

A TWO-LEVEL ITERATION METHOD FOR SOLUTION OF CONTACT PROBLEMS

Bahram NOUR-OMID* and Peter WRIGGERS**

Center for Pure and Applied Mathematics, University of California, Berkeley, CA 97420, U.S.A.

Received 1 November 1984

The merits and limitations of some existing procedures for the solution of contact problems, modeled by the finite element method, are examined. Based on the Lagrangian multiplier method, a partitioning scheme can be used to obtain a small system of equation for the Lagrange multipliers which is then solved by the conjugate gradient method. A two-level contact algorithm is employed which first linearizes the nonlinear contact problem to obtain a linear contact problem that is in turn solved by the Newton method. The performance of the algorithm compared to some existing procedures is demonstrated on some test problems.

1. Introduction

Presently, there are a number of procedures that are employed for the solution of the system of equations arising from contact problems. In this report we look at some of the commonly used methods and evaluate their performance. We concentrate on algorithms based on Lagrange multiplier and penalty methods. In the early finite element applications, contact problems were solved by the Lagrange multiplier approach [2, 7]. For each constraint condition a Lagrange parameter is introduced that appears in the list of the unknowns. Hence, the dimension of the resulting system of equations will increase. In addition, the associated tangent matrix is indefinite and has zero diagonal entries that pose some difficulties in the solution step. These shortcomings motivated the use of penalty formulation [8].

The penalty approach results in solutions that satisfy the contact conditions only approximately. The accuracy of the approximate solution depends strongly on the penalty parameter. The correct choice for this parameter is the essence of the algorithm.

The above two approaches have certain advantages as well as drawbacks. Here we attempt to eliminate some of these limitations. In Section 2 we state the different ways of modeling the constraint conditions for contact problems. All these models give rise to systems of equations that have identical structure. We then derive the Lagrange multiplier and the penalty methods and demonstrate the inter-relation between the two methods. A new method is constructed, based on the Lagrange multiplier method, in Section 3. The system of equations is partitioned into blocks. The primary unknowns are eliminated resulting in a small system of equations for

*This author gratefully acknowledges partial support by the Office of Naval Research under contract N00014-76-C-0013.

**Visiting Scholar, Department of Civil Engineering, University of California, Berkeley CA 94720, U.S.A. This author gratefully acknowledges support by the Deutsche Forschungsgemeinschaft.

the Lagrange parameters. These equations are then solved by the conjugate gradient method [6]. In Section 4 we use the above partitioning approach to construct the contact algorithm. The algorithm is a two-level iterative method which first linearizes the nonlinear contact problem to obtain a linear contact problem that is in turn solved by the Newton method. In Section 5 we compare the performance of the algorithm with some existing procedures using a number of different test problems.

2. Existing solution procedure for contact problems

2.1. Statement of the problem

In the context of linearized elasticity theory a number of different approaches have been developed for the solution of the equations arising in contact problems. These problems can be viewed as the minimization of the potential energy subject to certain kinematic constraints. The equilibrium state is achieved when the displacement field \mathbf{v} minimizes the potential energy

$$\bar{\pi}(\mathbf{v}) = \frac{1}{2} \int_{B^1 \cup B^2} W(\mathbf{v}) dV - \int_{B^1 \cup B^2} \rho \mathbf{b} \cdot \mathbf{v} dV - \int_{\partial_\sigma B^1 \cup \partial_\sigma B^2} \bar{\mathbf{t}} \cdot \mathbf{v} dA, \quad (2.1)$$

and satisfies the constraint condition on $\partial B^1 \cap \partial B^2$:

$$(\mathbf{v}^2 - \mathbf{v}^1) \cdot \mathbf{n} + \gamma \geq 0. \quad (2.2)$$

B^1 and B^2 denote the two bodies that come into contact; \mathbf{v}^1 and \mathbf{v}^2 are the displacement fields of B^1 and B^2 ; W is the strain energy; $\rho \mathbf{b}$ is the inertial forces; $\bar{\mathbf{t}}$ is the traction vector; \mathbf{n} is the unit vector normal to the contact surface and γ is the initial gap between the two bodies in the direction of \mathbf{n} .

The application of a standard finite element procedure leads to the discrete form of (2.1),

$$\bar{\pi}(\mathbf{u}) = \frac{1}{2} \mathbf{u}^T \mathbf{K} \mathbf{u} - \mathbf{u}^T \mathbf{f}, \quad (2.3)$$

where \mathbf{u} is the displacements vector of the nodes in the mesh; \mathbf{K} is the associated stiffness matrix; and \mathbf{f} is the vector of the nodal forces which is obtained from the last two integrals in (2.1).

The discrete counterpart of the constraint equation (2.2) can be given in a number of different forms depending on the modeling of the contact condition. We will restrict our attention to two-dimensional problems modeled by four-node elements.

Certain discretizations result in a mesh that assures node-to-node contact for the linear model. This leads to restrictive but simple constraint conditions. The model may be improved by allowing node-to-surface contact. For the class of Hertzian contact problems we assume a plane contact surface with a determined normal. In general, the contact surface is not plane and a reference body must be chosen to define the normal vector \mathbf{n} to the contact surface. The contact surface on the reference body is referred to as the master surface and that on the second body is called the slave surface. This leads to a family of models for choosing the

normal. In Fig. 1 we give four different ways of modeling the contact condition at each node. e_l is the l th column of the identity matrix and α is a surface coordinate defining the point of contact.

Our objective here is to examine different procedures for the solution of the equations arising from the contact problem. For this purpose, each of the constraint equations given in Fig. 1 can be stated in the form

$$\mathbf{b}^T \mathbf{u} \geq \gamma, \quad (2.4)$$

where \mathbf{b} depends on the choice of the model. Furthermore, the contact stresses normal to the contact surface, λ , must satisfy the condition

$$\lambda \leq 0. \quad (2.5)$$

Equations (2.4) and (2.5) can be combined to obtain

$$\lambda (\mathbf{b}^T \mathbf{u} - \gamma) = 0, \quad (2.6)$$

which is known as the Kuhn–Tucker condition.

2.2. Lagrange multiplier method

Using the Lagrangian multiplier approach, the contact condition, (2.4), can be added to the


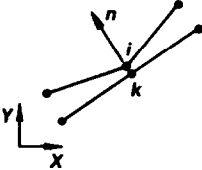
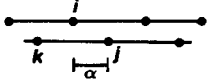
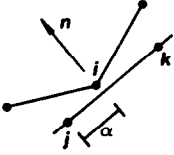
	Hertzian	General
Node-To-Node	 $(\mathbf{e}_i^T - \mathbf{e}_j^T) \mathbf{u} \geq \gamma$	 $\mathbf{n}^T \left(\begin{bmatrix} \mathbf{e}_x^T \\ \mathbf{e}_y^T \end{bmatrix}_i - \begin{bmatrix} \mathbf{e}_x^T \\ \mathbf{e}_y^T \end{bmatrix}_k \right) \mathbf{u} \geq \gamma$
Node-To-Surface	 $(\mathbf{e}_i^T - (1-\alpha)\mathbf{e}_j^T - \alpha\mathbf{e}_k) \mathbf{u} \geq \gamma$	 $\mathbf{n}^T \left(\begin{bmatrix} \mathbf{e}_x^T \\ \mathbf{e}_y^T \end{bmatrix}_i - (1-\alpha) \begin{bmatrix} \mathbf{e}_x^T \\ \mathbf{e}_y^T \end{bmatrix}_j - \alpha \begin{bmatrix} \mathbf{e}_x^T \\ \mathbf{e}_y^T \end{bmatrix}_k \right) \mathbf{u} \geq \gamma$

Fig. 1. Different constraint conditions for nodal contact.

relation for the discrete form of the potential energy, (2.3), to obtain

$$\pi(\mathbf{u}, \Lambda) = \bar{\pi}(\mathbf{u}) + \Lambda^t (\mathbf{B}^t \mathbf{u} - \boldsymbol{\gamma}), \quad (2.7)$$

where the last term contains all constraint conditions for the nodes in contact. Λ is a vector of the Lagrange multipliers, $\boldsymbol{\gamma}$ is a vector of the discrete initial gaps and

$$\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_c] \quad (2.8)$$

is a matrix containing the c vectors that represent the contact kinematics for each node. c is the total number of constraint conditions. The variation of π in (2.7) leads to the following set of equations:

$$[\mathbf{D}\pi(\mathbf{u}, \Lambda)]^t \delta \mathbf{u} = [\mathbf{D}\bar{\pi}(\mathbf{u})]^t \delta \mathbf{u} + \Lambda^t \mathbf{B}^t \delta \mathbf{u} = \mathbf{0}, \quad (2.9a)$$

$$[\mathbf{D}\pi(\mathbf{u}, \Lambda)]^t \delta \Lambda = [\mathbf{B}^t \mathbf{u} - \boldsymbol{\gamma}]^t \delta \Lambda = \mathbf{0}. \quad (2.9b)$$

By the fundamental theory of variations, (2.9) can be arranged in the matrix form

$$\begin{bmatrix} \mathbf{K} & \mathbf{B} \\ \mathbf{B}^t & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{u} \\ \Lambda \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \boldsymbol{\gamma} \end{Bmatrix}. \quad (2.10)$$

The coefficient matrix in the above equation is indefinite and nonsingular. Equation (2.10), therefore, has a unique solution. However, for some ordering of the unknowns a submatrix of the coefficient matrix is singular and care must be taken when solving these equations. This particularly poses a difficulty for three-dimensional problems [17].

REMARK 2.1. Adding $-(1/2\kappa)\Lambda^t \Lambda$ to (2.7) regularizes the problem. The effect is replacing the zero diagonal block in (2.10) with the diagonal matrix $-(1/\kappa)\mathbf{I}$. This approach is known as 'Augmented Lagrangian' method. The regularization does not solve the problem of the singularity of the submatrix, but helps in the solution step of (2.10) for an appropriate choice of κ . On the other hand the solution only satisfies the constraint condition, the second equation of (2.10), approximately. Instead, \mathbf{u} satisfies the perturbed condition

$$\mathbf{B}^t \mathbf{u} - \boldsymbol{\gamma} - \frac{1}{\kappa} \Lambda = \mathbf{0}. \quad (2.11)$$

Equation (2.10) must be solved as part of a Newton iteration. The size of the coefficient matrix varies with the number of active contact conditions. This results in an additional difficulty that can be avoided using special contact elements [7] and including the constraint condition for all the possible contact nodes in (2.10). The result is a larger system of equations than necessary.

The advantage of the Lagrange multiplier method is that the contact conditions are satisfied exactly.

2.3. Penalty formulation

An alternative procedure for the solution of the contact problem can be derived by satisfying the contact conditions only approximately. Adding a fictitious energy term, also called penalty term, to (2.3) we get

$$\pi(\mathbf{u}) = \bar{\pi}(\mathbf{u}) + \kappa[(\mathbf{B}'\mathbf{u} - \boldsymbol{\gamma})'(\mathbf{B}'\mathbf{u} - \boldsymbol{\gamma})], \quad (2.12)$$

where κ is the penalty parameter. For a more detailed derivation see [8]. Physically, κ represents the stiffness of a fictitious linear spring between any two points that are in contact. The variation of π leads to

$$[D\pi(\mathbf{u})]'\delta\mathbf{u} = [D\bar{\pi}(\mathbf{u})]'\delta\mathbf{u} + \kappa[\mathbf{u}'\mathbf{B}\mathbf{B}' - \boldsymbol{\gamma}'\mathbf{B}']\delta\mathbf{u} = 0. \quad (2.13)$$

Again, by the fundamental theory of variation, (2.13) becomes

$$[\mathbf{K} + \kappa\mathbf{B}\mathbf{B}']\mathbf{u} = \mathbf{f} + \kappa\mathbf{B}\boldsymbol{\gamma}. \quad (2.14)$$

Here the coefficient matrix is symmetric positive-definite. Its size remains unchanged during the Newton iteration. What changes is the profile of the matrix according to the number of contact conditions.

REMARK 2.2. When the penalty parameter is chosen to be too large it leads to numerical problems in the form of loss of accuracy in the solution of (2.14). On the other hand too small a choice for κ results in unacceptable penetration of one body into the other. These effects also arise in other applications of the penalty method as shown in [3, 4]. An analysis presented in [16] gives an estimate for the optimal penalty parameter in contact problems. This estimate depends on the computer precision, ε (ε is the smallest number in the computer that satisfies $1 + \varepsilon > 1$), the total number of unknowns, n , and the smaller stiffness, k , of the two elements that are in contact. The estimate of the penalty parameter is

$$\kappa = k_1/\sqrt{n\varepsilon}. \quad (2.15)$$

REMARK 2.3. The above formulation can also be derived from the augmented Lagrangian method. The elimination of the Lagrange parameters will lead directly to (2.14), where the penalty parameter depends on the perturbation in (2.11).

2.4. Use of static condensation

In most applications the contact region is small compared to the domain of the structure. This fact can be used to reduce the computational efforts for linear structures. In this case the equilibrium equation (2.10) can be arranged in the form

$$\begin{bmatrix} \mathbf{K}_I & \mathbf{K}_{IC} & \mathbf{0} \\ \mathbf{K}_{IC} & \mathbf{K}_C & \mathbf{B} \\ \mathbf{0} & \mathbf{B}' & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{u}_I \\ \mathbf{u}_C \\ \boldsymbol{\lambda} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f}_I \\ \mathbf{f}_C \\ \boldsymbol{\gamma} \end{Bmatrix}, \quad (2.16)$$

where C denotes all possible contact nodes and I are the remaining nodes. Since during the contact iteration \mathbf{u}_I depends only indirectly on \mathbf{A} it can be eliminated from (2.16). This leads to a full but much smaller system of equations

$$\begin{bmatrix} \hat{\mathbf{K}}_C & \mathbf{B} \\ \mathbf{B}^t & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{u}_C \\ \mathbf{A} \end{Bmatrix} = \begin{Bmatrix} \hat{\mathbf{f}}_C \\ \boldsymbol{\gamma} \end{Bmatrix}, \quad (2.17)$$

where

$$\hat{\mathbf{K}}_C = \mathbf{K}_C - \mathbf{K}_{IC}^t \mathbf{K}_I^{-1} \mathbf{K}_{IC}, \quad \hat{\mathbf{f}}_C = \mathbf{f}_C - \mathbf{K}_{IC}^t \mathbf{K}_I^{-1} \mathbf{f}_I.$$

$\hat{\mathbf{K}}_C$ is obtained at an intermediate step of the Gaussian elimination process and therefore its evaluation requires no additional computational effort. The advantage of static condensation is that a smaller system of equation has to be solved during the contact iteration. The static condensation process shown for the Lagrangian multiplier approach can also be used for the penalty method.

3. Partitioning method

The solution of the system of equations (2.10) can be obtained in a number of different ways. Direct methods are the most common procedures used [7], however, there are two difficulties with this scheme. As pointed out, the coefficient matrix of (2.10) is indefinite and may require special care during the factorization step. Since the contact problem is nonlinear we must perform a Newton–Raphson iteration to obtain the solution. Therefore, a factorization must be performed at each Newton step. In this section we derive a hybrid scheme that takes advantage of the fact that the part of the coefficient matrix associated with the degrees of freedom that do not come in contact remains unchanged during the iteration process. This is only true for elastic structures with small deformation. In the sequel, we state the corresponding procedure for nonlinear structural behavior. In the following section, \mathbf{K} represents either the total stiffness matrix or the condensed stiffness matrix given by (2.17).

Eliminating the displacements, \mathbf{u} , in (2.10) we obtain

$$\mathbf{B}^t \mathbf{K}^{-1} \mathbf{B} \mathbf{A} = \mathbf{B}^t \mathbf{K}^{-1} \mathbf{f} - \boldsymbol{\gamma}. \quad (3.1)$$

Then, the displacements are related to the contact forces, \mathbf{A} , through

$$\mathbf{u} = \mathbf{K}^{-1}(\mathbf{f} - \mathbf{B} \mathbf{A}). \quad (3.2)$$

The matrix $\mathbf{B}^t \mathbf{K}^{-1} \mathbf{B}$ appearing in (3.1) is symmetric, positive-definite and full. The size of this matrix depends on the number of nodes that are actually in contact which, in general, is much smaller than the total number of degrees of freedom.

The system of equations (3.1) may be solved using a direct procedure [15]. However, this approach requires the evaluation of the coefficient matrix in (3.1) which is not known explicitly. Therefore, an iterative method will be used that does not require computing the

Table 1

The conjugate gradient algorithm for evaluation of contact forces

Given an initial approximation Λ_0 then:

- (1) compute
 - (a) $p_0 = r_0 = B^T K^{-1} [f - B\Lambda_0] - \gamma$,
 - (b) $\rho_0 = r_0^T r_0$;
- (2) for $k = 0, 1, 2, \dots$ until convergence repeat:
 - (a) $d_k = B^T K^{-1} B p_k$,
 - (b) $\alpha_k = \rho_k / (d_k^T p_k)$,
 - (c) $\Lambda_{k+1} = \Lambda_k + \alpha_k p_k$,
 - (d) $r_{k+1} = r_k - \alpha_k d_k$,
 - (e) $\rho_{k+1} = r_{k+1}^T r_{k+1}$,
 - (f) if $\rho_{k+1} < \text{tol} \cdot \rho_0$ then terminate the loop,
 - (g) $\beta_k = \rho_{k+1} / \rho_k$,
 - (h) $p_{k+1} = r_{k+1} + \beta_k p_k$.

elements of this coefficient matrix. The conjugate gradient method (CG here after) can solve the linear system of equations (3.1) performing only the product $B^T K^{-1} B v$ for a given vector v . For a detailed description of the CG algorithm see [9]. The application of this method to (3.1) is given in Table 1. In this algorithm Λ_k is the k th approximation to the solution of (3.1); r_k is the residual associated with Λ_k and ρ_k is the square of the norm of this residual; p_k is the step direction and α_k is the step length that minimizes the potential energy associated with (3.1). d_k is the change in the residual vector due to the step. The CG iteration is terminated when the norm of the residual vector is reduced by a factor specified by tol . All these vectors are of length equal to the number of nodes in contact.

Theoretical results [9] show that the number of CG iterations is at most equal to the number of equations. However, in practice the number of iterations is greatly influenced by the conditioning of the coefficient matrix. The eigenvalues of the matrix $B^T K^{-1} B$ are the Ritz approximation to the eigenvalues of K^{-1} . Therefore, the spread of the eigenvalues of $B^T K^{-1} B$ is less than that of K^{-1} and hence better conditioned. It is this reason that makes the CG method effective when used to solve (3.1).

We compare the above method to a procedure which evaluates the coefficient matrix and then solves (3.1) using a direct scheme. We base this cost comparison on the operation counts. The number of operations for each method is given in Table 2. Here, c is the number of contact nodes; b and n are the half-bandwidth and the dimension of K respectively; and j is

Table 2

Operation count for each solution scheme

Method	Number of operations
Direct	$c(2bn + \frac{1}{3}c^2 + c)$
Conjugate gradient	$j(2bn + 5c)$

the number of CG iterations. When $j < c$, the CG method is faster than the direct method. In the case of small c ($c \approx 12$) and $j = c$ the direct method is slightly faster. Another advantage of the CG method is that it requires considerably less storage than the direct method, since the coefficient matrix is never computed explicitly. Furthermore, when a low level of accuracy is required j may be much less than c .

4. Contact algorithms

Our objective is to derive algorithms for the solution of contact problems which may also include material and/or geometrical nonlinearities in the structure. At present most algorithms are based on some form of Newton iteration which treats the nonlinearities due to contact conditions in the same way as those present in the structure itself. Here we develop algorithms that distinguish between these nonlinearities and therefore the contact equations are handled separately by a minor iteration.

4.1. Minor iteration

The inherent nonlinearities in contact problems are due to the fact that the correct contact area is not known *a priori*. To determine the contact area and the contact forces a minor Newton iteration is used. The linear system of equations arising at each step of the minor iteration is solved using the partitioning scheme given in Section 3. This minor iteration is described in Table 3.

The matrix $\mathbf{B}^{(i)}$ contains vectors that describe those contact conditions to be enforced. In general these vectors depend on the direction of the normal from a contact node to the contact surface. This normal is obtained by an orthogonal projection from a node on the slave surface onto the master surface [5]. In the large deformation case the change in the normal vectors must be considered. Since the nonlinear contact problem is solved as a sequence of linear

Table 3
Minor iteration

-
- (1) Given the displacements, $\bar{\mathbf{u}}$, the contact forces, $\bar{\mathbf{A}}$, a contact state, $\bar{\mathbf{B}}$, the residual force vector, \mathbf{r} , and the factors of the tangent matrix, \mathbf{K}_T from a previous major Newton step:

$$\text{set } \mathbf{u}^{(0)} = \bar{\mathbf{u}} \text{ and } \mathbf{B}^{(0)} = \bar{\mathbf{B}};$$

- (2) for $i = 0, 1, 2, \dots$ until convergence repeat;
 (a) determine all penetrating nodes through (2.4) and establish $\mathbf{B}^{(i)}$,
 (b) $\mathbf{g}^{(i)} = -\boldsymbol{\gamma} + \mathbf{B}^{(i)} \mathbf{u}^{(i)}$,
 (c) if $\|\mathbf{g}^{(i)}\| < \text{tol} \cdot \|\mathbf{g}^{(0)}\|$ then terminate,
 (d) solve using CG method $\mathbf{B}^{(i)T} \mathbf{K}_T^{-1} \mathbf{B}^{(i)} \boldsymbol{\Lambda}^{(i+1)} = \mathbf{g}^{(i)}$,
 (e) if any element of $\boldsymbol{\Lambda}^{(i+1)}$ is positive reset to zero,
 (f) $\Delta \mathbf{u}^{(i)} = -\mathbf{K}_T^{-1} \mathbf{B}^{(i)} \boldsymbol{\Lambda}^{(i+1)}$,
 (g) $\mathbf{u}^{(i+1)} = \mathbf{u}^{(0)} + \Delta \mathbf{u}^{(i)}$;

- (3) return $\mathbf{u}^{(i)}$, $\boldsymbol{\Lambda}^{(i)}$, and $\mathbf{B}^{(i)}$ in $\bar{\mathbf{u}}$, $\bar{\mathbf{A}}$, and $\bar{\mathbf{B}}$ respectively.
-

contact problems the structural nonlinearities can be ignored during the minor iteration. For this reason we can neglect effects such as the changes in the normal to the contact surface.

REMARK 4.1. The coefficient matrix in the system of equations, $\mathbf{B}^{(i)\top} \mathbf{K}_T^{-1} \mathbf{B}^{(i)} \mathbf{A}^{(i+1)} = \mathbf{g}^{(i)}$, at Step (2d) of the minor iteration may be replaced by an approximating matrix, $\mathbf{H}^{(i)}$. The object is to reduce the numerical effort in the solution of this equation. The price for this approximation is a larger number of iterations. Some choices for $\mathbf{H}^{(i)}$ can be found in [10]. These include:

- $\mathbf{H}^{(i)} = (1/\kappa^{(i)})\mathbf{I}$. Here, $\kappa^{(i)}$ is a penalty parameter which is increased during the iteration starting from a small value. This procedure is usually referred to as the augmented Lagrangian method.
- $\mathbf{H}^{(i)} = \mathbf{B}^{(i)\top} \mathbf{B}^{(i)}$. In the case of contact problems this is a diagonal matrix with nonzero terms ranging from 1 to 2 depending on the modeling of the contact condition. Due to the structure of $\mathbf{H}^{(i)}$ this leads to a poor rate of convergence.

With the above approximations the solution of the system of equations in Step (2d) can be obtained at little cost. However, the increase in the number of iterations will escalate the overall cost due to Step (2f).

4.2. Major iteration

When nonlinearities due to material and/or geometry are also present, all nonlinear effects must be considered. The Newton algorithm has been applied successfully to obtain the solution of such problems. Here we modify the Newton process by incorporating the minor iteration described above (see Table 4).

REMARK 4.2. As mentioned before the Lagrange parameter approach cannot avoid the difficulty that exists when one or more rigid-body modes are present in the finite element model. To sidestep this difficulty there are three possible procedures that one can follow:

- (i) Introduce as many springs as there are rigid-body modes into the finite element model.
- (ii) Solve the problem using a specified displacement instead of an applied force.

Table 4
Major iteration

-
- (1) Start from an initial approximation $\bar{\mathbf{u}}^{(0)}$, $\bar{\mathbf{A}}^{(0)} = \mathbf{0}$, and $\bar{\mathbf{B}}^{(0)} = \mathbf{0}$;
 - (2) for $m = 0, 1, 2, \dots$ until convergence do;
 - (a) compute the residual force, $\mathbf{r}^{(m)}$, due to $\bar{\mathbf{u}}^{(m)}$,
 - (b) if $\|\mathbf{r}^{(m)} + \bar{\mathbf{B}}^{(m)} \bar{\mathbf{A}}^{(m)}\| < \text{tol} \cdot \|\mathbf{r}^{(0)}\|$ then stop,
 - (c) compute the tangent matrix $\mathbf{K}_T^{(m)}$ at $\bar{\mathbf{u}}^{(m)}$,
 - (d) factorize $\mathbf{K}_T^{(m)}$ into \mathbf{LDL}^\top ,
 - (e) solve $\mathbf{K}_T^{(m)} \Delta \bar{\mathbf{m}}^{(m)} = -\mathbf{r}^{(m)} - \bar{\mathbf{B}}^{(m)} \bar{\mathbf{A}}^{(m)}$,
 - (f) update displacements $\bar{\mathbf{u}}^{(m+1)} = \bar{\mathbf{u}}^{(m)} + \Delta \bar{\mathbf{u}}^{(m)}$,
 - (g) perform the minor iteration.
-

(iii) During the LDL^T factorization of K_T one can replace the diagonal elements in D that are zero within the machine precision with terms that are of the order of the required accuracy. This is equivalent to removing the rigid-body modes.

An advantage of major–minor iteration is that it will never require more triangular factorizations than the standard Newton procedure. This is due to the fact that if only one minor iteration is performed, the algorithm reduces to the standard Newton scheme. As the numerical examples show, the minor iteration reduces the number of triangular factorizations and therefore results in a more efficient algorithm.

A further advantage is that the Steps (2c) and (2d) of the major iteration can be replaced with an updating scheme for the triangular factors of the tangent matrix such as the Broyden and BFGS methods [1, 11]. This will result in a further reduction in the number of triangular factorizations and may reduce the overall cost of the analysis. In general, updating procedures that retain the inertia¹ of the coefficient matrix such as those cited above cannot be applied to the usual formulation of contact problems by the Lagrange multiplier method, (2.10). The partitioning method permits the use of these updating schemes.

5. Numerical results

In this section we demonstrate the characteristics of the solution procedure described in Section 3 on several examples. The main objective is to compare the cost of different algorithms for the solution of contact problems. All numerical experiments were performed on a VAX 11/780 using the finite element program, FEAP [16], as a basis for all the computations.

EXAMPLE 5.1 (Parabolic beam). In this example a parabolic beam, described in Fig. 2, is modeled using four-node, plane-stress, quadrilateral elements. The finite element discretization results in a total of 860 equations. Four different procedures were used to solve this problem. These are:

– Lagrange multiplier method:

- (i) major–minor iteration or partitioning method based on (3.1), (3.2);
- (ii) full Newton method [7] based on (2.10);
- (iii) Newton–Lanczos method [12] based on (2.10);

– penalty method:

- (iv) full Newton method [5] based on (2.14).

The results of our comparisons are presented in Table 5. The solution times presented in this table include the evaluation of the stiffness matrices and residual force vectors as well as the solution of the system of equations. The partitioning method requires less computation time than the other methods. The Newton–Lanczos method, as implemented in [12], has a drawback since it requires additional storage for the preconditioning matrix when used for contact problems.

¹The inertia of a matrix is its number of negative, zero and positive eigenvalues.

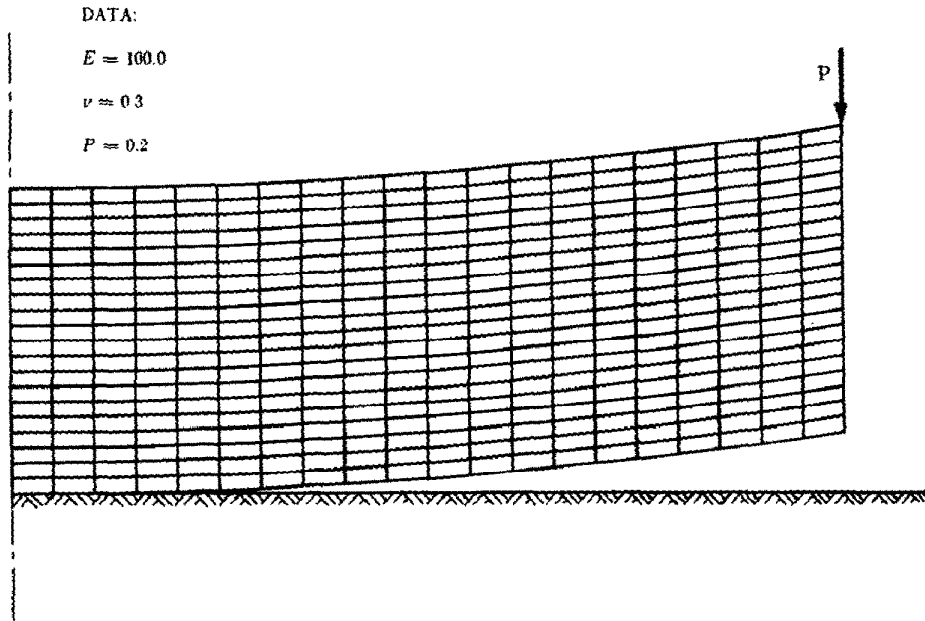


Fig. 2. Geometry and data of the parabolic beam.

The solution time for all the above methods may be reduced by performing static condensation as defined in Section 2.4. In this case the coefficient matrix of the reduced system is full and therefore the partitioning method can best take advantage of this. Note, a large part of the cost for the partitioning method (17.9 sec.) is due to the minor iteration and is consumed by the matrix-vector multiplication $\mathbf{K}^{-1}(\mathbf{B}\mathbf{v})$. When \mathbf{K} is condensed and much smaller this cost will almost be eliminated.

EXAMPLE 5.2 (Circular beam). As a second example we consider a circular beam in contact with a rigid foundation. The purpose of this is to demonstrate the influence of the penalty parameter on the total number of iterations. Due to the geometry of the problem (see Fig. 3) the middle of the beam will lift up. Using the symmetry condition, the beam is discretized by

Table 5
Cost comparisons of various methods for Example 5.1

	Lagrange multiplier			Penalty
	Part. method	Full Newton	Newton-Lanczos	Full Newton
Total solution time (sec.)	48.8	112.9	84.7	114.1
Number of iterations	3 minor 1 major	3	3	3
Number of factorizations	1	3	2	3

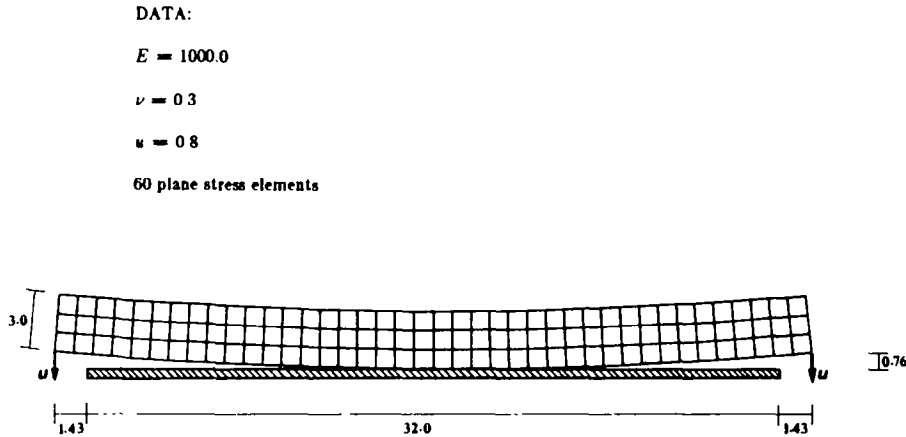


Fig. 3. Geometry and data of the circular beam problem.

60 plane-stress elements. Imposed displacements were specified at the two ends of the beam. The deformed shape of the beam is presented in Fig. 4.

This problem is solved with a range of penalty parameters and the results are presented in Table 6.

The optimum penalty parameter obtained from (2.15) is $\kappa_{\text{opt}} = 10^8$. Here optimality refers to accuracy in the constraint condition. The number of iterations for penalty parameters larger than κ_{opt} is the same as the number of iterations for κ_{opt} . Thus, there is little to be gained by overestimating the penalty parameter. For many applications lower accuracy may be sufficient. Then, underestimating the penalty parameter may translate into fewer iterations as can be seen in Table 6.

EXAMPLE 5.3 (Contact of two beams). The interaction of two cantilever beams due to contact is used to demonstrate the behavior of the methods for geometrically nonlinear problems. A detailed description of the problem is given in Fig. 5. Each beam is modeled by 10 finite deformation beam elements [13]. A prescribed vertical displacement of 25.0 units is applied at point A on the upper beam. The deformed shape (see Fig. 6) is obtained using a single load step. The partitioning method achieved convergence after 5 major and 6 minor iterations with a total solution time of 6.5 sec. In contrast, the penalty method required 9 iterations to converge. The total solution time for this method was 9.2 sec.

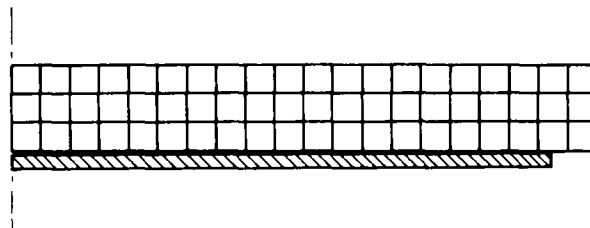


Fig. 4. Deformed configuration of the circular beam.

Table 6
Influence of penalty parameter on convergence

Method	Penalty parameter	Number of iterations	Maximum penetration
Partitioning		1 major 6 minor	0.0
Penalty	10	4	5.27×10^{-2}
	10^2	5	2.73×10^{-2}
	10^3	6	7.09×10^{-3}
	10^4	8	9.24×10^{-4}
	10^6	8	9.28×10^{-6}
	10^8	8	9.29×10^{-8}

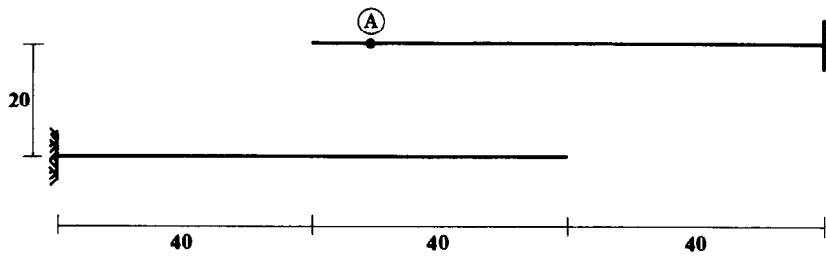


Fig. 5. Geometry and data of large-deformation beam problem.

DATA:

$$EI = 10^4$$

$$EA = 10^6$$

$$GA = 10^6$$

$$u_A = 25.0$$

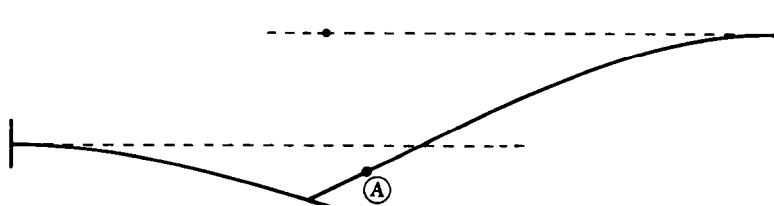


Fig. 6. Deformed configuration of the beam problem.

The observed reduction in the number of major iterations (from 9 to 5 iterations) is due to the minor iteration. At the beginning of the first iteration three nodes of the upper beam penetrate the lower beam. Then, the penalty method enforces the associated three constraint conditions. This is a poor approximation for the contact region and results in slow convergence. On the other hand, the minor iteration of the partitioning method releases two of the nodes in contact and therefore obtains a better approximation to the contact area and hence faster convergence.

References

- [1] C. Broyden, A new method of solving nonlinear simultaneous equations, *Comput. J.* 12 (1969) 94–99.
- [2] S.K. Chan and I.S. Tuba, A finite element method for contact problems of solid bodies: I. Theory and validation, *Internat. J. Mech. Sci.* 13 (1971) 627–639.
- [3] C.A. Felippa, Error analysis of penalty function techniques for constraint definition in linear algebraic systems, *Internat. J. Numer. Meths. Engrg.* 11 (1977) 709–728.
- [4] C.A. Felippa, Iterative procedures for improving penalty function solutions of algebraic systems, *Internat. J. Numer. Meths. Engrg.* 12 (1978) 821–836.
- [5] J.O. Hallquist, NIKE2D: An implicit, finite-deformation, finite-element code for analysing the static and dynamic response of two-dimensional solids, Rept. UCRL-52678, University of California, Lawrence Livermore National Laboratory, 1979.
- [6] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standards* 49 (1952) 409–436.
- [7] T.R.J. Hughes, R.L. Taylor, J.L. Sackman, A. Curnier and W. Kanoknukulchai, A finite element method for a class of contact–impact problems, *Comput. Meths. Appl. Mech. Engrg.* 8 (1976) 249–276.
- [8] N. Kikuchi and J.T. Oden, Contact problems in elastostatics, in: J.T. Oden and G.F. Carey, eds., *Finite Elements: Special Problems in Solid Mechanics*, Vol. IV (Prentice-Hall, Englewood Cliffs, NJ, 1984).
- [9] M.R. Li, B. Nour-Omid and B.N. Parlett, A fast solver free of fill-in for finite element problems, *SIAM J. Numer. Anal.* 19 (6) (1982) 1233–1242.
- [10] D.G. Luenberger, *Linear and Nonlinear Programming* (Addison-Wesley, Reading, MA, 2nd ed., 1984).
- [11] H. Matthies and G. Strang, The solution of nonlinear finite element equations, *Internat. J. Numer. Meths. Engrg.* 14 (1979) 1613–1626.
- [12] B. Nour-Omid, B.N. Parlett and R.L. Taylor, A Newton–Lanczos method for solution of nonlinear finite element equations, *Comput. & Structures* 16 (1–4) (1982) 241–252.
- [13] J.C. Simo, K.D. Hjelmstad and R.L. Taylor, Finite element formulations for problems of finite deformation of elasto-viscoplastic beams. Rept. No. UCB/SESM-83/01, Department of Civil Engineering, University of California, Berkeley, CA, 1983.
- [14] J.H. Wilkinson, *Rounding Errors in Algebraic Problems* (Prentice-Hall, Englewood Cliffs, NJ, 1963).
- [15] P. Wriggers, Zur Berechnung von Stoss- und Kontaktproblemen mit Hilfe der Finite-Element Methode, Bericht Nr. F81/1, Forschungs- und Seminarberichte aus dem Bereich der Mechanik der Universität, Hannover, 1981.
- [16] P. Wriggers and B. Nour-Omid, Solution methods for contact problems, Rept. No. UCB/SESM 84-09, University of California, Berkeley, CA, 1984.
- [17] O.C. Zienkiewicz, *The Finite Element Method* (McGraw-Hill, London, 3rd ed., 1977).