ij} - \omega\chi(\mathbf{U})_{rj}$
        else $)\mathbf{P})_{ii} \leftarrow (\mathbf{P})_{ii} - \omega\chi(\mathbf{U})_{rj}$
          $(\mathbf{P})_{jj} \leftarrow (\mathbf{P})_{jj} - \omega\chi(\mathbf{U})_{rj}$
```

or anisotropies, and so on. For each matrix $\mathbf{K}(h, r \ldots)$ a preconditioner $\mathbf{B}(h, r \ldots)$ can be computed; both families of matrices are said to be spectrally equivalent with respect to one or all of the above mentioned parameters if $\kappa(\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2})$ (or at least an upper bound of it) is independent of these parameters.

In addition to conditioning aspects, the preconditioner must satisfy feasibility requirements: since it will be used at each iteration, it must have ‘reasonable’ memory needs and yield a ‘reasonably’ low amount of computations. With these conditions, spectrally equivalent preconditioners are hard to find, causing $\kappa(\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2})$ to be often affected by many parameters. The Jacobi preconditioning ($\mathbf{B}=\mathrm{diag}(\mathbf{K})$) minimizes the memory needs but $\kappa(\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2})$ is then proportional to h^{-2} when $h \to 0$, which we denote $O(h^{-2})$. The same result is also true for element-by-element methods [4] and for basic approximate factorization preconditioners like incomplete Cholesky (IC), even if the leading proportionality constant is smaller. A whole family of approximate factorizations with the generic name XIC were derived from IC in order to reduce the dependency on h, as recalled in [1]. These methods force a ‘resemblance’ between the preconditioner $\mathbf{B}$ and the matrix that undergoes the factorization $\mathbf{K}$ by prescribing that the diagonal entries of $\mathbf{B}$ are computed such that

$$\mathbf{B}\mathbf{x} = \mathbf{K}\mathbf{x} + \Lambda\,\mathrm{diag}(\mathbf{K})\mathbf{x} \tag{2}$$

for a given positive vector $\mathbf{x}$ and a diagonal matrix Λ containing small perturbations. The way Λ is computed corresponds to a particular choice between the methods of the XIC family, which form what is called here the first trend. Their introduction goes back to Buleev [10], but the understanding of their power to improve the conditioning of Stieltjes matrices came from the analysis by Axelsson and his coworkers [11,12]. Subsequent improvement were brought to light a.o. by Beauwens [13,14], who suggested viewing the rowsum rule as a perturbed version of the rule $\mathbf{B}\mathbf{x}=\mathbf{K}\mathbf{x}$, as written under Eq. (2). Dynamic versions were introduced a.o. by Beauwens [15] and by Notay [5]. Among these, DRIC [15] seems to be currently the most robust method since its related conditioning is $O(h^{-1})$ while not being affected by the presence of anisotropies. The DRIC algorithm for finding an approximate factorization

$$\mathbf{B} = (\mathbf{P}+\mathbf{U})^{\mathrm{t}}\mathbf{P}^{-1}(\mathbf{P}+\mathbf{U})$$

of a given matrix $\mathbf{S}$, $\mathbf{P}$ being a diagonal matrix and $\mathbf{U}$ a strictly upper triangular matrix, is given in Table 2 and can be compared to the basic IC algorithm of Table 1. The parameter τ is computed a priori from $1-\tau=h_0$ where h_0 is a dimensionless measure of the mesh size,

$$h_0 = \frac{2}{d\sqrt{\text{number of nodes}}}$$

if the original problem is formulated in a d-dimensional space.

Positive definiteness of $\mathbf{S}$ does not ensure that IC and XIC are feasible algorithms, i.e. that all entries of the diagonal matrix $\mathbf{P}$ are positive. A sufficient condition is that $\mathbf{S}$ be a *Stielyes* matrix—symmetric positive definite with non-positive off-diagonal entries, which is almost never satisfied for stiffness matrices $\mathbf{K}$ in practical FE structural analysis. A solution to this bottleneck consists of building a Stieltjes matrix $\mathbf{S}$ from $\mathbf{K}$ such that these are spectrally equivalent; $\mathbf{S}$ then undergoes the XIC factorization. A method for finding a spectrally equivalent matrix $\mathbf{S}$, called DC-reduction, is proposed and validated in [1] for T-problems and very satisfactory numerical results have been obtained.

The DC-reduction proceeds in two steps: the so-called D- and C-reduction, respectively, due to Axelsson and Gustafsson [16] and Munksgaard [17]. First, a decoupled matrix $\mathbf{K}_{\mathrm{D}}$ is built from $\mathbf{K}$ by decoupling all translational degrees of freedom related to different axes:

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{xx} & \mathbf{K}_{xy} & \mathbf{K}_{xz} \\ \mathbf{K}_{yx} & \mathbf{K}_{yy} & \mathbf{K}_{yz} \\ \mathbf{K}_{zx} & \mathbf{K}_{zy} & \mathbf{K}_{zz} \end{bmatrix} \Longrightarrow \mathbf{K}_\mathrm{D} = \begin{bmatrix} \mathbf{K}_{xx} & 0 & 0 \\ 0 & \mathbf{K}_{yy} & 0 \\ 0 & 0 & \mathbf{K}_{zz} \end{bmatrix}$$

Next, the remaining positive entries in $\mathbf{K}_\mathrm{D}$ are shifted onto the diagonal to get a Stieltjes matrix $\mathbf{S}$, such that

$$\mathrm{offdiag}(\mathbf{S}) = \ \min(\mathrm{offdiag}(\mathbf{K}_D,\mathbf{0}))$$

and

$$\mathbf{S1} = \mathbf{K}_\mathrm{D}\mathbf{1}$$

if $\mathbf{1}$ is a vector whose components are all 1.

2.2. On the spectral equivalence of DC-reduction with respect to the mesh size for R-problems

The theoretical works that validate the DC-reduction, and more precisely the D-reduction step, are based on the Navier equations of elasticity which allow the modelling of any structure. However, FE users generally prefer to use other models for structures having a much smaller size following one direction. The reason is that using the Navier equations implies an explicit FE discretization along the thickness of the structure and involves:

- either the use of pretty flat—and therefore ill-conditioned—solid elements;
- or a large amount of well-conditioned solid elements (almost cubic), involving huge computational requirements.

Typical modellings for the above-mentioned structures, like Kirchhoff's or Mindlin–Reissner's, allow them to be considered as if they were two-dimensional without any explicit discretization along the thickness. The distribution of strains through the thickness is then assumed to satisfy given hypotheses (e.g. plane rotation of cross-sections). This is taken into account thanks to the introduction of rotational degrees of freedom representing the rotation of the cross-sections and therefore the strain evolution along the thickness. The equations obtained from the analysis of plates are more sophisticated than Navier's and are of fourth order instead of second order [6].

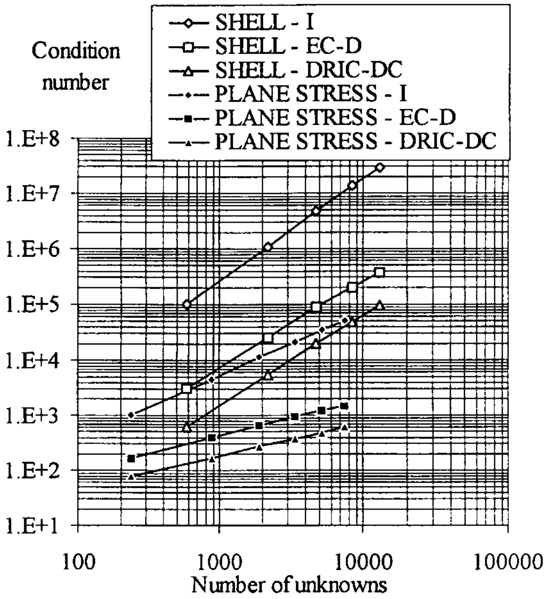

Fig. 1. Condition numbers obtained with I, EC-D and DRIC-DC preconditioners for regular plane stress and shell problems (t = 0.005 m).

Shlafman and Efrat [7] claimed that the scope of D-reduction could be extended to R-problems as well. Therefore, our preliminary developments [1], together with some unpublished numerical experiments on thick plates, gave us the hope that the DC-reduction would apply to R-problems as successfully as they did to T-problems. Further experiments with thin shells given below, discouraged us from considering further this approach. The first experiments investigate the quality of the D-reduction and the DRIC-DC preconditioner, through the condition numbers of

- $\mathbf{K}$, which is considered as the system matrix with preconditioner $\mathbf{I}$ (in this paper, a condition number is always related to a preconditioner);
- $\mathbf{K}_\mathrm{D}^{-1/2}\ \mathbf{K}\mathbf{K}_\mathrm{D}^{-1/2}$ which corresponds to the EC-D preconditioner (exact Cholesky factorization preceded by a D-reduction);
- $\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2}$ where $\mathbf{B}$ is the DRIC-DC preconditioner of order 1;

which are represented in Fig. 1 for a series of test problems including plane stress and shell analyses. The tested structures are regular meshes of four-node quadrilateral elements; the values of the unknowns at the corners are fixed to zero and there is a nodal load of 10^4 N at the centre of the structure, applied in all translational directions (x and y for the plane stress, x, y and z for the shell). The dimensions of the structures are always (1, 1, 0.005) where all lengths are expressed in meters. The Young modulus and Poisson ratio are 1.1 10^{11} N/m^2 and 0.3.

The comparison of $\kappa(\mathbf{K})$ and $\kappa(\mathbf{K}_\mathrm{D}^{-1/2}\mathbf{K}\mathbf{K}_\mathrm{D}^{-1/2})$ reflects the quality of the D-reduction, and thus partly of the DC-reduction. In both cases (plane stress and shells) the condition number is reduced and the difference between the two curves does not decrease for growing numbers of unknowns, tending to validate empirically the spectral equivalence of $\mathbf{K}$ and $\mathbf{K}_\mathrm{D}$ with respect to

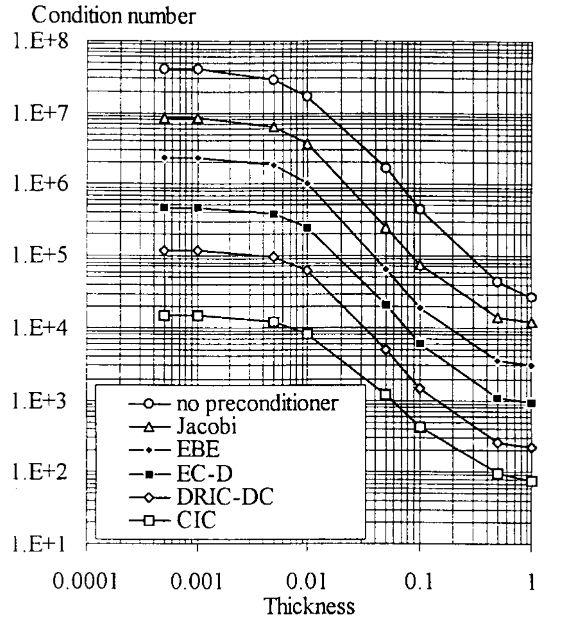

Fig. 2. Effect of the thickness t on the conditioning of the preconditioned system for various preconditioners.

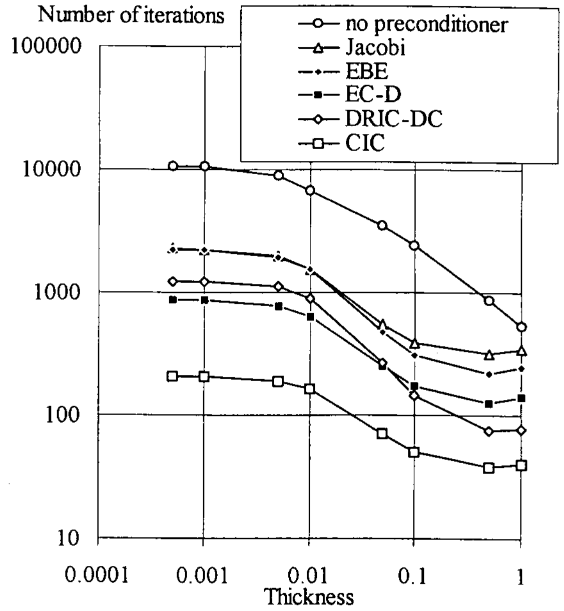

Fig. 3. Effect of the thickness t on the number of iterations obtained with various preconditioners.

the mesh size h. Unfortunately, the condition numbers of $\mathbf{K}_D^{-1/2}\mathbf{K}\mathbf{K}_D^{-1/2}$ for shells are much larger than those obtained for plane stress analyses and above all, they increase faster. In fact, the reduction seems to perform actually equally well as for plane stress but the initial matrix to reduce, **K**, is much worse conditioned. These comments on the EC-D preconditioner extend to DRIC-DC. The reason for the lack of efficiency of the DC-reduction for the R-problems has not yet been fully understood and deserves further investigation.

2.3. On the spectral equivalence of DC-reduction with respect to the thickness

It is easy to find numerically the origin of the ill-conditioning of **K** for shell problems by performing some experiments in which all parameters are fixed except the thickness t. This is shown in Fig. 2 which represents the condition number against the thickness for a square shell (already tested in the previous section) with 50 by 50 quadrilateral four-node Mindlin elements. In addition to the condition numbers obtained with I, EC-D and DRIC-DC preconditioners, those of

- $\mathbf{D}^{-1/2}\mathbf{K}\mathbf{D}^{-1/2}$ with $\mathbf{D}=\text{diag}(\mathbf{K})$, the Jacobi preconditioner;
- $\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2}$ where **B** is the CIC(10^{-3}) presented in section 3;
- $\mathbf{B}^{-1/2}\mathbf{K}\mathbf{B}^{-1/2}$ where **B** is the standard EBE preconditioner introduced by Hughes et al. [4];

are also presented in Fig. 2.

The important thing is that the condition number varies by a factor up to 1000 for DRIC-DC. Eq. (1) shows that an increase of the condition number by a factor 1000 may yield an increase of the number of iterations by a factor 30, which is roughly verified numerically from Fig. 3. The spectral equivalence is then not useful.

If the condition numbers for thin shells were 100 times smaller, the curves related to thin shells in Fig. 1 would be shifted down near the plane stress curves and the same performance as for plane stress analyses could be expected. This explains the above mentioned optimistic results obtained by the authors with thick shells that have condition numbers 1000 times smaller than those of thin shells.

3. A second trend : approximate factorizations aiming at robustness

3.1. The CIC method

As a conclusion of the previous section, the DC-reduction allowing the use of high-performance preconditioners like DRIC cannot be used to solve problems derived from discretizations involving rotational degrees of freedom. Moreover, the condition numbers reach considerably larger values when compared to those usually encountered in the plane stress case.

Table 3
The alternate IC factorization

```
For $i = 1 \ldots n$
  $(\mathbf{P})_{ii} \leftarrow (\mathbf{K})_{ii} - \sum_{r<i} (\mathbf{U})_{ri}^2 (\mathbf{P})_{rr}^{-1}$
  For $j = i + 1 \ldots n$
    $(\text{candidate})_{ij} \leftarrow (\mathbf{K})_{ij} - \sum_{r<i} (\mathbf{U})_{ri} (\mathbf{U})_{rj} (\mathbf{P})_{rr}^{-1}$
    If $(\psi)_{ij} = 1$
      then $(\mathbf{U})_{ij} \leftarrow (\text{candidate})_{ij}$
      else $(\mathbf{U})_{ij} \leftarrow 0$
```

We unfortunately did not find an alternate satisfactory reduction technique. This led us to consider methods that automatically produce positive definite preconditioners (when applied to any positive definite matrix) and more specifically

- explicit methods like Jacobi or EBE. Because Jacobi is the best diagonal preconditioner [9], any improvement over Jacobi must involve some offdiagonal structure. On the other hand no optimality result is known for sparsity patterns like those obtained through EBE. Direct numerical experimentation is thus required. But the numerical results already presented in Figs. 2 and 3 show that neither Jacobi nor EBE fits better than DRIC-DC to the solution of sytems derived from shell analyses;
- approximate factorization schemes that are *corrected* in order to always lead to a positive definite preconditioner. The remaining of this section is devoted to these CIC methods.

In order to avoid possible zero or negative entries in **P**, Jennings and Malik [2] introduced in 1977 a modified version of IC which has the nice property of always producing a positive definite factorization when applied to any symmetric positive definite matrix. The Jennings–Malik modification consists in adding to the diagonal entries of rows i and j corrections equal to the absolute value of rejected fill-in in position (i,j), properly rescaled. The method was subsequently improved by Ajiz and Jennings [3] through determining the fill-in pattern by value (according to some drop-tolerance criterion) rather than by position. Since Jennings and co-authors did not give a name to their method, we called it 'corrected' IC, or CIC, to avoid confusion with 'modified', 'relaxed' and other 'perturbed' IC methods (see [1]).

CIC is based on a splitting of the system matrix **K** of the form

$$\mathbf{K} = \mathbf{B} - \mathbf{E} \tag{3}$$

with **B** being defined by

$$\mathbf{B} = (\mathbf{P} + \mathbf{U})^{t} \mathbf{P}^{-1} (\mathbf{P} + \mathbf{U})$$

Table 4
The CIC factorization

```
For $i = 1 \ldots n$
  $(\mathbf{P})_{ii} \leftarrow (\mathbf{K})_{ii} - \sum_{r<i} (\mathbf{U})_{ri}^2 (\mathbf{P})_{rr}^{-1}$
  For $j = i + 1 \ldots n$
    $(\xi)_{ij} \leftarrow (\mathbf{K})_{ij} - \sum_{r<i} (\mathbf{U})_{ri} (\mathbf{U})_{rj} (\mathbf{P})_{rr}^{-1}$
    If $(\psi)_{ij} = 1$
      then $(\mathbf{U})_{ij} \leftarrow (\xi)_{ij}$
      else compute $r$
      $(\mathbf{P})_{\mathrm{ii}} \leftarrow (\mathbf{P})_{ii} + r^{2} |(\xi)_{ij}|$
      $(\mathbf{P})_{jj} \leftarrow (\mathbf{P})_{jj} + r^{-2} |(\xi)_{ij}|$
```

The positive definiteness of **B**, and thus its ability to be used as preconditioner in a conjugate gradient iteration, would be ensured if **E** was at least non-negative definite, since **K** is already supposed to be positive definite. Starting from the IC factorization scheme, reformulated in Table 3, modifications are introduced in order to satisfy this condition. In the IC scheme of Table 3, there are so-called 'candidates' eligible to become entries of the factored matrix. If all candidates were accepted, the factorization would be exact. Here, the factorization is incomplete due to the possible zero values found in the fill-in matrix ψ. If candidate (i, j) is rejected, i.e. if $(\psi)_{ij} = 0$, IC sets entry (i, j) of **U** to a zero value. This is the reason why **B** may be non-positive definite, which must be avoided by acting on **E**.

Indeed, the rejected candidates contribute to the error matrix **E**. Let $(\text{candidate})_{ij} = \xi$ be rejected, which causes $(\mathbf{E})_{ij} = -\xi$. We would like to find a way to correct matrix **E** such that this latter be non-negative definite. Assuming that no rejection is performed on the diagonal entries, only the off-diagonal part of **E** is fed by rejected entries and we remain free to alter the diagonal of **E**. The diagonal entries corresponding to the rejected entry at position (i, j) may always be written as the product of $|\xi|$ with, temporarily unknown, numbers a and b, which is suggested by writing **E** as

$$\mathbf{E} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots \\ \cdots & a\,|\,\xi\,| & \vdots & -\xi & \vdots \\ \cdots & \cdots & \ddots & & \vdots \\ \cdots & -\xi & \cdots & b\,|\,\xi\,| & \vdots \\ \cdots & \cdots & \cdots & \cdots & \ddots \end{bmatrix}$$

Matrix **E** is non-negative definite if $\langle \mathbf{v}, \mathbf{E}\mathbf{v} \rangle \geqslant 0$ for any **v**. The quadratic form expands as

$$|\,\xi\,|\,(a(\mathbf{v})_i^2 \pm 2(\mathbf{v})_i(\mathbf{v})_j + b(\mathbf{v})_j^2)$$

and taking $a = r^2$ and $b = r^{-2}$ gives

$$| \xi | (r(\mathbf{v})_i \pm r^{-1}(\mathbf{v})_j)^2$$

which is non-negative for any r and ensures the non-negative definiteness of **E**.

It has to be noticed that, due to Eq. (3), one cannot add $r^2|\xi|$ and $r^{-2}|\xi|$ on the main diagonal of **E** without adding them on the main diagonal of **B**, which gives the final form of the CIC algorithm of Table 4.

3.2. Choosing the rejection parameter

There exists some similarity between XIC and CIC factorizations. Both methods accept or reject off-diagonal entries according to some sparsity pattern. For the diagonal entries, XIC computes them following Eq. (2) for a given perturbation matrix while CIC uses Eq. (3) with a given definition of the rejection parameter. The rejection parameter r may besides be considered as a perturbation; the difference with XIC is that this perturbation

- is not small;
- is not designed for the enhancement of the convergence but to ensure the feasibility of the algorithm.

This difference is at the origin of the larger dependency of CIC with respect to parameters like h. Although it has not yet been possible to evaluate theoretically this dependency, numerical experiments tend to show that the obtained convergence is $O(h^{-2})$ with much smaller leading constant than for Jacobi, EBE or IC.

Of course, the value of r remains free. With $r = 1$ (the absolute values of) entries (i, j) that are rejected and do not contribute to the off-diagonal part of **B** are simply added on the corresponding (i, i) and (j, j) entries of the main diagonal.

Another possibility would be to set

$$r^2 = \sqrt{\frac{(\mathbf{P})_{ii}}{(\mathbf{P})_{jj}}}$$

as proposed in [3]. With this choice, the modification of the diagonal entries in CIC, as described in Table 4, becomes

$$(\mathbf{P})_{ii} \leftarrow (\mathbf{P})_{ii}\left(1 + \frac{| \xi |}{\sqrt{(\mathbf{P})_{ii}(\mathbf{P})_{jj}}}\right);$$

$$(\mathbf{P})_{jj} \leftarrow (\mathbf{P})_{jj}\left(1 + \frac{| \xi |}{\sqrt{(\mathbf{P})_{ii}(\mathbf{P})_{jj}}}\right)$$

meaning that the diagonal entries are increased proportionally to their initial value. There is currently no rigorous way to select r but our experience is that the latter generally gives CIC the fastest and smoothest convergence with respect to element size h and thickness t, even if the advantage is slight.

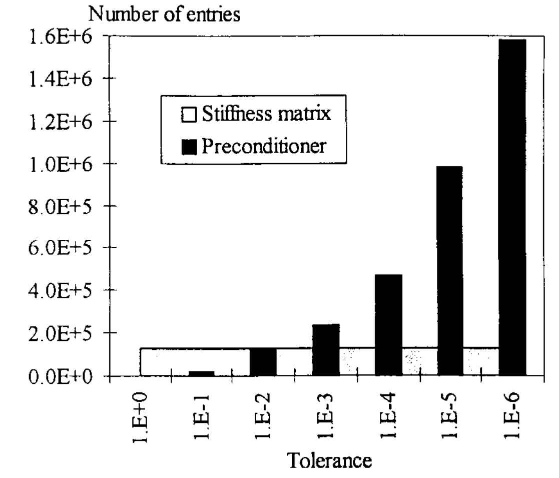

Fig. 4. Number of entries in the stiffness matrix and in the CIC(ϕ) preconditioner for varying values of the drop-tolerance ϕ, on the benchmark PARKING.

3.3. Choosing the fill-in pattern

The choice of the fill-in pattern ψ is still to be discussed. A first possibility is to use the concept of order, with increasing fill-in patterns ψ_0, ψ_1, ... ψ_m of order 0, 1 ... m corresponding to the definitions

- $\psi_0 \equiv \{(\psi_0)_{ij} = 1 \Leftrightarrow i = j\}$
- $\psi_1 \equiv \{(\psi_1)_{ij} = 1 \Leftrightarrow (\mathbf{K})_{ij} \neq 0\}$
- $\psi_m \equiv \{(\psi_m)_{ij} = 1 \Leftrightarrow (\psi_{m-1})_{ij} = 1$ or $\exists r < i, j$ such that $(\psi_{m-1})_{ri} = 1$ and $(\psi_{m-1})_{rj} = 1\}$

Only the connectivity of the entries of the matrix to be factored influences such fill-in patterns. For low-order fill-ins, the factorization reflects local effects only, increasing the order creates a connectivity between unknowns that are more distant with respect to the connectivity of the matrix to be factored, so that the incomplete factorization tends to exact elimination. This kind of fill-in pattern was found very effective even at very low (0 or 1) orders for structures including only translational degrees of freedom [1] when XIC-DC preconditioners were used.

Increasing the order is not a solution to enhance significantly the efficiency of XIC-DC on R-problems since it has been established that the thickness has an important effect while not modifying the connectivity of the system matrix. A more accurate fill-in pattern would thus not (only) be based on connectivity but rather on the value of the entries of the factored matrix. Such a fill-in pattern is built by setting $(\psi)_{ij}$ to 1 if and only if the candidate corresponding to entry (i, j) is larger than some value, following for instance

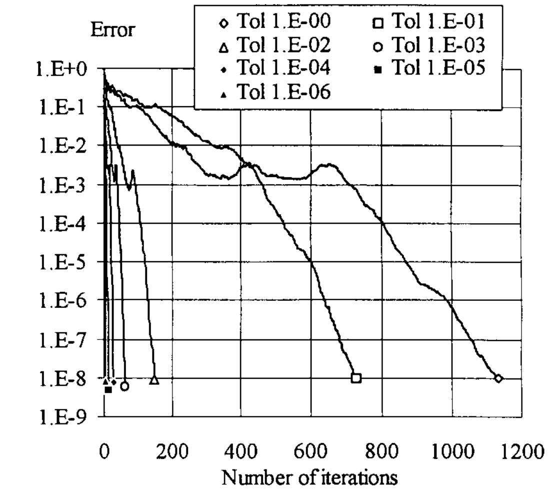

Fig. 5. Effect of the drop-tolerance ϕ on the convergence for the CIC(ϕ) preconditioner, illustrated on the benchmark PARKING.

$$(\text{candidate})_{ij}^2 > \phi(\mathbf{P})_{ii}^2$$

or

$$(\text{candidate})_{ij}^2 > \phi(\mathbf{P})_{ii}(\mathbf{P})_{jj}$$

with ϕ being a given drop-tolerance. It is our experience that the latter test gives better results; see [3,8] for additional comments about this topic.

Here also, there is no optimal value for ϕ and one has to proceed empirically. Figs. 4 and 5 represent the effect of ϕ on memory needs and convergence respectively, for a benchmark named PARKING already used in [1] which represents the plane stress study of a concrete parking floor. In Fig. 4 the growing memory needs of the preconditioner **B** are compared to those of the stiffness matrix **K** (not affected by ϕ). Since we are interested in solving large systems and **B** is stored in RAM with **K**, it would be reasonable to keep the size (i.e. the number of non-zero entries) of **B** under or close to that of **K**. On the other hand, the plot of Fig. 5 shows that a decrease of ϕ from 1 to 10^{-3} produces impressive accelerations of the convergence while values beyond 10^{-3} yield only small enhancements at a much larger cost in memory.

The results obtained with this benchmark typify the general behaviour of CIC, so that all the results in this paper are produced with 10^{-3}, which is denoted CIC(10^{-3}).

4. More numerical results

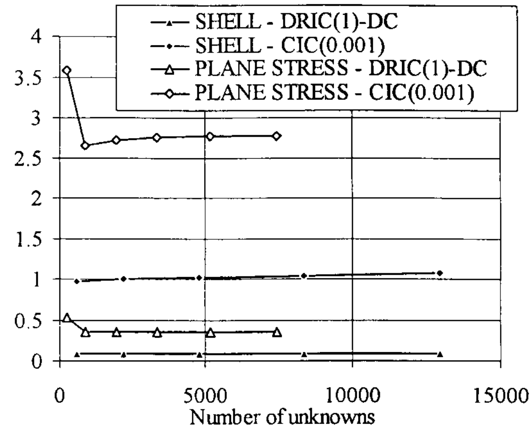

Fig. 6. Ratio of the number of entries in the preconditioner and the system matrix for DRIC(1)-DC and CIC(10^{-3}) on regular thin shell ($t = 0.005$) and regular plane stress structures.

The tests performed in the previous sections were designed to highlight the effect of some parameters on the behaviour of the studied preconditioners. From here, the numerical experiments will serve for the discussion of the advantages and drawbacks of DRIC-DC and CIC. The efficiency of both techniques will be compared to that of the high-performance frontal solver of the industrial FE software SAMCEF v5.1. All tests were run on a SUN SPARC20/514-50 with 128 Mb RAM.

The first major difference between DRIC-DC and

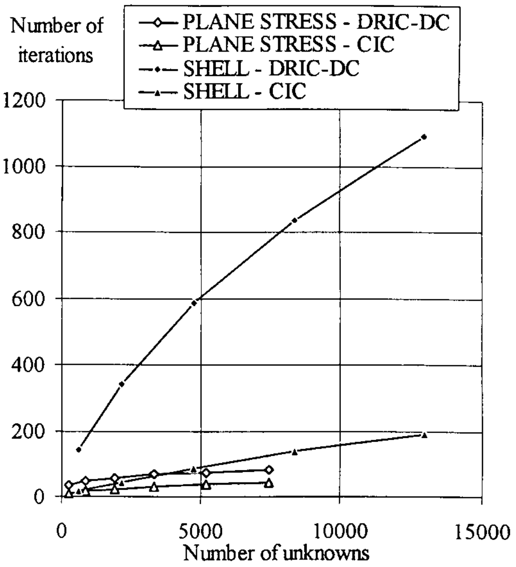

Fig. 7. Number of iterations for DRIC(1)-DC and CIC(10^{-3}) on regular thin shell (t = 0. 005) and regular plane stress structures.

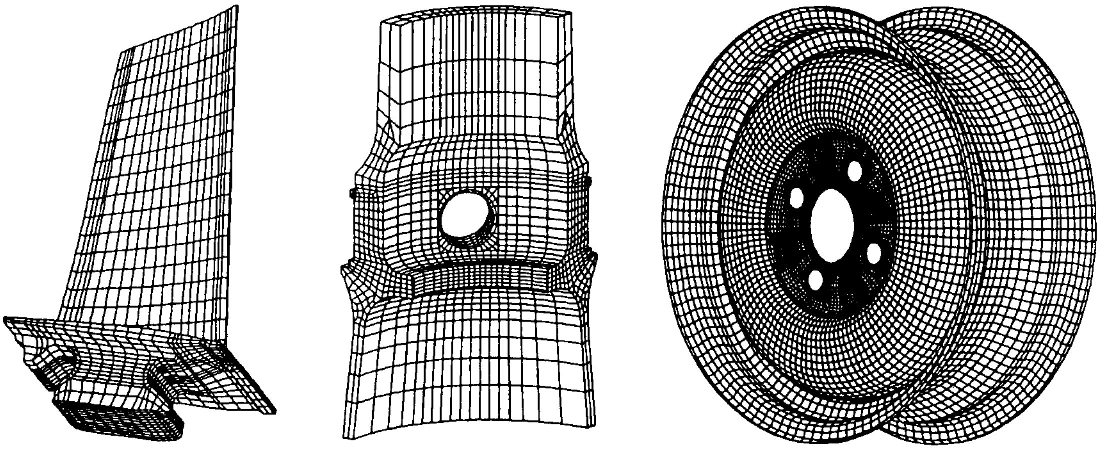

Fig. 8. The BLADE, SHAFT and RIM industrial grids.

CIC is their memory needs. The preconditioner **B** must be stored in the RAM because it is used at each iteration and it is well known that swapping can increase dramatically the solution time. The system matrix **K** is also used at each iteration and must also be stored in the RAM. These two matrices are the main memory requirements; the place taken by **K** is not available to **B**, so the ratio of the number of nonzero entries of **B** and that of **K** gives a good idea of the memory requirements of a preconditioner. This ratio is represented in Fig. 6 for a series of thin shell and plane stress structures.

DRIC(1)-DC always gives values of the ratio under 1 (and even under 0.5); it is thus a quite economical preconditioner. CIC(10^{-3}) leads to a ratio around 1 for shell problems, which is acceptable, but the ratio increases to 2.5 for plane stress structures, which is no longer reasonable with a view to the solution of large problems.

Fig. 7 represents the numbers of iterations obtained for regular shell and plane stress problems against the number of unknowns. The number of iterations of CIC is always smaller than that of DRIC-DC. For plane stress however the enhancement obtained by switching from DRIC-DC to CIC is obtained at a much too large cost with respect to memory requirements.

Regular plane grids allow curves to be plotted and the effect of h and t on the quality of the preconditioners to be studied. They do not, however, reflect the complexity of the grids encountered in industrial practice. Fig. 8 represents industrial models of, respectively, a turbine blade, a quarter of a turbine shaft and the rim of a car. The blade and the shaft are meshed with *solid* elements of various shapes, with linear and quadratic shape functions while the rim is meshed with triangular and quadrilateral Marguerre shell elements. Additional information on the number of unknowns, elements and frontwidth is given in Table 5.

Table 6 presents a series of numerical values related to the solution of these industrial benchmarks by DRIC-DC and CIC, and by the frontal solver FRONT of the FE software SAMCEF v5.1 which is generally admitted to be efficient. This gives an idea of the quality of our preconditioners compared to currently implemented direct solvers on which industrial FE softwares generally rely because of their robustness. The first thing to notice in Table 6 is that DRIC-DC is not able to deal with the rim (the error has not been reduced sufficiently during the first 7500 iterations). In contrast, CIC obtained the solution in 409 iterations, but suffers from its large RAM requirements when considering the 'solid‘ benchmarks: CIC(10^{-3}) requires 2.36 times more RAM for the blade problem and was not able to solve the shaft problem within 120 Mb. When switching to CIC(10^{-2}), the CPU time is

Table 5
Characteristics of the proposed benchmarks

Benchmark	Number of unknowns	Number of elements	Frontwidth
BLADE	38657	5084	2619
SHAFT	49119	5622	1499
RIM	59490	9928	1023

Table 6
Comparison of DRIC(1)-DC, CIC(10^{-3}) and FRONT

	Front			DRIC(1)-DC				CIC			
Benchmark	CPU (s)	RAM (Mb)	Disk (Mb)	CPU (s)	RAM (Mb)	Disk (Mb)	Iterations	CPU (s)	RAM (Mb)	Disk (Mb)	Iterations
BLADE	15396	67.3	538.9	2897	37.7	37.7	1663	3651	89.2	37.7	881
SHAFT	5614	29.3	495.9	1021	54.2	56.9	236	3081[b]	89.2[b]	56.9[b]	962[b]
RIM	3102	13.0	475.6	[a]	23.0	26.4	[a]	1365	66.2	26.4	409

[a] = No convergence was obtained after 7500 iterations. [b] = More than 120 Mb RAM were required by CIC(10^{-3}), results produced with CIC(10^{-2}).

not satisfactory when compared to that obtained with DRIC-DC. Secondly, when the most effective preconditioner is chosen according to the nature of the problem to be solved (T- or R-), the solution is always obtained within:

- a smaller CPU time;
- a much smaller amount of disk space;
- and a larger (while reasonable) amount of RAM;

than with the frontal solver. Concerning the disk space, only the elementary stiffness matrices are needed for the iterative solver to build the assembled stiffness matrix. This explains that the same disk space is required for both preconditioners.

5. Conclusions

In this paper, we concentrated on the solution of linear systems arising from FE discretizations with rotational degrees of freedom. Some preconditioners have been discussed in the context of a conjugate gradient iteration. It has been shown that due to the strong ill-conditioning of R-problems stiffness matrices, Jacobi and element-by-element preconditioners yield too slow convergence rates, so that one has to go to approximate factorizations.

A DC-reduction had been successfully introduced to allow the use of high-performance perturbed modified approximate factorizations (XIC) in the context of discretizations of the Navier equations of elasticity. It has been shown that XIC-DC could perform as well for R-problems provided the thickness has about the same value as the 'in-plane' dimensions of the elements, a condition seldom fulfilled. It has been demonstrated that CIC gives satisfying results (fast convergence and small CPU times) for thin shells while its larger RAM requirements make it too expensive for T-problems and it has therefore been proposed to switch between DRIC-DC and CIC according to the nature of the problem. This allowed an efficient frontal solver to always be outperformed.

Finally, the numerical experiments proposed in this paper give sufficient information to use the solver as a 'black box' by selecting appropriate values for the parameters on which the preconditioners depend. Indeed, the main parameter that affects the preconditioner concerns the fill-in pattern. Fill-ins based on the concept of order and connectivity have been validated in [1] for T-problems and order 0 and 1 were determined to enable fast convergence. The interest of a fill-in pattern based on the rejection of small entries has been discussed here for shell analyses and a drop-tolerance criterion of 10^{-3} has been found to give fast convergence while limiting the memory needs of CIC.

Acknowledgements

This work is funded by the F.R.I.A. (Fonds pour la formation à la Recherche dans l'Industrie et l'Agriculture) and the F.N.R.S. (Fonds National de la Recherche Scientifique). The authors are most appreciative for the support of SAMTECH S.A. (Liège, Belgium) and TECHSPACE AERO S.A. (Herstal, Belgium) and wish to thank them for having provided the industrial meshes presented in this paper.

References

[1] Saint-Georges P, Warzée G, Notay Y, Beauwens R. High-performance PCG solver for FEM structural analyses. Int J Numer Meth Engng 1996;39:1133–60.

[2] Jennings A, Malik G. Partial elimination. J Inst Math Appl 1977;20:307–16.

[3] Ajiz MA, Jennings A. A robust incomplete Cholesky conjugate gradient algorithm. Int J Numer Meth Engng 1984;20:949–66.

[4] Hughes TJR, Levit I, Winget J. An element by element solution algorithm for problems of structural and solid mechanics. Comput Meth Appl Mech Engng 1983;36:241–54.

[5] Notay Y. DRIC: a dynamic version of the RIC method. J Numer Linear Algebra Appl 1994;1:511–32.

[6] Zienkiewicz OC, Taylor RL. The finite element method. vol. 2, 4th ed. McGraw-Hill, London, 1991.

[7] Shlafman S, Efrat I. Using Korn's inequality for an efficient iterative solution of structural analysis problems. In: Beauwens R, de Groen P, editors. Iterative Methods in Linear Algebra. Amsterdam: North-Holland, 1992. p. 575–81.

[8] Dickinson JK, Forsyth PA. Preconditioned conjugate gradient methods for three-dimensional linear elasticity. Int J Numer Meth Engng 1994;37:2211–34.

[9] Greenbaum A, Rodrigue GH. Optimal preconditioners of a given sparsity pattern. BIT 1989;29:610–34.

[10] Buleev NI. A numerical method for the solution of two-dimensional and three-dimensional equations of diffusion. Math Sb 1960;51:227–238; English translation in report BNL-TR- 551, Brookhaven, National Laboratory, Upton, NY, 1973.

[11] Axelsson O. A generalized SSOR method. BIT 1972;13:443–67.

[12] Gustafsson I. A class of first order factorization methods. BIT 1978;18:142–56.

[13] Beauwens R. Upper eigenvalue bounds for pencils of matrices. Linear Algebra Applic 1984;62:87–104.

[14] Beauwens R. On Axelsson's perturbations. Linear Algebra Applic 1985;68:221–42.

[15] Beauwens R, Modified incomplete factorization strategies. In: Axelsson O, Kolotilina L, editors. Preconditioned Conjugate Gradient Methods. Lectures Notes in Mathematics, vol. 1457. Berlin: Springer, 1990. p. 1–16.

[16] Axelsson O, Gustafsson I. Iterative methods for the solution of the Navier equations of elasticity. Comput Meth Appl Mech Engng 1978;15:241–58.

[17] Munksgaard N. Solving sparse symmetric sets of linear equations by preconditioned conjugate gradients. ACM Trans Math Softw 1980;6:206–19.