Stress based finite element methods for solving contact problems: Comparisons between various solution methods

François Kuss *, Frédéric Lebon

Laboratoire de Mécanique et d’Acoustique, 31 Chemin Joseph Aiguier, F-13402 Marseille Cedex 20, France

A R T I C L E   I N F O

Article history:
Received 21 January 2007
Accepted 28 November 2008
Available online 12 January 2009

Keywords:
Contact
Friction
Equilibrium finite elements
Augmented Lagrangian
Condensation

A B S T R A C T

This paper deals with numerical methods for solving unilateral contact problems with friction. Although these problems are usually defined in terms of the displacement, a stress based approach to the problem is developed here. The “equilibrium” finite elements method is therefore used. Using these elements make it possible to satisfy the local equilibrium condition a priori, but on the other hand, prescribed and contact forces have to be introduced using Lagrangian multipliers. The problem obtained is therefore a non-linear, constrained problem and the global system matrix is non-positive definite. Various solution algorithms are thus proposed and compared. Comparisons between the classical method and that developed here show that the stress formulation gives very satisfactory results in terms of the stresses.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Here we present a non-classical approach for solving contact problems with friction. These problems are usually approached using the classical displacement based finite element method, with which many solution methods have been proposed and proved to be efficient. With approaches of this kind, the compatibility conditions are fulfilled, but not the local equilibrium conditions. A stress formulation for the problem, satisfying the local equilibrium conditions, is proposed, first because the stresses are usually the parameter of interest in engineering contexts, secondly with a view to using this approach for mesh refinement and limit analysis purposes. This paper is an extended and improved version of the paper [1] presented at the Fifth International Conference on Engineering Computational Technology. In the second section, the classical unilateral contact problem and its stress formulation are presented.

The method of discretizing the stress formulation presented in Section 3 involves equilibrium finite elements. The element used is described, it was found to be more suitable for contact and friction problems, since it makes it possible to impose the contact and friction conditions both strictly and simply. Some convergence results are presented in the appendix.

In Section 3, we show how equilibrium finite elements can be used to satisfy the local equilibrium conditions a priori, while imposing the use of degrees of freedom which are not directly stresses, prescribed forces and contact forces therefore have to be introduced using Lagrangian multipliers. The global system, which is presented in Section 4, is a non-linear problem with constraints due to the contact and friction conditions, and its matrix system is non-positive definite due to the Lagrangian multipliers involved. Various solvers are then presented: three based on condensation procedures and one on the Uzawa algorithm. In Section 5, the results obtained using the classical method and that presented here are compared. The validity of the present method is confirmed by these comparisons: it gives better results than the classical one in terms of the stresses. Lastly, the solvers presented here are compared. Various preconditioners are first compared by applying them to the stress discretization procedure. The computation times and memory requirements of each of the solvers applied to various problems are then compared.

2. The contact problem

2.1. Mechanical problem

In this section, the classical contact problem shown in Fig. 1 is presented. We consider a deformable body Ω in receding contact with a rigid foundation. The boundary Γ is split into three parts, ΓD, ΓF and ΓC so that Γ = ΓD ∪ ΓF ∪ ΓC. We take u = (ui) to denote the displacement field, ε = εi(u) = 1/2(∂ui/∂xj + ∂uj/∂xi) to denote the linearised strain tensor, σ = σi(u) = Aijklεjkl(u) to denote the stress tensor and x = A−1 to denote the flexibility tensor. The body is in contact with the foundation at ΓC. The contact is governed by the Signorini unilateral conditions and the quasi Coulomb’s friction law. At the contact boundary, the displacement and the stress vector are
decomposed by introducing \( n \), the outward normal unit vector to \( \Gamma \):

\[
\begin{align*}
 u_n &= u_n n, \quad u_t = u - u_n \cdot n, \\
 \sigma_n &= \sigma_0 n, \quad (\sigma_t)_i = \sigma_0 n_i - \sigma_n n_i.
\end{align*}
\]

The body is subjected to volume forces \( F = (f_i) \) in \( \Omega \) and to surface forces \( f = (f) \) on \( \Gamma_s \) (if possibly be zero at a part of \( \Gamma_s \)). It is assumed that \( F \in [L^2(\Omega)]^d \) and \( f \in [L^2(\Gamma_s)]^d \). The displacement is prescribed on \( \Gamma_D \). The mechanical problem involving an elastic body in contact with a rigid obstacle is classically written as follows:

**Problem (P).** Find \( u \) and \( \sigma \) such that

\[
\begin{align*}
 \sigma &= A e(u) \quad \text{in} \ \Omega, \\
 \text{div} \sigma &= -F \quad \text{in} \ \Omega, \\
 u &= u_0 \quad \text{on} \ \Gamma_D, \\
 \sigma \cdot n &= f \quad \text{on} \ \Gamma_s, \\
 u_n &\leq 0, \quad \sigma_n \leq 0, \quad u_n \sigma_n = 0 \quad \text{on} \ \Gamma_C, \\
 |\sigma_i| &\leq \mu \sigma_n \quad \text{on} \ \Gamma_e \quad \text{and} \quad |\sigma_i| < -\mu \sigma_n \Rightarrow u_i = 0, \\
 |\sigma_i| &< -\mu \sigma_n \Rightarrow \exists \lambda \geq 0, u_i = -\lambda \sigma_i,
\end{align*}
\]

where \( \mu \) is the friction coefficient. Classically, this problem is treated using a kinematically admissible formulation, written in terms of the displacements. Many studies have been carried out on these lines (for further details, see [2]). We adopt a statically admissible formulation of the problem presented above.

![Fig. 1. The classical contact problem.](image)

### 2.2. Dual (stress) formulation

First, we define the statically admissible sets:

\[
\begin{align*}
 H &= \{ \sigma = (\sigma_i) : \sigma_i \in L^2(\Omega) \}, \\
 H_{ef} &= \{ \sigma \in H : \sigma_{ij} + F_i = 0 \quad \text{in} \ \Omega, \ \sigma_{ij} n_j = f_i \quad \text{on} \ \Gamma_f \}, \\
 \Sigma (\tau) &= \{ \sigma \in H_{ef} : \sigma_n \leq 0, |\sigma_i| \leq \mu \tau_i \quad \text{on} \ \Gamma_c \}
\end{align*}
\]

to ensure the symmetry of the stress tensor, the local equilibrium conditions and the contact and friction conditions.

The variational formulation presented in [3] which is written in terms of the stresses and is known as the dual formulation, is used here:

**Problem (P).** Find a stress field \( \sigma : \Omega \rightarrow H \) such that

\[
\begin{align*}
 \{ \sigma \in \Sigma (\tau), \\
 b(\sigma, \tau - \sigma) &\geq l(\tau - \sigma) \quad \forall \tau \in \Sigma (\sigma)
\end{align*}
\]

with \( b(\sigma, \tau) = \int_\Omega A^{-1} \sigma \cdot \tau \, dx \) and \( l(\tau) = \int_{\Gamma_D} u_0 \tau \cdot n \, ds \).

### 3. Discretization of stress formulations

#### 3.1. Discrete formulation

In this part, which deals with the discretization of the dual problem, the problem is reduced to that of the complementary energy for the sake of simplification. The problem is then expressed without any contact or friction; it is discretized using the equilibrium finite elements method introduced by Fraeijs de Veubeke in [4] and recently used by Kempeneers et al. in [5]. The elementary stress field is obtained from an Airy stress function. This type of element has been used in some numerical cases: by Zavelani-Rossi [6] in a study on plane structures in the case of plasticity and cracks, by Wiedkowsk and al. [7] in the elastoplastic analysis of plane structures, and by Bisegna and et al. [8,9] to solve the Signorini and Coulomb problem in the case of a plane elastic structure. In the last part, we describe how the prescribed and contact forces are accounted in this system.

The problem presented above can be written starting with the complementary energy and solved using a relaxation algorithm. We then write the complementary energy functional \( \pi \) as follows:

\[
\pi = \frac{1}{2} b (\tau, \sigma) - l(\tau).
\]

The discretization of the problem leads to transforming the functional \( \pi \) into a sum of elementary functionals:

\[
\pi = \sum_{e=1}^{N} \pi_e \quad \text{with} : \quad \pi_e = \frac{1}{2} b (\tau^e, \sigma^e) - l(\tau^e) \quad \text{and}
\]

\[
\begin{align*}
 b (\sigma^e, \tau^e) &= \int_{\Gamma_e} \sigma^e_{ij} \tau^e_{ij} \, dx_e, \\
 l(\tau^e) &= \int_{\Gamma_D} \tau^e_{ij} u_0 n_i \, ds_e, \quad \text{if the element edge} \ \Gamma_D
\end{align*}
\]

In each element, one has to satisfy the local equilibrium condition, to satisfy the continuity of the stress vector between two neighbouring elements \( K \) and \( K' \) in order to satisfy the local equilibrium condition on the whole structure, and to introduce the external (prescribed or contact) forces if the edge of the element is in \( \Gamma_f \) or \( \Gamma_c \). These conditions can be written respectively as follows:

\[
\begin{align*}
 \text{div} \ \vec{\tau}_e + F &= 0 \quad \text{in} \ K, \\
 \vec{\tau}_e \vec{n} &= \vec{F} \vec{n} \quad \text{if} \ \Gamma_K = \Gamma_K'.
\end{align*}
\]

![Fig. 2. The Hsieh–Clough–Tocher (HCT) element.](image)
2. Elementary stress field

In order to reduce the computational cost, the elementary stress field interpolation is constructed in such a way that it fulfills the condition (7) a priori. The stress field in an element \( K \) is defined by the curl of an Airy stress function \( \Phi = \Phi(x, y) \), such that:

\[
\begin{align*}
\tau_{xx}^e &= \Phi_{xy}(x, y), \\
\tau_{yy}^e &= -\Phi_{xx}(x, y), \\
\tau_{xy}^e &= \Phi_{yx}(x, y) + \psi,
\end{align*}
\]

With \( \psi \) such that \( F_i = -\psi_n \), where \( F_i \) denotes the volume forces in the direction \( i \). For the sake of simplicity, the problem will be written with no volume forces in the following part of the paper.

The Airy stress function is interpolated on an element by:

\[
\Phi^e = [Y^e(x, y)]^t [\phi^e],
\]

where \( \phi^e \) are the degrees of freedom. The stress field is then linked to the degrees of freedom by differentiating the Airy stress function:

\[
[\tau^e] = \begin{bmatrix} \tau_{xx}^e \\ \tau_{yy}^e \\ \tau_{xy}^e \end{bmatrix} = [B^e]^t [\phi^e].
\]

Using Airy stress functions also makes it possible to meet condition (8) very simply. This condition requires the stress vector to be continuous across elements. Since the stress vector is completely defined on an edge by the value of the Airy stress function and that of its first derivatives, \( C^1 \) regular finite elements were used to ensure the continuity of the stress vector, by performing nodal degrees of freedom assembly.

Here, the triangular composite Hsieh–Clough–Tocher (HCT) (Fig. 2) element is used. This element is built using three Airy subfunctions which allow to define three linear stress fields on each sub-element. The elementary stress field is therefore linear in parts and defined by 12 degrees of freedom. It can be shown that using this element also simplifies the procedure used to impose the contact and friction conditions, since the stress vector is linear on the edges of the element.

3.3. Contact and prescribed forces

At this point, conditions (9) and (10) have not yet been imposed. Since the degrees of freedom are not forces, this cannot be done directly; the stress vector is obtained by projecting the stress field onto an element edge:

\[
T(M(\bar{s}), n) = [n]^t [B^e(\bar{s})]^t [\phi^e],
\]

where \( s \) is the coordinate of a point \( M \) on the edge and \( [n] \) is the projection matrix associated with the outward normal vector \( \hat{n} \) of the edge:

\[
[n] = \begin{bmatrix} 1 & 0 & m \\ 0 & 1 & m \\ 0 & 0 & 1 \end{bmatrix}, \quad \hat{n} = \begin{bmatrix} 1 \\ m \end{bmatrix} \text{ in } (x, y).
\]

We then introduce the force vector \( [T^e] \). To completely define the stress vector on one edge, one has to apply its value at different points on an element edge, the number of which depends on the degree of the stress interpolation function. In the case of HCT, since both components are linear, four values are required.

These values are then placed on the force vector \( [T^e] \), which can be directly related to the degrees of freedom by the force interpolation matrix \( [C^e] \):

\[
[T^e] = [C^e]^t [\phi^e].
\]

Conditions (9) and (10) are then directly imposed by introducing the vector \( [F^e] \) (respectively \( [F^e_c] \)), which gives the discrete values of the external forces \( f \) prescribed on the element edges (contact and friction forces \( f_c \), respectively):

\[
[F^e] = [C^e]^t [\phi^e],
\]

\[
[F^e_c] = [C^e]^t [\phi^e].
\]

The external forces are then prescribed by fixing the values of \( [F^e] \).

3.4. Elementary problem

The integral \( I(\tau^e) \) is discretized as follows:

\[
I(\tau^e) = \int_{\Gamma_p} \langle [\tau^e]^t [n] \rangle [u] ds_x = [\phi^e]^t \int_{\Gamma_p} [C^e]^t [N]^t [u] ds_x,
\]

where \( [N] \) is the displacement interpolation matrix and \( [u] \) the nodal displacements. Using Gauss integration formulas, we obtain:

\[
I(\tau^e) = [\phi^e]^t [q^e].
\]

The flexibility matrix \( [S^e] \) is obtained by Gauss integration and the elementary functional becomes:

\[
\pi_e = \frac{1}{2} [\phi^e]^t [S^e][\phi^e] - [\phi^e]^t [q^e],
\]

\[
[F^e] = [C^e]^t [\phi^e], \quad [F^e_c] = [C^e]^t [\phi^e].
\]

The last two conditions are included using Lagrangian multipliers. We obtain the following complementary energy functional:

\[
\pi^* = \frac{1}{2} [\phi^e]^t [S^e][\phi^e] - [\phi^e]^t [q^e] + [\lambda](([C^e][\phi] - [F]) + [\xi^e](([C^e][\phi] - [F]))).
\]

Cancelling the first variation of this functional with respect to the variables \( \phi, \lambda, \xi^e \), and \( F_c \) gives the global matrix system:

\[
\begin{bmatrix} S & C^t_F & C^t_C \\ C_F & 0 & 0 \\ C_C & 0 & -I \end{bmatrix} \begin{bmatrix} \phi \\ \lambda \\ \xi^e \end{bmatrix} = \begin{bmatrix} q \\ F \\ F_c \end{bmatrix}.
\]

At this point, the contact and friction conditions have to be applied. In the primal case, when dealing with displacements, the matrix is positive definite and the system can be solved using relaxation algorithms. Here the matrix is not positive definite. We therefore compared two ways of solving this system: by using a Gauss–Seidel algorithm to solve a smaller condensed system, which can be obtained using various condensation strategies, and by solving the problem, given by the augmented Lagrangian formulation, using the Uzawa algorithm.
procedure, and \( n_f \) be the number of prescribed forces (which usually amount to less than 20% of \( n_s \) in many contact problems) and \( n_c \) be the number of contact forces (which usually amount to less than 10% of \( n_s \) in many contact problems). Table 1 summarises the characteristics of each matrix and gives storage schemes that can possibly be used.

4.1. Condensation on the contact boundary

The goal of the condensation procedure is to obtain the reduced (condensed) system:

\[
[D\{F\}C] = [G]
\]

to be solved with a relaxation algorithm. This condensed system can be obtained using various methods: the “blind condensation” method, or condensation using either the SYMMLQ solver or a three-step solver.

4.1.1. Blind condensation

The first method of obtaining this system consists in applying the basic idea of condensation, starting with system (28) and taking the first three groups of unknowns as the internal unknowns \( \phi_i \).

The symbolic system to be reduced is therefore:

\[
\begin{bmatrix}
M_{11} & M_{12} \\
M_{21} & 0
\end{bmatrix}
\begin{bmatrix}
\phi_i \\
U_i
\end{bmatrix}
= 
\begin{bmatrix}
U_i \\
0
\end{bmatrix}.
\]

By solving:

\[
\phi_i = M_{11}^{-1}U_i - M_{11}^{-1}M_{12}F_C
\]

the following condensed system is obtained:

\[
M_{21}M_{11}^{-1}M_{12}F_C = M_{21}M_{11}^{-1}U_i \quad \text{denoted } [D\{F\}C] = [G],
\]

where

\[
M_{11} = \begin{bmatrix}
S & C_F^T \\
C_F & 0 & 0 \\
C_C & 0 & 0
\end{bmatrix}.
\]

Since the computational cost of inverting \( M_{11} \) could be very high, (31) can be solved using a linear solver: \( M_{11} \) is factorised once and \( \phi_i \) is obtained by \((1+n_c)\) triangular matrix multiplications. Since the matrix \( M_{11} \) is not positive definite and cannot be stored using an efficient storage scheme, the first method is very costly and it has not been used here.

4.1.2. SYMMLQ condensation

Since the slowness of the above condensation strategy is due to the inefficiency of the direct solvers applied to non-positive definite matrices, the SYMMLQ iterative method was used to solve the previous \((1+n_c)\) systems. The SYMMLQ method has been presented in [12]: it is a conjugate gradient like method for solving symmetric indefinite linear systems. It solves the projected system and keeps the residual vectors orthogonal to all the previous ones. The efficiency of this algorithm depends heavily on the choice of the preconditioner. An efficient diagonal block preconditioner is used in [13]. Let us simplify the notation of (33) as follows:

\[
M_{11} = \begin{bmatrix}
S & C^T \\
C & 0
\end{bmatrix}.
\]

The preconditioner matrix \( P \) is the block diagonal matrix:

\[
P = \begin{bmatrix}
S' & 0 \\
0 & -T
\end{bmatrix}.
\]

where \( S' \) is an approximation for \( S \). Various preconditioners were tested in this study: incomplete Cholesky factorizations of \( S \), the SSOR preconditioner and, since the SYMMLQ solver is used \((1+n_c)\) times during the condensation process, complete Cholesky factorization of \( S \) was also tested. The second block preconditioner used here is such that \( T = CS^{-1}C^T \).

4.1.3. Three-step condensation

An efficient condensation method was developed to be able to take the structure of each matrix into account. Note that in what follows, the brackets will be omitted from the matrices to simplify the notations. The matricial system (28) can be written as a system of four equations as follows:

\[
\begin{align*}
S\phi + C_F^T \lambda + C_F^T \lambda' &= q, \\
C_F \phi &= F, \\
C_C \phi &= F_C, \\
\lambda' &= 0.
\end{align*}
\]

The condensation procedure involves three main steps:

Step 1: Evaluating \( \phi \) so that

\[
\phi = S^{-1}q - S^{-1}C_F^T \lambda - S^{-1}C_F^T \lambda'
\]

can be achieved in two ways:

- Using a direct solver: \( S \) is factorized once by performing LDL\(^T\) factorization with skyline storage and the products \( S^{-1}q \), \( S^{-1}C_F^T \lambda \) and \( S^{-1}C_F^T \lambda' \) are obtained by \((1+n_f+n_c)\) system resolutions. The costliest part of this step is therefore carried out only once.
- Using a conjugate gradient (CG) algorithm: an efficient preconditioner is computed once and used to quickly obtain the products given above as the result of the \((1+n_f+n_c)\) system resolutions. \( \phi \) is then introduced into lines 2 and 3 of (36).

Step 2: The system then becomes:

\[
\begin{align*}
C_F S^{-1}q - C_F S^{-1}C_C^T \lambda - C_F S^{-1}C_C^T \lambda' &= F, \\
C_C S^{-1}q - C_C S^{-1}C_F^T \lambda - C_C S^{-1}C_F^T \lambda' &= F_C, \\
\lambda' &= 0.
\end{align*}
\]

\( \lambda \) is obtained by performing \((1+n_c)\) resolutions of a full and symmetric \( n_f \) by \( n_f \) matrix. Here again, the matrix is factorised only once.

Step 3: \( \lambda \) is then introduced into line 2 of (38) and \( \lambda' \) is obtained in the same way (by performing \((1+n_c)\) resolutions of a full and symmetric \( n_c \) by \( n_c \) matrix). Lastly, \( \lambda' \) is introduced into the last line of (38) and the condensed system is obtained. Note that the last two steps involve full and symmetric matrices, but since \( n_f \) and \( n_c \) are small in comparison with \( n_s \), their factorization is not too costly.

4.1.4. Preconditioners

Since the efficiency of iterative methods depends strongly on the preconditioning procedure, the following preconditioners have been used to accelerate both the SYMMLQ and CG methods.

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Size</th>
<th>Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S )</td>
<td>( n_f ) by ( n_f )</td>
<td>Symmetric (skyline or sparse)</td>
</tr>
<tr>
<td>( C_F )</td>
<td>( n_f ) by ( n_f )</td>
<td>Sparse</td>
</tr>
<tr>
<td>( C_C )</td>
<td>( n_c ) by ( n_c )</td>
<td>Sparse</td>
</tr>
</tbody>
</table>
4.1.5. Incomplete Cholesky factorization

When direct solvers are used, the symmetric positive definite matrix $S$ is decomposed into a lower and an upper matrix so that:

$$ S = LL^T.$$  (39)

The aim of incomplete Cholesky factorization is to find an approximation for $L$ named $L'$. The classical incomplete Cholesky factorisation strategies are reviewed in [14]. We keep only the most important terms of $L'$. For this purpose, two methods are used:

- The drop tolerance factorization strategy: The term $L'(i,j)$ is kept only if $L'(i,j) > 10^{-1} \cdot \sqrt{L(i,i) \cdot L(j,j)}$.
- The filling level factorization strategy: The largest $n_3 \times n_3$ terms in line $i$ of $L'$ are kept, where $n_3$ is the size of line $i$ of $L$ and $n_3$ is a parameter on which the filling level of $L'$ depends.

With both methods, two strategies can be used, and the least important terms of $L'$ can be removed during the factorization procedure or removed after performing the full Cholesky factorization procedure. The former strategy is usually used in order to minimise the computational cost of $L'$ and the memory requirements, but this sometimes makes it necessary to add small diagonal terms $\delta$ to $S$. Since the preconditioner is used many times in the condensation process, the second strategy can be a useful means of maximizing the efficiency of the preconditioner, although the computational cost of the preconditioner will be greater in this case.

In order to compare those two strategies, the first one was applied to the filling level factorization strategy, and the second one to the drop tolerance factorization strategy.

4.1.6. SSOR preconditioner

The matrix $S$ is decomposed so that:

$$ S = D - E - F^T,$$  \hspace{1cm} (40)

Until convergence:

For $i = 1$ to $n_f$:

$$ F_C^i = \frac{1}{D(i,i)} [G(i) - \sum_{j=1}^{i-1} D(i,j) F_C(j) - \sum_{j=i+1}^{n_f} D(i,j) F_C(j)]$$

- If $F_C(i)$ is normal: If $F_C^2(i) \geq 0$
  - Then: $F_C(i) = 0$
  - Else: $F_C(i) = F_C^2$

- If $F_C(i)$ is tangent: If $|F_C^2| > \mu|F_Cv|$ (Where $F_Cv$ is the normal force corresponding to $F_C(i)$)
  - Then: $F_C(i) = \mu \text{ sign}(F_C^2) |F_Cv|$
  - Else: $F_C(i) = F_C^2$

End of loop

Convergence if $i_C^k$ small

End of loop

Fig. 3. The Gauss-Seidel algorithm.
5. Numerical results

The algorithms presented here have been implemented in the computer code LMGC90 (http://www.lmgc.univ-montp2.fr/~du-bois/LMGC90/) and tested on the case of the steel tooth presented in Fig. 5, with displacements \( u_x = 0 \) cm and \( u_y = 0.01 \) cm prescribed on its left edge, which is in contact with a rigid foundation where \( \mu = 0.2 \).

We first compare the accuracy of the results obtained using the method presented here and those obtained using the classical displacement based method, and we then compare the efficiency of each of the solvers tested.

5.1. Validity of the method

In order to check the validity of the present dual method, we compare the stress fields obtained using this method with the stress fields obtained using the classical displacement based method, and we then compare the efficiency of each of the solvers tested.

The stress fields obtained using the classical displacement based finite element method on various homogeneous meshes are obtained using the computer code LMGC90, which is based on the NSCD solver [16], with linear triangular elements.

The relative error \( e \) is computed by performing numerical integration on the Gauss points of the reference mesh. The evolution of the error is shown in Fig. 6 in the case of the present method (named DUAL) and the displacement based method (named PRIMAL) with various homogeneous meshes.

These results clearly show that both methods converge towards the same reference solution. Comparisons on the relative error obtained with the DUAL and PRIMAL methods with an equivalent number of degrees of freedom show that the present method gives more accurate results in terms of the stresses in the case of the example tested.

5.2. Comparisons between numerical methods

5.2.1. Descriptions of benchmarking examples and algorithms parameters

In this section, we test the present methods of resolution on the examples given in Figs. 5 (Mesh0 to Mesh3) and 7 (Mesh4 to Mesh7) with the various sizes presented in Table 2. This second...
example is similar to the first one, but involves four separated contact boundaries and different ratios to when not particularly refined meshes are used. All the meshes were previously renumbered using the Cuthill McKee algorithm (see [17]) and the matrices were equilibrated using diagonal equilibration method to improve the convergence rates and the accuracy.

The following methods were then compared:

- The SYMMLQ condensation strategy.
- The three-step condensation method using two different approaches at the first level of condensation: a skyline LDL\(^T\) solver and a conjugate gradient solver. These two methods will be named 3STEPD and 3STEPCG below.
- Solving the augmented Lagrangian formulation using the Uzawa algorithm.

With the CG and SYMMLQ methods, the following approximations \( S' \) of \( S \) were tested as preconditioners:

- The drop tolerance incomplete Cholesky factorization (DTIC) procedure with the tolerance parameter \( c \in [3; 6] \).
- The filling level incomplete Cholesky factorization (FLIC) procedure with the filling level factor \( n_f \in [1; 5] \).
- The SSOR preconditioner (SS) with the relaxation parameter \( \omega \in [0; 2] \).
- Exact Cholesky factorization of \( S \) (SKY).

When DTIC, FLIC and SSOR preconditioners were used, the upper and lower parts of these preconditioners were stored using sparse line storage schemes. Both parts were stored in order to accelerate the matrix vector products. When the exact Cholesky factorization procedure was used, both the upper and lower parts were stored for the same reason and a skyline storage line scheme was used. When using the skyline LDL\(^T\) solver, both the lower \((-E)\) and upper \((-E^T)\) parts of \( S \) were stored using a skyline storage line scheme. Since the computation of line \( i \) of \( L \) does not require the \((i - 1)\) lines of \( S \), matrices \( L \) and \( L^T \) can be stored in \(-E\) and \(-E^T\), respectively.

To use Uzawa’s algorithm, one has to choose the parameters \( \rho \) and \( r \) defined in Section 4.3. Since the parameter \( r \) defines the importance of the product matrices \( C_C^T \) and \( C_C^T \) in comparison with \( S \), matrices \( C_C \) and \( S \) have been multiplied by factors \( p_F \) and \( p_S \) so that:

\[
\begin{align*}
    p_F &= \frac{\text{mean}(S)}{\text{mean}(C)}, \\
p_S &= \frac{\text{mean}(S)}{\text{mean}(C_C)}
\end{align*}
\]

This makes it possible to give any \( r \) the same weight in all the examples tested. On the other hand, since we had no rules for choosing \( \rho \) and \( r \), we tested various combinations of these two parameters in \([1; 20; 40; 60; 80; 100; 200; 300]\) in each case and took the best sets for the subsequent comparisons.

These computations were performed on an Intel P4 1.7 Ghz computer with 1 Gb RAM memory in order to assess the performances on a medium-range computer.

5.2.2. Influence of the preconditioner

Before comparing these methods, we analyzed the influence of preconditioners on the SYMMLQ and 3STEPCG methods. Since similar behaviour was found to occur in each case and with both methods, we focused on results obtained with the SYMMLQ method on the Mesh3 example given in Table 3. This table gives the name of each preconditioner, the value of the parameter \( c \) with DTIC, \( n_f \) with FLIC and \( \omega \) with SSOR, the time taken to compute the preconditioner matrix, the time required to perform the condensation, the time required to complete the entire resolution (condensation and Gauss–Seidel methods), the size that would be used by the lower part of the full Cholesky factorization preconditioner, the size of the lower part of the preconditioner, and the maximum RAM memory used.

First, if we look at the DTIC, it can be seen that DTIC3 gives a good compromise between memory requirements and speed, and that DTIC6 gives the best results, since it is similar, in terms of size, to the full Cholesky factorization. It is worth noting that the RAM value of DTIC6 is greater than the one of SKY because that the skyline storage scheme is lighter than the sparse storage scheme, when dealing with the same number of components.

### Table 2

<table>
<thead>
<tr>
<th>Name</th>
<th>( n_0 )</th>
<th>( n_f )</th>
<th>( n_c )</th>
<th>Name</th>
<th>( n_0 )</th>
<th>( n_f )</th>
<th>( n_c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesh0</td>
<td>761</td>
<td>96</td>
<td>38</td>
<td>Mesh4</td>
<td>8360</td>
<td>392</td>
<td>200</td>
</tr>
<tr>
<td>Mesh1</td>
<td>1973</td>
<td>144</td>
<td>80</td>
<td>Mesh5</td>
<td>19128</td>
<td>588</td>
<td>480</td>
</tr>
<tr>
<td>Mesh2</td>
<td>3105</td>
<td>176</td>
<td>120</td>
<td>Mesh6</td>
<td>26538</td>
<td>708</td>
<td>576</td>
</tr>
<tr>
<td>Mesh3</td>
<td>6287</td>
<td>232</td>
<td>160</td>
<td>Mesh7</td>
<td>50483</td>
<td>920</td>
<td>640</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Name</th>
<th>Parameter</th>
<th>Tprec (s)</th>
<th>Tcond (s)</th>
<th>Ttotal (s)</th>
<th>Full size (nb terms)</th>
<th>Lprec size (nb terms)</th>
<th>Maximum (Mb)</th>
<th>( \delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTIC2</td>
<td>2</td>
<td>5.92</td>
<td>258.66</td>
<td>258.72</td>
<td>1264705</td>
<td>171548</td>
<td>31.84</td>
<td></td>
</tr>
<tr>
<td>DTIC3</td>
<td>3</td>
<td>6.61</td>
<td>91.50</td>
<td>91.56</td>
<td>1264705</td>
<td>353749</td>
<td>36.66</td>
<td></td>
</tr>
<tr>
<td>DTIC4</td>
<td>4</td>
<td>9.73</td>
<td>58.54</td>
<td>58.60</td>
<td>1264705</td>
<td>655198</td>
<td>43.58</td>
<td></td>
</tr>
<tr>
<td>DTIC5</td>
<td>5</td>
<td>13.45</td>
<td>43.50</td>
<td>43.56</td>
<td>1264705</td>
<td>1005769</td>
<td>51.59</td>
<td></td>
</tr>
<tr>
<td>DTIC6</td>
<td>6</td>
<td>15.52</td>
<td>41.10</td>
<td>41.16</td>
<td>1264705</td>
<td>1204640</td>
<td>56.14</td>
<td></td>
</tr>
<tr>
<td>FLIC1</td>
<td>1</td>
<td>5.99</td>
<td>546.54</td>
<td>546.60</td>
<td>1264705</td>
<td>80018</td>
<td>29.74</td>
<td>3.5 \times 10^{-2}</td>
</tr>
<tr>
<td>FLIC2</td>
<td>2</td>
<td>5.98</td>
<td>292.95</td>
<td>293.01</td>
<td>1264705</td>
<td>159908</td>
<td>31.64</td>
<td>5 \times 10^{-1}</td>
</tr>
<tr>
<td>FLIC3</td>
<td>3</td>
<td>13.11</td>
<td>259.99</td>
<td>259.05</td>
<td>1264705</td>
<td>339607</td>
<td>33.31</td>
<td>5 \times 10^{-1}</td>
</tr>
<tr>
<td>FLIC4</td>
<td>4</td>
<td>19.23</td>
<td>285.38</td>
<td>285.44</td>
<td>1264705</td>
<td>317324</td>
<td>35.85</td>
<td>5 \times 10^{-1}</td>
</tr>
<tr>
<td>FLIC5</td>
<td>5</td>
<td>24.65</td>
<td>315.80</td>
<td>315.86</td>
<td>1264705</td>
<td>393895</td>
<td>37.60</td>
<td>5 \times 10^{-1}</td>
</tr>
<tr>
<td>SKY</td>
<td>6.91</td>
<td>14.50</td>
<td>14.56</td>
<td>14.56</td>
<td>1264705</td>
<td>1264705</td>
<td>47.87</td>
<td></td>
</tr>
<tr>
<td>SS02</td>
<td>0.2</td>
<td>2.46</td>
<td>1571.28</td>
<td>1571.34</td>
<td>1264705</td>
<td>80018</td>
<td>30.67</td>
<td></td>
</tr>
<tr>
<td>SS04</td>
<td>0.4</td>
<td>2.61</td>
<td>1258.93</td>
<td>1258.99</td>
<td>1264705</td>
<td>80018</td>
<td>30.67</td>
<td></td>
</tr>
<tr>
<td>SS06</td>
<td>0.6</td>
<td>2.64</td>
<td>1081.68</td>
<td>1081.74</td>
<td>1264705</td>
<td>80018</td>
<td>30.67</td>
<td></td>
</tr>
<tr>
<td>SS08</td>
<td>0.8</td>
<td>2.62</td>
<td>975.84</td>
<td>975.90</td>
<td>1264705</td>
<td>80018</td>
<td>30.67</td>
<td></td>
</tr>
<tr>
<td>SS10</td>
<td>1</td>
<td>2.60</td>
<td>921.31</td>
<td>921.37</td>
<td>1264705</td>
<td>80018</td>
<td>30.67</td>
<td></td>
</tr>
<tr>
<td>SS12</td>
<td>1.2</td>
<td>2.65</td>
<td>926.26</td>
<td>926.32</td>
<td>1264705</td>
<td>80018</td>
<td>30.67</td>
<td></td>
</tr>
<tr>
<td>SS14</td>
<td>1.4</td>
<td>2.70</td>
<td>984.78</td>
<td>984.84</td>
<td>1264705</td>
<td>80018</td>
<td>30.67</td>
<td></td>
</tr>
<tr>
<td>SS16</td>
<td>1.6</td>
<td>2.78</td>
<td>1136.87</td>
<td>1136.93</td>
<td>1264705</td>
<td>80018</td>
<td>30.67</td>
<td></td>
</tr>
<tr>
<td>SS18</td>
<td>1.8</td>
<td>2.88</td>
<td>1457.94</td>
<td>1458.00</td>
<td>1264705</td>
<td>80018</td>
<td>30.66</td>
<td></td>
</tr>
</tbody>
</table>
The last column gives the value $\delta$ added to the diagonal of $S$ when the FLIC preconditioner was used. This procedure involves adding $\Delta \delta$ to the diagonal when the algorithm fails to compute the preconditioner and the computation has to be restarted. The times given in Table 3 take these restarts into account and $\Delta \delta = 0.005$. It can be seen that the strategy which consists in taking the $(i - 1)$ first lines of $L$ to compute the $i$th line usually involves restarts and therefore does not make for any computational time savings.

The optimum parameter $\omega$ used for the SSOR preconditioner was 1 in all the cases tested here. This preconditioner gives a good approximation for $S$, but it is less efficient than the incomplete Cholesky preconditioner when working on problems of this kind.

The SKY preconditioner stored on a skyline scheme eventually gave the best results. First since choosing the exact factorization procedure for $S$ gives an excellent approximation of $P$, the block preconditioner; the SYMMLQ algorithm converges in 1 or 2 iterations with this preconditioner. Secondly, the skyline storage scheme plays an important role: to confirm this point, the DTIC preconditioner was tested with a threshold value $c = 10^{-21}$; this preconditioner therefore gives very similar results to those obtained with the full Cholesky preconditioner. But due to the sparse storage scheme involved, the SYMMLQ method takes longer with the DTIC preconditioner than with the SKY preconditioner.

5.2.3. Comparisons between algorithms

In the previous section, efficient preconditioners were selected in order to reduce the number of methods compared. In this part, we use only the DTIC and SKY preconditioners with iterative methods. The results obtained with the following methods on all the examples are compared in Figs. 8 and 9:

- The SYMMLQ condensation method with the DTIC3, DTIC6 and SKY preconditioners, which was named SDTIC3, SDTIC6 and SSKY, respectively.
- The 3STEP condensation method with the direct $LDL^T$ and the conjugate gradient solver, using the DTIC3 and DTIC6 preconditioners, which was named 3STEPD, 3STEPCG3 and 3STEPCG6, respectively.
- The solution of the augmented Lagrangian formulation using the Uzawa algorithm, which was named ALM.

The computational times are summarized in Fig. 8.

The memory requirements of all the methods tested here are summarized in Fig. 9.

First, by comparing SDTIC3 with 3STEPCG3 and SDTIC6 with 3STEPCG6, it can be seen that the SYMMLQ method and the 3STEP-CG method are both equivalent when equivalent preconditioners are used, but these four strategies are less efficient than the others. As a general rule, iterative methods can only be as efficient as the 3STEPD and ALM methods when highly efficient preconditioners are used, as in the case of the SSky strategy. But these methods involve the storage of both the global matrix and the preconditioner; they therefore require a greater amount of memory than the 3STEPD and ALM methods.

The 3STEPD and ALM methods therefore seem to be the most efficient of these methods. In all the examples tested, these two methods required similar computational times, although the ALM method was found to be slightly faster.
The main advantage of the ALM method is that it requires very little memory, whereas the 3STEPD methods involve the storage of intermediate matrices at each of the condensation steps. This drawback of the 3STEPD method could be overcome using an appropriate file storage strategy.

The main advantage of the 3STEPD method is its robustness, since unlike the ALM method, it is not parameter dependent. The computation times are therefore predictable, since they depend mainly on the profile of the flexibility matrix. On the other hand, this method can easily be parallelized.

6. Conclusion

In this paper, an adapted version of the equilibrium finite element method is presented for solving contact and friction problems. With the element presented, the interpolated stress fulfills the equilibrium conditions a priori.

The use of equilibrium finite elements requires the prescribed and contact forces to be introduced into the matrix system using Lagrangian multipliers. The matrix system is thus non-positive definite and the global system is non-linear and constrained. Various solution algorithms based on condensation and on an augmented Lagrangian formulation have been proposed to solve this particular system.

The validity of the equilibrium finite element method was confirmed in the case of various examples. Results obtained in the case of a tooth using both the classical displacement based finite element method and that proposed here were compared in the last section. The finite element method presented here gives more accurate results in terms of the stresses than the classical one. This can be explained by the fact that the contact and friction conditions are directly imposed on the stress field in the case of the present method.

Lastly, all the numerical methods developed so far for solving the global system were compared. Two of these methods turned out to be particularly efficient. The first one, which is based on the use of a suitable condensation strategy and of direct solvers, is very robust but needs some memory management. The second method, which is based on the Uzawa algorithm applied to an augmented Lagrangian formulation requires little memory, but some difficult parameters have to be determined to make it efficient.

Appendix A. Some convergence results

In this paragraph, some convergence results obtained using the finite elements method in a simplified framework are presented. We introduce the space

$$W = H(div; \Omega) = \{ \tau \in H; div\tau \in [L^2(\Omega)]^d \}. \tag{49}$$

To simplify the problem, the domain $$\Omega$$ is assumed to be polygonal. We take $$\mathcal{T}_h$$ to denote a triangulation of $$\Omega$$ such that $$\Omega = \mathop{\cup}_{T \in \mathcal{T}_h} T$$. $$\mathcal{T}_h$$ is assumed to be regular, so that all the angles of the element $$T \in \mathcal{T}_h$$ are bounded far from zero and there exists a positive constant $$\varepsilon$$ such that the length of any side of any $$T \in \mathcal{T}_h$$ is at least $$\varepsilon h$$, where $$h$$ is the characteristic length of the elements.
In a general framework, spaces \( W_h \) and \( V_h \) are chosen so that: if \( \tau \in W_h \) and \( (d\nu, \tau) = 0 \ \forall \nu_h \in V_h \) then \( d\nu = 0 \) in \( \Omega \).

(50)

In addition, we assume that it is possible to construct an interpolation operator \( \pi_h : W \to W_h \) fulfilling:

\( (d\nu, \pi_h \sigma) = (d\nu, \sigma) \ \forall \nu_h \in V_h \).

An element is defined in keeping with the above hypothesis. For all \( T \in T_h \), the following finite dimensional space of piecewise linear stress tensors is defined:

\[ W_T = H(d\nu; T) \]

and we write

\[ W_h = \{ \tau \in W_h ; \nu = [L^2(\Omega)]^2 \}, \]

where

\[ \bar{W}_h = \{ \tau \in H; \nu_T \in W_T \forall T \in T_h \}. \]

(54)

We now adopt some restrictive hypotheses. \( F \) (the weight, for example) is assumed to be constant and \( f \) (the linear force distribution) to be linear. Here we take the case of Tresca friction: in problems \( (P_i) \) and \( (P_i^r) \), the sets \( \Sigma(\sigma) \) and \( \Sigma'(\sigma) \) are replaced by the sets \( \Sigma(\beta) \) and \( \Sigma'(\beta) \), respectively, where \( \beta \) and \( \beta_r \) are given. \( \beta \) (resp. \( \beta_r \)) are assumed to be concave or piecewise linear. Using a classical result obtained by Carlet [10], one can obtain:

Elasticity: there exists a constant \( C_h \) independent of \( h \) such that:

\[ \|\sigma - \sigma_h\|_W \leq C_h \|\sigma\|_2. \]

Tresca \( (P_h) \) there exists a constant \( C_h \) independent of \( h \) such that:

\[ \|\sigma - \sigma_h\|_W \leq C_h \|\sigma\|_2. \]

Tresca \( (P_r) \) there exists a constant \( C_h \) independent of \( h \) such that:

\[ \|\sigma - \sigma_h\|_{-1,2,2,2} \leq C_h \|\sigma\|_2. \]

(57)

These results are proved in [11].

Comment: In the case of the Signorini–Coulomb problem, the hypotheses adopted in theorem (7.1) are generally not satisfied. It is therefore necessary to use external approximations [10]. This problem still remains to be solved.

A.1. Accuracy of the Block preconditioner

In this section, we explain why the Block preconditioner presented in Section 4.1.2 is very accurate when the full Cholesky factorization method is used as the first block of the preconditioner. Let us first recall the condensation principle. To obtain expression (31), one has to solve a series of systems:

\[ M_{11}U_i = U_i \text{and} \]

\[ M_{11}X = M_{12} \text{ treated by } n_c \text{ operations:} \]

\[ \text{do } i = 1 \text{ to } n_c : \]

\[ M_{11}(i,:) = M_{12}(i,:) \] (where \( X(:,i) \) denotes the column i of \( X \))

enddo

These \( (1 + n_c) \) systems can be solved using the SYMMLQ method with the block preconditioner (35). When the full Cholesky factorization procedure is used as the first block of the preconditioner, one of the \( n_c \) systems solved by applying the preconditioner (35) is:

\[ \begin{bmatrix}
S & 0 \\
0 & T
\end{bmatrix}
\begin{bmatrix}
X_1(:,i) \\
X_2(:,i)
\end{bmatrix}
= 
\begin{bmatrix}
M_{11}(i,:) \\
M_{12}(i,:)
\end{bmatrix}, \]

(59)

which is equivalent to:

\[ \begin{bmatrix}
SX_1(:,i) = M_{11}(i,:), \\
-C^tC X_2(:,i) = M_{12}(i,)
\end{bmatrix}. \]

(60)

The exact system is:

\[ \begin{bmatrix}
S & C^t \\
C & 0
\end{bmatrix}
\begin{bmatrix}
X_1(:,i) \\
X_2(:,i)
\end{bmatrix}
= 
\begin{bmatrix}
M_{11}(i,:) \\
M_{12}(i,:)
\end{bmatrix}, \]

(61)

which is equivalent to:

\[ \begin{bmatrix}
SX_1(:,i) = M_{11}(i,:) - C^tC X_2(:,i) \\
-C^tC X_2(:,i) = M_{12}(i,)
\end{bmatrix}. \]

(62)

Note that this could be a two-step condensation strategy.

With this choice of preconditioner, systems (60) and (62) can be seen to be very similar; the SYMMLQ method therefore converges after only one or two iterations when this preconditioner is used.

References