REPORT NO.
UCB/SESM-85/09

STRUCTURAL ENGINEERING AND
STRUCTURAL MECHANICS

AN AUGMENTED LAGRANGIAN FORMULATION FOR THE FINITE ELEMENT SOLUTION OF CONTACT PROBLEMS

by

JOSEPH A. LANDERS

and

ROBERT L. TAYLOR

DECEMBER 1985

DEPARTMENT OF CIVIL ENGINEERING
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

An Augmented Lagrangian Formulation for the Finite Element Solution of Contact Problems

Joseph A. Landers† and Robert L. Taylor‡

Department of Civil Engineering,
University of California, Berkeley.

ABSTRACT

A solution method for small deformation frictionless contact problems is developed. First, a review of various solution methods for contact problems is given. Next, a detailed discussion is made of the augmented Lagrangian method. Then, the finite element implementation is discussed. Finally, sample problems and their solutions are presented to demonstrate the usefulness of the method. Both the static and dynamic solution algorithms are described in this report.

1. Introduction

While solutions to contact problems have been available for years [1], techniques applicable to a wide range of problems with a minimum amount of user intervention have been quite limited. This is true not only of static contact problems, but also in dynamic impact situations as well.

In this report, a study is first made of various frictionless, small deformation contact algorithms. The advantages and disadvantages of each are discussed. Next, the augmented Lagrangian solution method is presented in the context of finite element contact problems. This discussion covers both the static and dynamic algorithms. The generality of the augmented Lagrangian method is shown by examining how the other methods may be derived from it. Several examples are presented, and comparisons are made to work done by previous researchers. Finally, recommendations are made for future studies in contact problems.

---

† Research Assistant
‡ Professor and Chairman

This work was funded by a grant from the Naval Civil Engineering Laboratory at Port Hueneme, California. The authors gratefully acknowledge their support.

2. Contact Solution Algorithms

Contact solution algorithms are generally based upon either the classical Lagrange procedure or the penalty function method. In the context of linearized elasticity, finite element solutions can be viewed as the minimization of the potential energy functional presented in equation (1).

$$\Pi(\underline{\mathbf{u}}) = \int_{B_1 \cup B_2} W(\underline{\mathbf{u}})\, dv - \int_{B_1 \cup B_2} \rho \mathbf{b} \cdot \underline{\mathbf{u}}\, dv - \int_{dB_1 \cup dB_2} \bar{\mathbf{t}} \cdot \underline{\mathbf{u}}\, da \tag{1}$$

Where W is the stored energy function, $\rho\mathbf{b}$ is the body force vector, $\bar{\mathbf{t}}$ is the surface traction specified on the boundary and $\underline{\mathbf{u}}$ is the displacement field. The integrals are taken over both bodies and their boundaries. Application of standard finite element procedures leads to equation (2), the discrete form of the functional.

$$\Pi(\mathbf{u}) = \tfrac{1}{2}\mathbf{u}^t \mathbf{K}\mathbf{u} - \mathbf{u}^t \mathbf{R} \tag{2}$$

Where $\mathbf{K}$ is the system stiffness matrix, $\mathbf{f}$ is the vector of nodal forces, and $\mathbf{u}$ is the displacement vector of nodes in the mesh. In the case of finite element contact problems, a further requirement states that penetration of the two bodies does not take place.

$$(\hat{\mathbf{u}}_{B_1} \cup \hat{\mathbf{u}}_{B_2}) \cdot \mathbf{n} = 0 \tag{3}$$

Where $\hat{\mathbf{u}}_{B_1}$ and $\hat{\mathbf{u}}_{B_2}$ are the positions in either deformed body, taken over the contact interface, and $\mathbf{n}$ is the normal between the bodies.

Application of equation (2) subject to the constraint of equation (3) is equivalent to an optimization problem, which may be stated as:

$$\textit{minimize}: \Pi(\mathbf{u}) \tag{4.a}$$

$$\textit{with the constraint}: \mathbf{g}^t \mathbf{u} = 0 \tag{4.b}$$

The purpose of this report is to discuss solution schemes for the expressions listed in equations (4).

2.1. The Penalty Method

Optimization problems that involve constraints, such as those presented in equation (4.b), are frequently converted to unconstrained problems by using these constraint equations to establish relationships between the unknowns [2]. By writing equation (2) with the addition of a fictitious energy term, a new functional is defined.

$$\Pi'(\mathbf{u}) = \Pi(\mathbf{u}) + \frac{\mu}{2}\mathbf{u}^t \mathbf{g}\mathbf{g}^t \mathbf{u}\,, \quad \textit{where}: \mu > 0 \tag{5}$$

The last term contains the constraint, $\mathbf{g}^t\mathbf{u}$, times a penalty term, μ. Because

solution to finite element problems involve the minimization of equation (5), the answer will tend to have a small value for $\mathbf{g}^t\mathbf{u}$. Taking the first variation of (5) equation (6) is found.

$$\delta\mathbf{u}^t\mathbf{K}\mathbf{u} - \delta\mathbf{u}^t\mathbf{R} + \delta\mathbf{u}^t\mu\mathbf{g}\mathbf{g}^t\mathbf{u} = \mathbf{0} \tag{6}$$

By the fundamental lemma of variational equations, this becomes:

$$[\mathbf{K} + \mu\mathbf{g}\mathbf{g}^t]\mathbf{u} = \mathbf{R} \tag{7}$$

Several comments are applicable to the penalty method. First, the correct choice of the penalty parameter, μ, is the essence of the algorithm. Since the penalty approach satisfies the contact conditions only approximately, valid results from equation (4) are highly dependent on the value of μ.

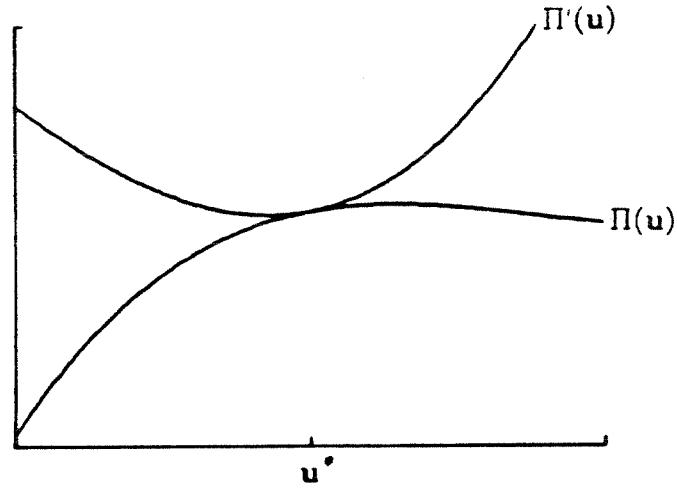

Figure 1: The effect of the penalty term, one dimensional case.

In Figure 1, a plot of the two functionals is shown. The addition of the penalty parameter to equation (2), as stated in equation (5), can be seen as rendering the functional convex [3]. This is true even if $\Pi(\mathbf{u})$ is not already convex, for points near the solution. Also note that in Figure 2, the minimum of the functional with the penalty term will approach the minimum of the functional without the penalty term only as $\mu\to\infty$.

Issues such as $\mu\to\infty$ bring forward a major problem with the penalty method: the matrices used in the solution of the finite element problems may exhibit poor numerical conditioning. While large penalty values insure a tighter tolerance for the constraint, these values make an accurate solution of the linear system of equations much more difficult.

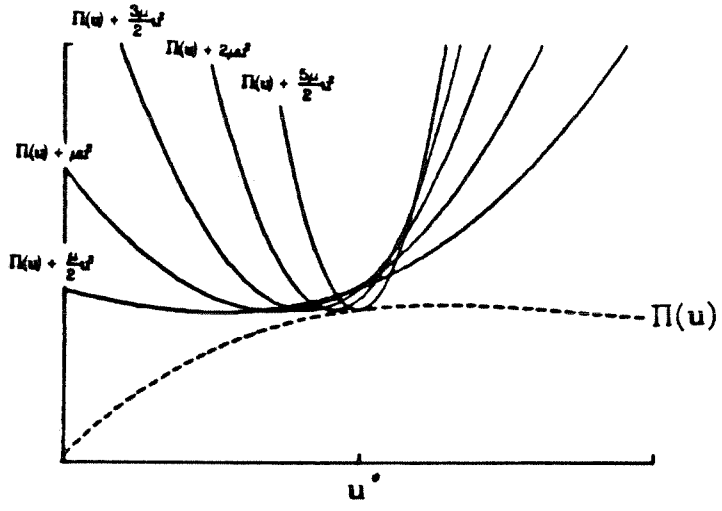

Figure 2: The effect of increasing the penalty parameter.

The following problem will serve as a basis of comparison for the various contact solution algorithms. This system consists of two springs connected in series. One end is fixed, while at the other, a concentrated force is applied. The constraint requires that the displacements of nodes 1 and 2 remain equal. Figure 3 illustrates the system and external loading. By applying the penalty method, the matrices corresponding to equation (7) are:

$$\mathbf{K} = \begin{bmatrix} 2k & -k \\ -k & k \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 0 \\ q \end{bmatrix} \quad \mathbf{g} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{8}$$

This leads to a linear system of equations:

$$\begin{bmatrix} 2k + \mu & -(k + \mu) \\ -(k + \mu) & k + \mu \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ q \end{bmatrix} \tag{9}$$

This system has the solution:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \dfrac{q}{k} \\ \dfrac{q}{k}\left(\dfrac{2k + \mu}{k + \mu}\right) \end{bmatrix} \tag{10}$$

The exact solution is shown in equation (11).

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \dfrac{q}{k} \\ \dfrac{q}{k} \end{bmatrix} \tag{11}$$

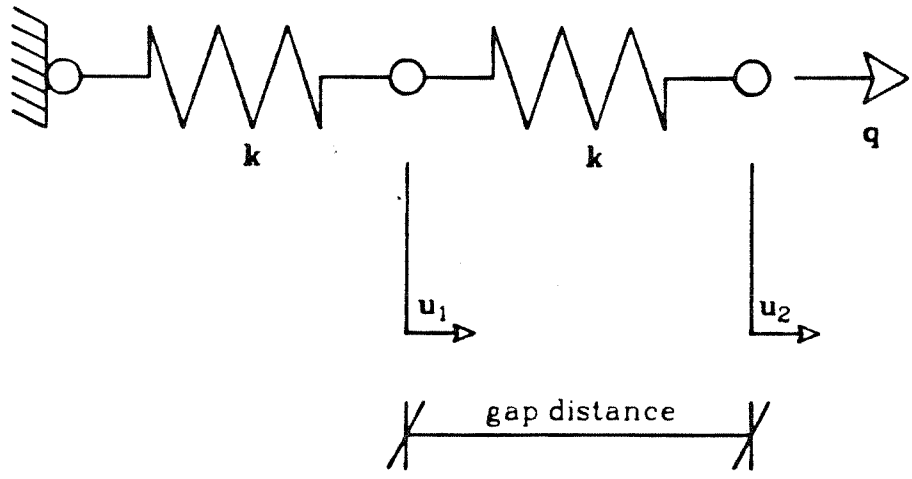

Figure 3: A sample finite element system.

Note that the condition number of the stiffness matrix can easily be found to be:

$$cond(\mathbf{K}) = \frac{\lambda_{max}}{\lambda_{min}} = \frac{3k + 2\mu + \sqrt{5k^2 + 8k\mu + 4\mu^2}}{3k + 2\mu - \sqrt{5k^2 + 8k\mu + 4\mu^2}} \tag{12}$$

As Figures 4 and 5 illustrate, increasing penalty values lead to more accurate solutions, but they also lead to more poorly conditioned matrices. Because of the extreme importance of μ, a natural question to ask is: how can an optimal value for the penalty parameter μ be found?

In order to find an answer to this question, the errors involved in solution of the finite element contact problem must be examined in some detail. Two sources of error effect the accuracy of the penalty method: the large perturbation that results from a small penalty parameter, and the rounding errors due to large penalty values in finite precision arithmetic on the computer [4]. A study of the perturbation errors is done first.

Applying the Sherman-Morrison formula to equation (7), the following expression for the displacements by the penalty method is obtained.

$$\mathbf{u}_p = \left[\mathbf{K}^{-1} - \frac{\mu \mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{\mathbf{I} + \mu \mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}}\right]\mathbf{R} \tag{13}$$

The exact solution can be found by letting $\mu \to \infty$:

$$\mathbf{u}_e = \left[\mathbf{K}^{-1} - \frac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}}\right]\mathbf{R} \tag{14}$$

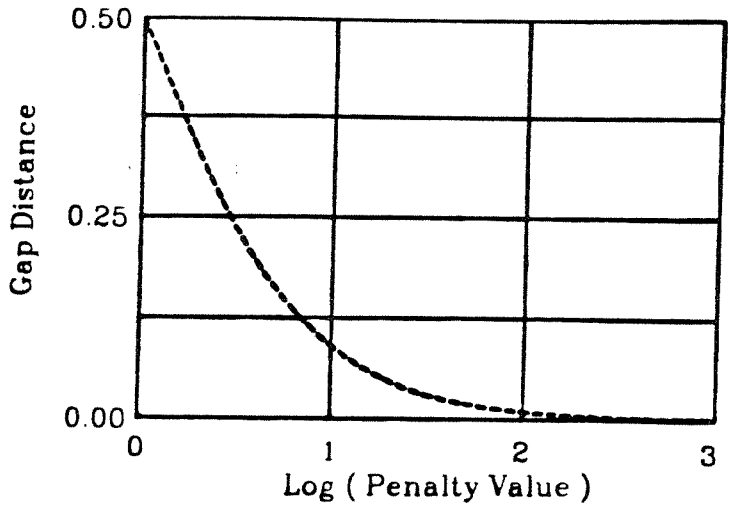

Figure 4: The effect of μ on the gap distance.

The error in **u** is:

$$\mathbf{u}_p - \mathbf{u}_e = \left[\frac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}} - \frac{\mu\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{1+\mu\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}}\right]\mathbf{R} \tag{15}$$

Upon expanding the last term as a series, this becomes:

$$\mathbf{u}_p - \mathbf{u}_e = \left[\frac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}} - \frac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}} + \frac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{\mu(\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g})^2} - \frac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{\mu^2(\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g})^3} + \cdots\right]\mathbf{R} \tag{16}$$

Canceling the first two terms, dropping items in μ with powers greater than one, and taking norms of both sides (16) can be written as:

$$\left\|\mathbf{u}_p - \mathbf{u}_e\right\| \approx \frac{1}{\mu}\left\|\frac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}\mathbf{R}}{\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}}\right\| \tag{17}$$

Dividing both sides by $\left\|\mathbf{u}_e\right\|$, (17) is transformed to:

$$\frac{\left\|\mathbf{u}_p - \mathbf{u}_e\right\|}{\left\|\mathbf{u}_e\right\|} \approx \frac{1}{\mu}\frac{\left\|\dfrac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}\mathbf{R}}{\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}}\right\|}{\left\|\left[\dfrac{\mathbf{K}^{-1}\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g} - \mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}}{\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}}\right]\mathbf{R}\right\|} \tag{18}$$

Simplifying, equation (19) is found:

$$\frac{\left\|\mathbf{u}_p - \mathbf{u}_e\right\|}{\left\|\mathbf{u}_e\right\|} \le \frac{1}{\mu}\left\|\frac{\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}\mathbf{R}}{\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}}\right\|\frac{1}{\left\|\mathbf{K}^{-1}\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}\,\mathbf{R}\right\| + \left\|\mathbf{K}^{-1}\mathbf{g}\mathbf{g}^t\mathbf{K}^{-1}\mathbf{R}\right\|} \tag{19}$$

For some constant a, this expression can be written as:

$$\frac{\|\mathbf{u}_p - \mathbf{u}_e\|}{\|\mathbf{u}_e\|} \leq \frac{1}{\mu} \frac{a}{\|\mathbf{g}^t \mathbf{K}^{-1} \mathbf{g}\|} \tag{20}$$

In practice, for a reasonably large μ, a has a value near unity.

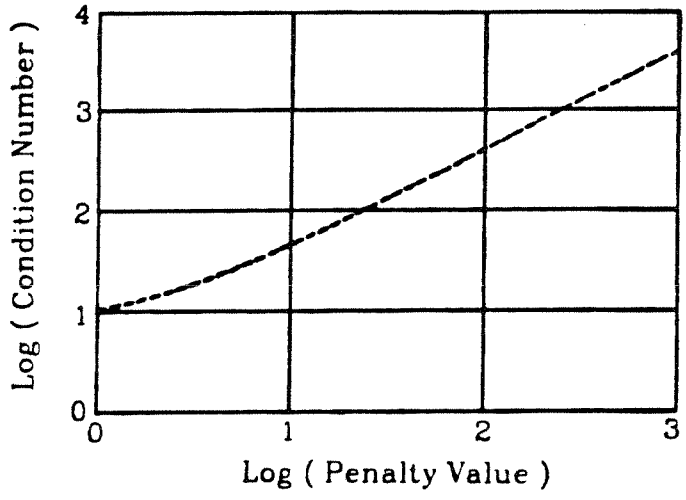

Figure 5: The effect of μ on the condition number.

A standard approximation for rounding errors is given in [5] as an expression of the form:

$$\frac{\|\mathbf{u}_f - \mathbf{u}_p\|}{\|\mathbf{u}_p\|} \leq b\, 2^{-t} cond(\mathbf{K}) \tag{21}$$

where $\mathbf{u}_f$ represents the result computed using finite precision arithmetic, $\mathbf{u}_p$ identifies the result computed if exact arithmetic is used. The symbol b represents some constant, while t represents the number of binary digits of floating point precision. On the DEC VAX* series of computers, using double precision instructions, t has a value of 56. An approximation for the condition number of the system matrix **K** is:

$$cond(\mathbf{K}) \approx \frac{\mu}{K_{min}} \tag{22}$$

Here, K_{min} is smallest diagonal stiffness matrix coefficient that is in the contact

* VAX is a trademark of the Digital Equipment Corporation

region. Hence, the total error is represented by summing both, as in equation (23).

$$\frac{\left\| u_p - u_e \right\|}{\left\| u_e \right\|} + \frac{\left\| u_f - u_p \right\|}{\left\| u_p \right\|} \approx \frac{1}{\mu} \frac{a}{\left\| g^t K^{-1} g \right\|} + \frac{\mu\, b\, 2^{-t}}{K_{min.}} \tag{23}$$

Differentiating this expression and setting it equal to zero, the minimum is:

$$\mu = \left(\frac{a\, K_{min.}}{b\, 2^{-t} \left\| g^t K^{-1} g \right\|} \right)^{1/2} \tag{24}$$

Assuming that the following approximation holds:

$$\left\| g^t K^{-1} g \right\| \approx \frac{1}{K_{min.}} \tag{25}$$

equation (24) then becomes:

$$\mu \approx \left(\frac{c\, K_{min.}^2}{2^{-t}} \right)^{1/2} \tag{26}$$

for some constant c. In practice, c has a value near unity when μ is reasonably large. Equation (26) represents a recommendation for the penalty value, within the limitations presented in its derivation.

To summarize, penalty methods have several advantages. Despite the conditioning and accuracy problems, the resulting solution requires no extra efforts beyond adding the contact components to the stiffness matrix. Furthermore, these methods are easy to implement and have a certain intuitive appeal to engineers. Finally, the contact force conjugate to the contact constraint, although not explicitly found by this method, can be computed using an extra procedure.

2.2. The Lagrange Multiplier Method

The classical Lagrange methods were among the first methods used to solve contact problems. With this method, the functional of equation (2) is amended with an additional term. This defines the new system listed in equation (27).

$$\Pi'(\mathbf{u},\boldsymbol{\lambda}) = \Pi(\mathbf{u}) + \boldsymbol{\lambda}^t \mathbf{g}^t \mathbf{u} \tag{27}$$

Taking the variation of (27) yields the following two equations:

$$\delta \mathbf{u}^t \mathbf{K}\mathbf{u} - \delta \mathbf{u}^t \mathbf{R} + \delta \mathbf{u}^t \mathbf{g} \boldsymbol{\lambda} = 0 \tag{28.a}$$

$$\delta \boldsymbol{\lambda}^t \mathbf{g}^t \mathbf{u} = 0 \tag{28.b}$$

By the fundamental lemma of variational equations, this leads to the system:

$$\begin{bmatrix} \mathbf{K} & \mathbf{g} \\ \mathbf{g}^t & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix} \tag{29}$$

The set of equations expressed by equation (29) may be indefinite but non-singular [2]. Unlike the standard penalty method, the dimensions of the matrices increase with this approach. Additionally, as illustrated by equation (29), the matrix contains a zero sub-matrix along its diagonal. In a practical finite element system, these zero diagonal terms are not automatically available in a sub-matrix, but are scattered throughout the system stiffness matrix. Thus, care must be exercised when the solution is found; special procedures should be used which effectively re-order the system of equations.

Despite these disadvantages, classical Lagrange multiplier methods allow the contact conditions to be satisfied exactly. Also, the λ value has a straightforward physical significance; it is the value of the contact force conjugate to the constraint condition.

An illustration of how the classical Lagrange method is used in contact problems is demonstrated by again considering the two spring problem. Applying equation (29), the linear system of equations can easily be found:

$$\begin{bmatrix} 2k & -k & 1 \\ -k & k & -1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ q \\ 0 \end{bmatrix} \tag{30}$$

This system has the solution:

$$\begin{bmatrix} u_1 \\ u_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} \frac{q}{k} \\ \frac{q}{k} \\ -q \end{bmatrix} \tag{31}$$

The results are the exact values for the displacement and contact force.

2.3. The Perturbed Lagrangian Method

An interesting extension to the classical Lagrange method is offered in references [6] and [7]. Here the functional described by equation (27) is modified with an additional term:

$$\Pi'(\mathbf{u},\boldsymbol{\lambda}) = \Pi(\mathbf{u}) + \boldsymbol{\lambda}^t \mathbf{g}^t \mathbf{u} - \frac{1}{2\mu}\boldsymbol{\lambda}^t\boldsymbol{\lambda}, \quad where: \mu > 0 \tag{32}$$

The effect of the last term is to fill in the zero sub-matrix of equation (29). The resulting stiffness matrix no longer contains the zeros along the diagonal, the classical Lagrange functional is recovered by allowing $\mu \to \infty$. Superficially, it might be observed that this approach still suffers from one problem of the classical Lagrange method; an increase in the number of degrees of freedom of the system. However, this problem can be circumvented in a clever manner when

the method is applied to finite element systems. To see how this may be done, consider the first variation of equation (32).

$$\delta \mathbf{u}^t \mathbf{K}\mathbf{u} - \delta \mathbf{u}^t \mathbf{R} + \int_{\delta_c} \lambda \delta \mathbf{u}^t \mathbf{g}\, d\alpha = 0 \tag{33.a}$$

$$\int_{\delta_c} \delta\lambda \left\{ \frac{-\lambda}{\mu} + \mathbf{g}^t \mathbf{u} \right\} d\alpha = 0 \tag{33.b}$$

Where the integral extends over the area of contact between the two bodies. Since $\delta\lambda$ is arbitrary and μ is a given constant, equation (33.b) can be written as:

$$\frac{1}{\mu} \int_{\delta_c} \lambda \, d\alpha = \int_{\delta_c} \mathbf{g}^t \mathbf{u} \, d\alpha \tag{34}$$

In the framework of finite element analysis, the displacement field is generally approximated over each element region by some polynomial. If this approximation is linear, the strain field is constant function. Thus assuming that the contact force, λ, is constant over the region of interest, the above expression can be transformed to:

$$\lambda = \frac{\mu}{A_c} \int_{\delta_c} \mathbf{g}^t \mathbf{u} \, d\alpha \tag{35}$$

Because of the linear approximation, the integral can be evaluated by means of the trapezoidal rule. Hence, λ, can be approximated to:

$$\lambda = \mu \mathbf{g}^t \mathbf{u} \tag{36}$$

The point of this exercise is that equation (33.a), after applying the fundamental lemma of variational equations, becomes the expression presented in equation (37).

$$\mathbf{K}\mathbf{u} - \mathbf{R} + \mathbf{g}\lambda = 0 \tag{37}$$

Applying the results of (36), a new expression is found.

$$\mathbf{K}\mathbf{u} - \mathbf{R} + \mu \mathbf{g}\mathbf{g}^t \mathbf{u} = 0 \tag{38}$$

Thus, the two equations in (33) can be reduced to a single system that does not increase the number of degrees of freedom. Equation (38) is essentially that of equation (7), except for the following points. First, an explicit expression for the contact force is directly available through the use of equation (36). Second, while a solution is sought, using the Lagrange approach, to the following:

$$\textit{minimize}: \Pi(\mathbf{u}) \tag{39.a}$$

$$\textit{with the constraint}: \mathbf{g}^t \mathbf{u} = 0 \tag{39.b}$$

What instead is obtained by using the perturbation is:

$$\textit{minimize}: \Pi(\mathbf{u}) \tag{40.a}$$

$$\textit{with the constraint}: \mathbf{g}^t\mathbf{u} - \frac{\boldsymbol{\lambda}}{\mu} = 0 \tag{40.b}$$

Clearly, one can see that the solutions are equivalent only as $\mu \to \infty$, but as was seen in the standard penalty approach, increasing μ leads to numerical difficulties.

Again, the two spring problem is presented within the context of the perturbed Lagrangian method in order to illustrate how it may be applied to finite element systems. The system of linear equations is:

$$\begin{bmatrix} 2k + \mu & -(k + \mu) \\ -(k + \mu) & k + \mu \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ q \end{bmatrix} \tag{41}$$

This has the solution:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \frac{q}{k} \\ \frac{q}{k}\left(\frac{2k + \mu}{k + \mu}\right) \end{bmatrix} \qquad \boldsymbol{\lambda} = -q\left(\frac{\mu}{k + \mu}\right) \tag{42}$$

Note that as $\mu \to \infty$, the exact solution is found.

2.4. The Augmented Lagrangian Method

Until now, the methods considered have had some considerable implementation drawbacks. The penalty and perturbed Lagrangian methods had solutions which are highly dependent on the penalty parameter μ. Although it is possible through the use of an expression like that presented in equation (26) to get a reasonably good value for μ based on conditioning considerations, such a value says nothing about how *accurate* a given solution might be. Moreover, the classical Lagrange method, while it promises the exact results, requires that the resulting *larger* system of equations effectively be re-ordered during the solution of the system. One might ask if it is possible to develop a method which, while retaining a relatively easy implementation, minimizes the disadvantages of the penalty and Lagrange methods. To this end, consider the following functional in equation (43).

$$\Pi'(\mathbf{u},\boldsymbol{\lambda}) = \Pi(\mathbf{u}) + \boldsymbol{\lambda}^t\mathbf{g}^t\mathbf{u} + \frac{\mu}{2}\mathbf{u}^t\mathbf{g}\mathbf{g}^t\mathbf{u}\,, \quad \textit{where}: \mu > 0 \tag{43}$$

Equation (43) can be viewed as a classical Lagrangian method:

$$\textit{minimize}: \Pi(\mathbf{u}) + \frac{\mu}{2}\mathbf{u}^t\mathbf{g}\mathbf{g}^t\mathbf{u} \tag{44.a}$$

$$\textit{with the constraint}: \mathbf{g}^t\mathbf{u} = 0 \tag{44.b}$$

because the addition of the penalty term to equation (44.a) does not change the optimal values of $\mathbf{u}$ or the Lagrange multipliers $\boldsymbol{\lambda}$. Also, note that the system can also be seen from the penalty viewpoint:

$$\textit{minimize}: \Pi(\mathbf{u}) + \boldsymbol{\lambda}^t \mathbf{g}^t \mathbf{u} \tag{45.a}$$

$$\textit{with the constraint}: \mathbf{g}^t \mathbf{u} = \mathbf{0} \tag{45.b}$$

Observe that, if the correct values of the Lagrange multipliers $\boldsymbol{\lambda}^*$ are used, equations (45) can be written:

$$\textit{minimize}: \Pi(\mathbf{u}) + (\boldsymbol{\lambda}^*)^t \mathbf{g}^t \mathbf{u} \tag{46.a}$$

$$\textit{with the constraint}: \mathbf{g}^t \mathbf{u} = \mathbf{0} \tag{46.b}$$

Or, by applying the penalty approach:

$$\Pi'(\mathbf{u},\boldsymbol{\lambda}) = \Pi(\mathbf{u}) + (\boldsymbol{\lambda}^*)^t \mathbf{g}^t \mathbf{u} + \frac{\mu}{2} \mathbf{u}^t \mathbf{g} \mathbf{g}^t \mathbf{u} \tag{47}$$

Taking the first variation of equation (47) and noting that since $\boldsymbol{\lambda}^*$ is the exact value, the result must be equal to zero:

$$\delta \mathbf{u}^t \mathbf{K} \mathbf{u} - \delta \mathbf{u}^t \mathbf{R} + \delta \mathbf{u}^t \mathbf{g} \boldsymbol{\lambda}^* + \mu \delta \mathbf{u}^t \mathbf{g} \mathbf{g}^t \mathbf{u} = \mathbf{0} \tag{48.a}$$

$$(\delta \boldsymbol{\lambda}^*)^t \mathbf{g}^t \mathbf{u} = \mathbf{0} \tag{48.b}$$

Since $\delta \mathbf{u}^t$ and $(\delta \boldsymbol{\lambda}^*)^t$ are arbitrary, equations (48) have the solution:

$$\mathbf{K} \mathbf{u} - \mathbf{R} + \mathbf{g} \boldsymbol{\lambda}^* = \mathbf{0} \tag{49.a}$$

$$\mathbf{g}^t \mathbf{u} = \mathbf{0} \tag{49.b}$$

The formulas (49) represents a set of equations for an *exact* penalty method. For details on exact penalty methods, see [3].

Since the correct values of $\boldsymbol{\lambda}$ are not known in advance, a procedure must be developed to find them, if the functional of equation (43) is to be of any use. Until now, the algorithms developed did not require iteration if the contact area between the two bodies remained constant. With the augmented Lagrangian method, an iteration is required to compute the correct values of $\mathbf{u}$ and $\boldsymbol{\lambda}$. At first this procedure might seem to be at a disadvantage, since extra computation is required. However, as will be shown, there are several advantages to this approach. The system can yield a value for $\boldsymbol{\lambda}$ within a given tolerance, and hence, $\mathbf{u}$ can be computed to a specified value. This did not occur with the penalty or perturbed Lagrangian methods. Second, the number of equations of the system does not increase. This was not the case in the classical Lagrangian method. Third, convergence to a specified value does not require that $\mu \to \infty$. This can help mitigate the severe conditioning problems associated with using a large penalty value. Fourth, and most important, in finite element problems, the contact area is generally not known in advance. This area must be found as the solution progresses. Hence, an iterative procedure is always necessary for

the finite element solution by the penalty, Lagrangian, and perturbed Lagrangian methods previously presented.

A first order update method for $\boldsymbol{\lambda}$ can be written as in equation (50).

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \mu \mathbf{g}^t \mathbf{u} \tag{50}$$

The motivation for such an updating scheme can be found if equation (45) is written as:

$$\textit{mimimize}: \Pi(\mathbf{u}) + \boldsymbol{\lambda}_k^t \mathbf{g}^t \mathbf{u} \tag{51.a}$$

$$\textit{with the constraint}: \mathbf{g}^t \mathbf{u} = \mathbf{0} \tag{51.b}$$

Viewing this from a Lagrange standpoint:

$$\delta\left[\Pi(\mathbf{u}) + \boldsymbol{\lambda}_k^t \mathbf{g}^t \mathbf{u} + (\boldsymbol{\lambda}^* - \boldsymbol{\lambda}_k)^t \mathbf{g}^t \mathbf{u}\right] = 0 \tag{52}$$

A typical step in the augmented Lagrangian method involves the minimization of:

$$\delta\left[\Pi(\mathbf{u}) + \boldsymbol{\lambda}_k^t \mathbf{g}^t \mathbf{u} + \frac{\mu}{2} \mathbf{u}^t \mathbf{g} \mathbf{g}^t \mathbf{u}\right] = 0 \tag{53}$$

So, comparing equations (52) and (53), a good approximation is:

$$\mu \mathbf{g}^t \mathbf{u} \approx \boldsymbol{\lambda}^* - \boldsymbol{\lambda}_k \tag{54}$$

Rearranging terms, $\boldsymbol{\lambda}_{k+1}$ is found.

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \mu \mathbf{g}^t \mathbf{u} \tag{55}$$

This is exactly the updating method proposed by equation (50).

Considering the two spring problem described in Figure 3 for the augmented Lagrangian method, the following system of linear equations results:

$$\begin{bmatrix} 2k + \mu & -(k + \mu) \\ -(k + \mu) & k + \mu \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ q \end{bmatrix} - \lambda \begin{bmatrix} 1 \\ -1 \end{bmatrix} \tag{56}$$

Note that in this example, $\boldsymbol{\lambda}$ is one dimensional and that, initially it is set to zero.

$$\lambda_0 = 0 \tag{57}$$

So, the equation to update $\boldsymbol{\lambda}$ is simply:

$$\lambda_{k+1} = \lambda_k + \mu (u_1 - u_2) \tag{58}$$

Solving the linear system, and updating the value of λ leads to the following series of equations.

$$\lambda_1 = \frac{-\mu q}{k + \mu} \tag{59.a}$$

$$\lambda_2 = \frac{-\mu q}{k + \mu}\left\{1 + \frac{k}{k + \mu}\right\} \tag{59.b}$$

$$\lambda_3 = \frac{-\mu q}{k+\mu}\left\{1 + \frac{k}{k+\mu} + \left[\frac{k}{k+\mu}\right]^2\right\} \tag{59.c}$$

Or, in general, equations (59) can be written as:

$$\lambda_n = \frac{-\mu q}{k+\mu}\left\{1 + \frac{k}{k+\mu} + \left[\frac{k}{k+\mu}\right]^2 + \left[\frac{k}{k+\mu}\right]^3 + \ldots + \left[\frac{k}{k+\mu}\right]^{n-1}\right\} \tag{60}$$

This has the form:

$$\lambda_n = \varphi(1 + \alpha + \alpha^2 + \alpha^3 + \ldots + \alpha^{n-1}) \tag{61}$$

for the following substitutions:

$$\varphi = \frac{-\mu q}{k+\mu} \qquad \alpha = \frac{k}{k+\mu} \tag{62}$$

An infinite geometric series, this converges to:

$$\lambda^* = \varphi\left[\frac{1}{1-\alpha}\right] = \frac{-\mu q}{k+\mu}\left[\frac{1}{1-\frac{k}{k+\mu}}\right] = -q \tag{63}$$

A requirement for convergence is that:

$$|\alpha| < 1 \quad or \quad \left|\frac{k}{k+\mu}\right| < 1 \tag{64}$$

Therefore, it is seen that this system will converge for *any* positive value of μ. At convergence, the displacements can easily be found as:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} \frac{q}{k} \\ \frac{q+\lambda^*}{k}\left\{\frac{2k+\mu}{k+\mu}\right\} - \frac{\lambda^*}{k} \end{bmatrix} = \begin{bmatrix} \frac{q}{k} \\ \frac{q}{k} \end{bmatrix} \tag{65}$$

These values represent the exact result.

For a detailed convergence proof of the augmented Lagrangian method in the general case, see [3]. The example previously presented displays the power of the method for the following reasons. First, convergence to a specified tolerance is found without requiring $\mu \to \infty$. Second, the correct displacements *and* contact forces are available without the need for extra calculations. Also note that, in practice, the penalty parameter does not have to be held fixed during the iteration: it can slowly increase as the solution nears convergence. Third, the burden of accuracy does not depend directly on μ but rather on the number of iterations used to compute $\boldsymbol{\lambda}$. Furthermore, the algorithm does not have to be run until $\boldsymbol{\lambda}^*$ is found, but rather it needs only to be computed until $\boldsymbol{\lambda}$ is within a specified tolerance of $\boldsymbol{\lambda}^*$. Finally, each iteration leads to an

improved estimate of the solution.

In order to study further the convergence properties of this method, the duality viewpoint of equation (43) will be examined. Recall that duality means that the functional which results from the problem:

$$\textit{minimize}: \Pi(\mathbf{u}) + \frac{\mu}{2}\mathbf{u}^t\mathbf{g}\mathbf{g}^t\mathbf{u} \tag{66.a}$$

$$\textit{with the constraint}: \mathbf{g}^t\mathbf{u} = \mathbf{0} \tag{66.b}$$

will lead to the system of equations:

$$\mathbf{K}\mathbf{u} - \mathbf{R} + (\boldsymbol{\lambda}^*)^t\mathbf{g} + \mu\mathbf{g}\mathbf{g}^t\mathbf{u} = 0 \tag{67}$$

when the Lagrange approach is used. By invoking a convexity assumption on the Hessian of equation (67), an equivalent dual problem can be written:

$$\varphi(\boldsymbol{\lambda}) = \textit{minimum}\left\{\Pi(\mathbf{u}) + \boldsymbol{\lambda}^t\mathbf{g}^t\mathbf{u} + \frac{\mu}{2}\mathbf{u}^t\mathbf{g}\mathbf{g}^t\mathbf{u}\right\} \tag{68}$$

Specifically, the dual function is used since it is a function of only one variable, $\boldsymbol{\lambda}$, and hence there is a corresponding simplification in the problem statement. The dual problem requires that the Hessian of the solution of the solution to equations (66) be positive definite: this is a convexity assumption. Note that from the results of the penalty method, and in fact from the results of Figure 1, the problem is locally convex near the solution point for a reasonably large value of μ. Also worth noting is that equation (68) is not used in the solution to the finite element problem since the minimization requires a solution of an unconstrained problem in terms of the *unknown* displacements. Updating schemes for $\mathbf{u}$ corresponding to equation (50) are much more difficult to develop and use.

Using $\mathbf{f}(\mathbf{x}(\boldsymbol{\lambda}))$ to represent the standard finite element system, $\mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}))$ to represent the constraint, and $\mathbf{x}(\boldsymbol{\lambda})$ as the vector which minimizes equations (66), the dual of equation (68) can be written as:

$$d(\boldsymbol{\lambda}) = f(\mathbf{x}(\boldsymbol{\lambda})) + \boldsymbol{\lambda}^t\mathbf{h}(\mathbf{x}(\boldsymbol{\lambda})) + \frac{\mu}{2}\mathbf{h}^t(\mathbf{x}(\boldsymbol{\lambda}))\mathbf{h}(\mathbf{x}(\boldsymbol{\lambda})) \tag{69}$$

Equation (69) has the gradient:

$$\nabla d(\boldsymbol{\lambda}) = \nabla_{\boldsymbol{\lambda}}\mathbf{x}(\boldsymbol{\lambda})\left\{\nabla_{\mathbf{x}}\mathbf{f}(\mathbf{x}(\boldsymbol{\lambda})) + \nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}))\boldsymbol{\lambda} + \mu\nabla_{\mathbf{x}}\mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}))\mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}))\right\} + \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda})) \tag{70}$$

$$= \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}))$$

The term in braces vanishes at the stationary point, because it corresponds the first variation of the functional given by equation (45). The Hessian of the dual is now calculated in order to examine the convergence properties of the

augmented Lagrangian method.

$$\nabla^2 d = \nabla_\lambda \mathbf{x}(\lambda) \nabla_x \mathbf{h}(\mathbf{x}(\lambda)) \tag{71}$$

If the term in braces of equation (70) is described as:

$$\nabla_x L(\mathbf{x}(\lambda),\lambda) = \nabla_x \mathbf{f}(\mathbf{x}(\lambda)) + \lambda^t \nabla_x \mathbf{h}(\mathbf{x}(\lambda)) + \mu \nabla_x \mathbf{h}(\mathbf{x}(\lambda)) \mathbf{h}(\mathbf{x}(\lambda)) = \mathbf{0} \tag{72}$$

then, upon differentiating this expression one finds (73).

$$\nabla_\lambda \mathbf{x}(\lambda) \nabla^2_{xx} L(\mathbf{x}(\lambda),\lambda) + \nabla^2_{\lambda x} L(\mathbf{x}(\lambda),\lambda) = \mathbf{0} \tag{73}$$

Equation (73) can be further simplified to:

$$\nabla_\lambda \mathbf{x}(\lambda) \nabla^2_{xx} L(\mathbf{x}(\lambda),\lambda) + \nabla_x \mathbf{h}^t(\mathbf{x}(\lambda)) = \mathbf{0} \tag{74}$$

Finally, $\nabla_\lambda \mathbf{x}(\lambda)$ can be found by solving equation (74).

$$\nabla_\lambda \mathbf{x}(\lambda) = -\nabla_x \mathbf{h}^t(\mathbf{x}(\lambda)) \nabla^2_{xx} L^{-1}(\mathbf{x}(\lambda),\lambda) \tag{75}$$

Substituting this result into equation (71), the Hessian is listed in (76).

$$\begin{aligned} \nabla^2 d(\lambda) &= -\nabla_x \mathbf{h}^t(\mathbf{x}(\lambda)) \nabla^2_{xx} L^{-1}(\mathbf{x}(\lambda),\lambda) \\ &= -\nabla_x \mathbf{h}^t(\mathbf{x}(\lambda)) \left\{ \nabla^2_{xx} \mathbf{f}(\mathbf{x}(\lambda)) + \mu \nabla_x \mathbf{h}(\mathbf{x}(\lambda)) \nabla_x \mathbf{h}^t(\mathbf{x}(\lambda)) \right\}^{-1} \nabla_x \mathbf{h}(\mathbf{x}(\lambda)) \end{aligned} \tag{76}$$

The steepest ascent formula for maximizing $d(\lambda)$ of equation (69) can be written as:

$$\lambda_{k+1} = \lambda_k + \alpha \nabla d(\lambda_k) \tag{77}$$

Or, by substituting the results of (70), equation (77) becomes:

$$\lambda_{k+1} = \lambda_k + \alpha \mathbf{h}(\mathbf{x}(\lambda_k)) \tag{78}$$

And, from reference [3], a relationship between the values of λ and the convergence ratio is:

$$\frac{\lambda_{k+1} - \lambda^*}{\lambda_k - \lambda^*} \le r(\mu) = \max \left(\left| 1 - \mu W \right|, \left| 1 - \mu w \right| \right) \tag{79}$$

Where W is the maximum eigenvalue of $\nabla^2 d(\lambda)$, w is the minimum eigenvalue of $\nabla^2 d(\lambda)$, the function $r(\mu)$ is the *convergence ratio* and μ^* is the optimal value of μ. This optimal value can be computed to be:

$$\mu^* = \frac{2}{W + w} \tag{80}$$

And, the convergence ratio at that value is:

$$r(\mu^*) = \frac{W - w}{W + w} \tag{81}$$

Similarly, the Hessian of the primal function can be computed to be:

$$\nabla^2 p(\mathbf{u}) = \left\{\nabla_x \mathbf{h}^t(\mathbf{x}(\mathbf{u}))\left[\nabla^2_{xx}\mathbf{f}(\mathbf{x}(\mathbf{u}),\boldsymbol{\lambda}(\mathbf{u})) + \mu\nabla_x \mathbf{h}(\mathbf{x}(\mathbf{u}))\nabla_x \mathbf{h}^t(\mathbf{x}(\mathbf{u}))\right]^{-1}\nabla_x \mathbf{h}(\mathbf{x}(\mathbf{u}))\right\}^{-1} \qquad (82)$$

$$-\mu\,\mathbf{I}$$

Substituting and simplifying, the Hessians of both the dual and primal functionals become:

$$\nabla^2 d(\boldsymbol{\lambda}) = -\mathbf{g}^t\left[\mathbf{K} + \mu\mathbf{g}\mathbf{g}^t\right]^{-1}\mathbf{g} \qquad (83)$$

$$\nabla^2 p(\mathbf{u}) = \left[\mathbf{g}^t\left[\mathbf{K} + \mu\mathbf{g}\mathbf{g}^t\right)^{-1}\mathbf{g}\right]^{-1} - \mu\,\mathbf{I} \qquad (84)$$

Next, consider the following rank one update formula:

$$\mu\mathbf{g}^t(\mathbf{K} + \mu\mathbf{g}\mathbf{g}^t)^{-1}\mathbf{g} = \mathbf{I} - (\mathbf{I} + \mu\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g})^{-1} \qquad (85)$$

Multiplying both sides by $(\mathbf{I} + \mu\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g})$ on the right leads to equation (86).

$$\mu\mathbf{g}^t(\mathbf{K} + \mu\mathbf{g}\mathbf{g}^t)^{-1}\mathbf{g}(\mathbf{I} + \mu\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}) = \mu\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g} \qquad (86)$$

Suppose $\mathbf{e}$ is an eigenvector of $\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}$ with an eigenvalue a. Post-multiplying equation (86) by $\mathbf{e}$ and simplifying, it can be written as equation (87).

$$\mu\mathbf{g}^t(\mathbf{K} + \mu\mathbf{g}\mathbf{g}^t)^{-1}\mathbf{g}(1 + \mu a)\mathbf{e} = \mu a\,\mathbf{e} \qquad (87)$$

This equation has the form:

$$\mu\mathbf{A}(1 + \mu a)\mathbf{e} = \mu a\,\mathbf{e} \qquad (88.a)$$

or:

$$\mathbf{A}k\,\mathbf{e} = a\,\mathbf{e} \qquad (88.b)$$

So, $\mathbf{e}$ is also an eigenvector of A. Thus $\mathbf{g}^t(\mathbf{K} + \mu\mathbf{g}\mathbf{g}^t)^{-1}\mathbf{g}$ and $\mathbf{g}^t\mathbf{K}^{-1}\mathbf{g}$ have the same eigenvectors. Using this information in equation (87) and the symbol b to represent and eigenvalue of $\mathbf{g}^t(\mathbf{k} + \mu\mathbf{g}\mathbf{g}^t)^{-1}\mathbf{g}$, a relationship between eigenvalues a and b can be written:

$$b(1 + \mu a) = a \qquad (89.a)$$

or:

$$b = \frac{a}{1+\mu a} \qquad (89.b)$$

By substituting these values, the convergence ratio is:

$$r(\mu) = \frac{\frac{1}{w} + \mu}{\frac{1}{w} + \mu} \qquad (90)$$

Clearly, as $\mu \to \infty$, the convergence ratio goes to unity, implying arbitrarily fast convergence. By application of the analysis of the penalty method, it is known that from equations (13) to (26), rounding errors also play a part. Thus, a tradeoff situation has developed. Although increasing the penalty parameter increases the rate of convergence, it also leads to more poorly conditioned set of equations. Since the augmented Lagrangian will converge, within certain restrictions, for any value of the penalty parameter it is clear that "best" value to use is the one which will minimize the rounding errors. Thus, this optimal value for the penalty parameter μ has exactly the same form as equation (26).

As a further justification for the first order update given by equation (50), consider the application of Newton's method, and an approximation to the Hessian of the dual functional. The value of $\boldsymbol{\lambda}_{k+1}$ can be thought of as minimizing the second order expansion of the dual around $\boldsymbol{\lambda}_k$:

$$d_k(\boldsymbol{\lambda}) = d(\boldsymbol{\lambda}_k) + \nabla d^t(\boldsymbol{\lambda}_k)(\boldsymbol{\lambda} - \boldsymbol{\lambda}_k) + \tfrac{1}{2}(\boldsymbol{\lambda} - \boldsymbol{\lambda}_k)^t \nabla^2 d(\boldsymbol{\lambda}_k)(\boldsymbol{\lambda} - \boldsymbol{\lambda}_k) \tag{91}$$

Taking the derivative and setting it equal to zero:

$$\nabla d_k(\boldsymbol{\lambda}) = \nabla d(\boldsymbol{\lambda}_k) + \nabla^2 d(\boldsymbol{\lambda}_k)(\boldsymbol{\lambda} - \boldsymbol{\lambda}_k) = 0 \tag{92}$$

This leads to the update formula:

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k - \left[\nabla^2 d(\boldsymbol{\lambda}_k)\right]^{-1} \nabla d(\boldsymbol{\lambda}_k) \tag{93}$$

By equation (70), $\nabla d(\boldsymbol{\lambda}_k)$ is known, and from equation (76) $\nabla^2 d(\boldsymbol{\lambda}_k)$ is also known. Equation (83) represents the system of equations for the contact problem: this expression could be substituted into equation (93). As a further simplification, note that for a reasonably large μ, equation (83) is approximately $-\frac{\mathbf{I}}{\mu}$. Hence, equation (93) can be expressed as:

$$\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \mu \mathbf{g}^t \mathbf{u} \tag{94}$$

This is exactly the method proposed in equation (50).

The augmented Lagrangian procedure can be interpreted geometrically in Figure 6. Consider the primal functional given in equation (43). By addition of the penalty term, the system is convex near the solution, $\mathbf{u}^*$, $\boldsymbol{\lambda}^*$. For a given set of starting values for $\mathbf{u}_0$ and $\boldsymbol{\lambda}_0$ the solution proceeds as follows. The value of $\mathbf{u}_k$ is found and is used to update the Lagrange parameter $\boldsymbol{\lambda}_{k+1}$. This value corresponds to the negative slope of the functional at the point $\mathbf{u}_k$. The new value of $\boldsymbol{\lambda}_{k+1}$ is then used to compute a new $\mathbf{u}_{k+1}$. This process may be continued until the values $\mathbf{u}_{k+n}$ and $\boldsymbol{\lambda}_{k+n}$ are within a specified tolerance to $\mathbf{u}^*$ and $\boldsymbol{\lambda}^*$.

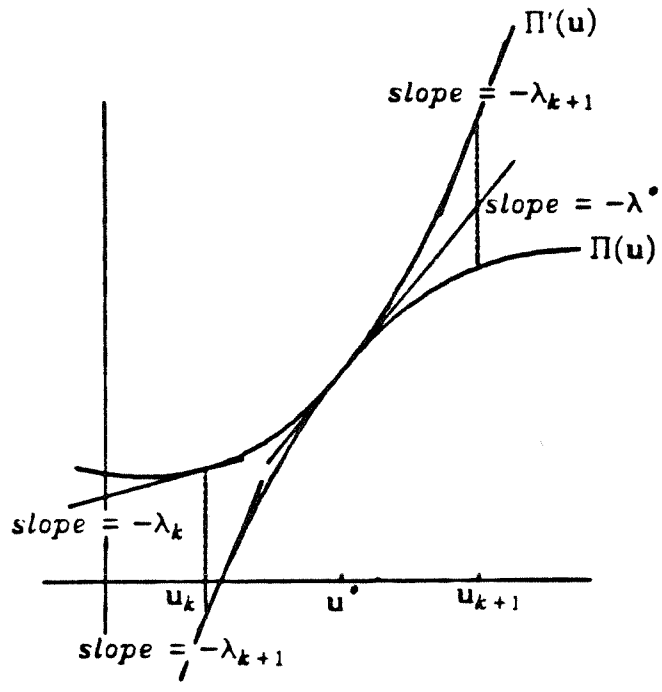

Figure 6: Graphical interpretation of the augmented Lagrangian method.

3. Finite Element Implementation

The finite element method may be used to construct solution procedures for the four contact schemes discussed in Section 2. In this section, only the frictionless, two-dimensional, small deformation kinematics case will be presented. Since the algorithms for the penalty and classical Lagrange methods are so well known, the discussion of these methods will be confined to their relationship with the augmented Lagrangian scheme. Details on the perturbed Lagrangian algorithm are given in references [6] and [7].

The augmented Lagrangian method has been implemented in the finite element program FEAP [9]. Details of the contact kinematics can be found in [6] and [7]. In these references, the contact interface is first split into a number of segments along an intermediate surface, as shown in Figure 7. Assuming four node isoparametric elements, these segments are unambiguously defined by the geometry of the adjacent elements of the bodies in contact. Figure 8 shows a typical segment.

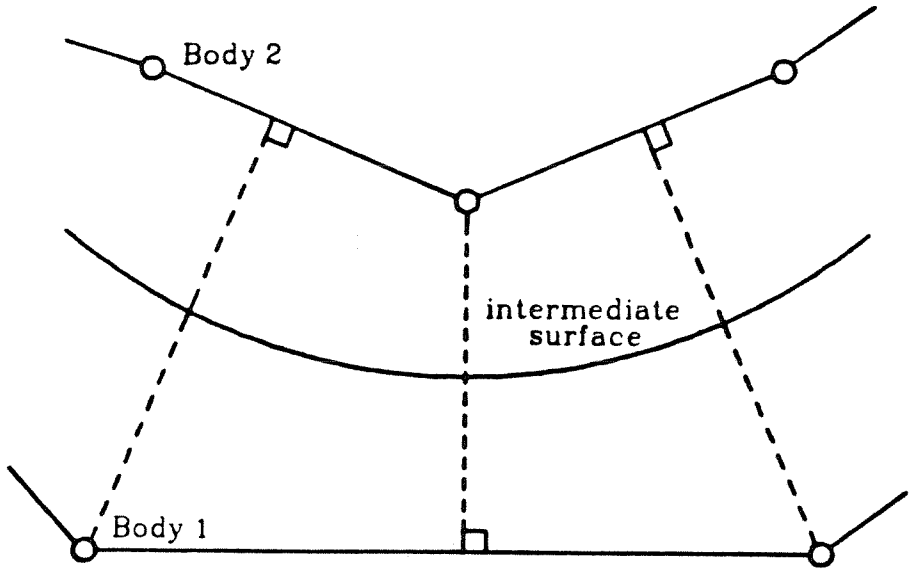

Figure 7: Discretization into contact segments.

Variable	Description
$\mathbf{x}_j^i$	Element node on body i at node j
$\bar{\mathbf{x}}^i$	Projection onto body i of $\mathbf{x}_j^i$
$\mathbf{u}_j^i$	Displacement of body i at node j
$\bar{\mathbf{u}}^i$	Projection onto body i of $\mathbf{u}_j^i$
$\mathbf{n}^i$	Normal to the contact segment on body i

Table 1: A select list of variables and their description.

The notation in Table 1 is used to obtain the projection of contact segments described in (95).

$$\bar{\mathbf{x}}^i = (1 - \alpha^i)\mathbf{x}_1^i + \alpha^i \mathbf{x}_2^i \tag{95}$$

The variables in equation (95) are defined by:

$$\alpha^1 = \frac{(\mathbf{x}_2^2 - \mathbf{x}_1^1) \cdot \mathbf{t}^1}{\| \mathbf{x}_2^1 - \mathbf{x}_1^1 \|} \tag{96.a}$$

$$\alpha^2 = \frac{(\mathbf{x}_2^1 - \mathbf{x}_1^2) \cdot \mathbf{t}^2}{\| \mathbf{x}_2^2 - \mathbf{x}_1^2 \|} \tag{96.b}$$

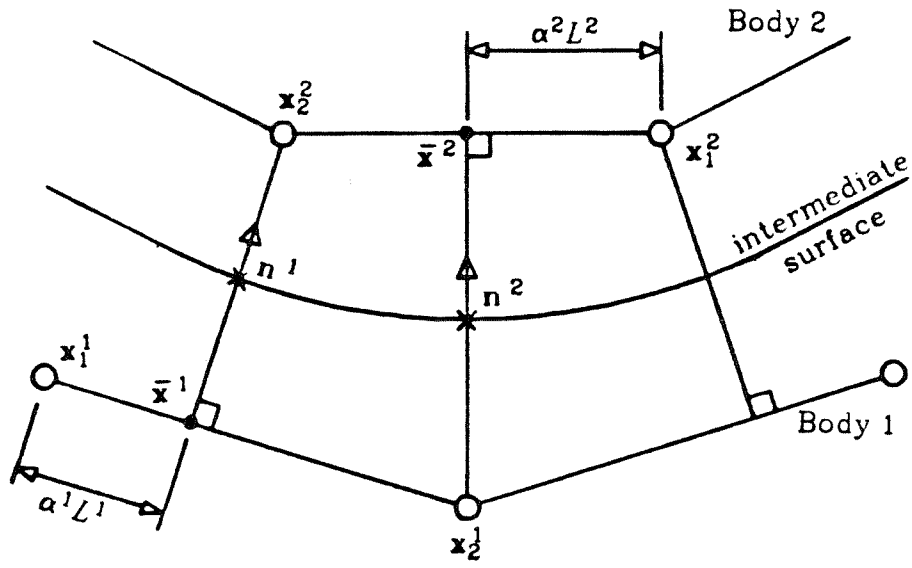

Figure 8: A typical contact segment.

$$\mathbf{t}^i = \frac{(\mathbf{x}_2^i - \mathbf{x}_1^i)}{|\mathbf{x}_2^i - \mathbf{x}_1^i|} \tag{96.c}$$

From equations (96), the displacement can be described as:

$$\bar{\mathbf{u}}^i = (1 - \alpha^i)\mathbf{u}_1^i + \alpha^i \mathbf{u}_2^i \tag{97}$$

And, hence, the gaps in the contact segments can be represented by:

$$g^1 = (\mathbf{u}_2^2 - \bar{\mathbf{u}}^1) \cdot \mathbf{n}^1 \tag{98.a}$$

$$g^2 = (\mathbf{u}_2^1 - \bar{\mathbf{u}}^2) \cdot \mathbf{n}^2 \tag{98.b}$$

The normal $\mathbf{n}^i$ is defined by:

$$\mathbf{n}^i = \mathbf{e}_3 \otimes \mathbf{t}^i \tag{99}$$

By applying a linear interpolation to the displacement field described by equation (97), a consistent approximation is obtained with the four node iso-parametric element. In the context of this approximation, the contact pressure within each segment is a constant. Note that this approach is not based on nodal penetration of the adjacent body, but rather an *average* penetration across a contact segment. Thus, in the solution to a finite element contact problem, the constraint is satisfied in an "average" sense over a segment instead of "point-wise" for a contact node. By using the expressions in equation (98) for the interpolation functions $\mathbf{g}$, in any of the equations (5), (27), (32) or

(43) for $\mathbf{g}^t\mathbf{u}$, this averaging can be incorporated into the solution methods previously discussed.

Consider equation (5), which represents the penalty method. Assuming two bodies in contact, the functional can be written as in equation (100).

$$\Pi'(\mathbf{u}^1,\mathbf{u}^2) = \Pi(\mathbf{u}^1) + \Pi(\mathbf{u}^2) + \int_{\delta_c} \frac{\mu}{2}\mathbf{g}^t\mathbf{g}\, da \tag{100}$$

Where the first two terms of equation (100) are of the same form as equation (1):

$$\Pi(\mathbf{u}_1) = \int_{B_1} \mathbf{W}(\mathbf{u})\, dv_1 - \int_{B_1} \rho\mathbf{b}\cdot\mathbf{u}\, dv_1 - \int_{dB_1} \bar{\mathbf{t}}\cdot\mathbf{u}\, da_1 \tag{101.a}$$

$$\Pi(\mathbf{u}_2) = \int_{B_2} \mathbf{W}(\mathbf{u})\, dv_2 - \int_{B_2} \rho\mathbf{b}\cdot\mathbf{u}\, dv_2 - \int_{dB_2} \bar{\mathbf{t}}\cdot\mathbf{u}\, da_2 \tag{101.b}$$

By substituting equation (98) into equation (100), the following expression is obtained:

$$\Pi'(\mathbf{u}) = \Pi(\mathbf{u}) + \frac{\mu}{2}\int_{\delta_c} \left\{\left[\bar{\mathbf{u}}^2 - \bar{\mathbf{u}}^1\right]\cdot\bar{\mathbf{n}}\right\}^t \left\{\left[\bar{\mathbf{u}}^2 - \bar{\mathbf{u}}^1\right]\cdot\bar{\mathbf{n}}\right\} da \tag{102}$$

The kinematically admissible variations are:

$$\mathbf{V}^1 = \left\{\eta_1 : \mathbf{v}^1(\Omega_1)\rightarrow\mathbf{R}^3\,,\ \eta_1 \mid_{dB_1} = \mathbf{0}\right\} \tag{103.a}$$

$$\mathbf{V}^2 = \left\{\eta_2 : \mathbf{v}^2(\Omega_2)\rightarrow\mathbf{R}^3\,,\ \eta_2 \mid_{dB_2} = \mathbf{0}\right\} \tag{103.b}$$

Applying the Gateaux derivative to equation (102), and invoking stationarity:

$$\mathbf{D}_{\mathbf{v}_\alpha^1}\Pi\cdot\eta_1 = \frac{d}{d\mathbf{v}^1}\Pi(\mathbf{v}^1 + \alpha\eta_1)\mid_{\alpha=0}\cdot\frac{d}{\alpha}(\mathbf{v}^1 + \alpha\eta_1)\mid_{\alpha=0} + \tag{104.a}$$

$$\frac{\mu}{2}\int_{\delta_c}\frac{d}{d\mathbf{v}^1}\left\{\left[\left(\mathbf{v}^2 + \alpha\eta_2\right) - \left(\mathbf{v}^1 + \alpha\eta_1\right)\right]\cdot\bar{\mathbf{n}}\right\}^t$$

$$\left\{\left[\left(\mathbf{v}^2 + \alpha\eta_2\right) - \left(\mathbf{v}^1 + \alpha\eta_1\right)\right]\cdot\bar{\mathbf{n}}\right\}\Bigg|_{\alpha=0} da = 0$$

$$\mathbf{D}_{\mathbf{v}_\alpha^2}\Pi\cdot\eta_2 = \frac{d}{d\mathbf{v}^2}\Pi(\mathbf{v}^2 + \alpha\eta_2)\mid_{\alpha=0}\cdot\frac{d}{\alpha}(\mathbf{v}^2 + \alpha\eta_2)\mid_{\alpha=0} + \tag{104.b}$$

$$\frac{\mu}{2}\int_{\delta_c}\frac{d}{d\mathbf{v}^2}\left\{\left[\left(\mathbf{v}^2+\alpha\eta_2\right)-\left(\mathbf{v}^1+\alpha\eta_1\right)\right]\cdot\bar{\mathbf{n}}\right\}^t$$

$$\left\{\left[\left(\mathbf{v}^2+\alpha\eta_2\right)-\left(\mathbf{v}^1+\alpha\eta_1\right)\right]\cdot\bar{\mathbf{n}}\right\}\Bigg|_{\alpha=0} d\alpha = 0$$

Or, simplifying and combining these two equations the following expression is obtained.

$$D_{\mathbf{v}^1}\Pi\cdot\eta_1 + D_{\mathbf{v}^2}\Pi\cdot\eta_2 + \mu\int_{\delta_c}(\eta_2-\eta_1)\cdot\bar{\mathbf{n}}\left\{\left[\mathbf{v}^2-\mathbf{v}^1\right]\cdot\bar{\mathbf{n}}\right\}^t d\alpha = 0 \qquad (105)$$

Applying the linear finite element discretization, equation (105) becomes:

$$\sum_{e}^{E_{total}} G_e + \sum_{s}^{S_{total}}\left\{\mu\sum_{i=1}^{4}\eta_i\cdot\mathbf{c}_i\sum_{j=1}^{4}\mathbf{c}_j\cdot\mathbf{u}_j\right\} = 0 \qquad (106)$$

Where the first sum represents the standard contributions of stiffness and force to the finite element system, and the second sum represents the contact constraint contribution. The vector $\mathbf{c}_i$ can be expressed with reference to Figure 8 as follows.

$$\mathbf{c}_1 = -\tfrac{1}{2}(1-\alpha^1)\,\mathbf{n}^1 \qquad (107.a)$$

$$\mathbf{c}_2 = +\tfrac{1}{2}(\mathbf{n}^2-\alpha^1\mathbf{n}^1) \qquad (107.b)$$

$$\mathbf{c}_3 = +\tfrac{1}{2}(\mathbf{n}^1-\alpha^2\mathbf{n}^2) \qquad (107.c)$$

$$\mathbf{c}_4 = -\tfrac{1}{2}(1-\alpha^2)\,\mathbf{n}^2 \qquad (107.d)$$

In equation (106), the sum is taken over all of the elements representing the discretization of both bodies for the first term, and the sum over all segments in contact for the second term. A typical contribution of a contact element to the system stiffness matrix can be described as:

$$\mathbf{K}^s_{contact} = \mu\begin{bmatrix}\mathbf{c}_1\\ \mathbf{c}_2\\ \mathbf{c}_3\\ \mathbf{c}_4\end{bmatrix}\begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \mathbf{c}_4\end{bmatrix} \qquad (108)$$

Which, for each segment in contact, represents a rank one update to the system stiffness matrix.

Some difficulties are present with this kinematic formulation, particularly when higher order interpolations are used or an extension to three dimensions is made. The small deformation, linear kinematics implementation has a

reasonable geometrical definition, as described in Figures 7 and 8, for most cases. However, the extension to multiple contact segments along a single slide segment, or the general application of symmetric boundary conditions is presently unclear. Figure 9 shows some problems that may arise even for the four node case. Here, the interpolation method breaks down because projections are not simple enough to be described by the application of of the method shown in Figure 8.

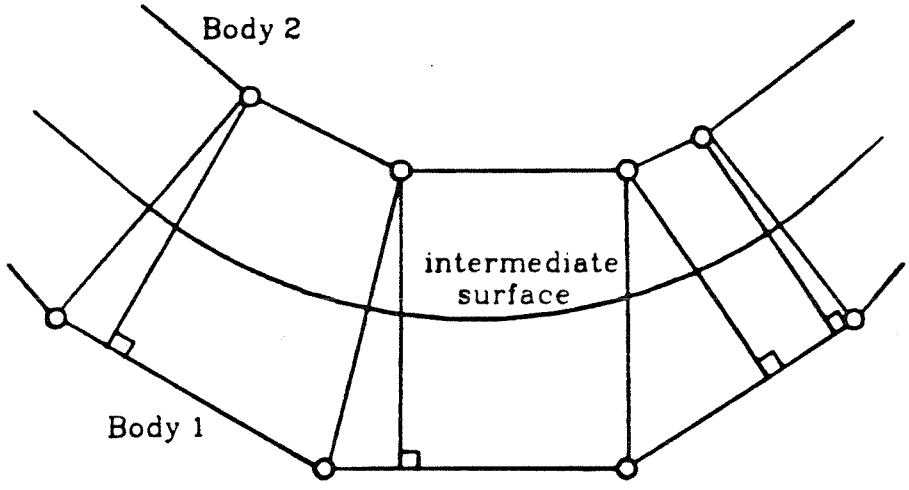

Figure 9: Possible problems with the slide line logic.

Furthermore, ambiguity arises in the case of higher order interpolations of the intermediate surface. While this problem could be overcome by a complex calculation, it is not clear that the test for penetration could be done efficiently. For example, determining if a node is penetrating a body requires the use of a concave clipping algorithm. For boundaries represented by polynomials greater than first order, the calculation can be very expensive [10]. Moreover, in order to extend this approach to three dimensions, techniques from constructive solid geometry (CSG) would have to be employed. This means that the penetration calculations might generally take longer than the solution of the system of equations for the entire finite element system [11].

The finite element implementation of the augmented Lagrangian method can now be discussed in more detail. By applying the procedure given in the last section to the functional expressed in equation (43), a finite element system of the form of equations (106) can be derived. This system is shown in equation (109.a) in matrix notation.

$$\sum_{e}^{E_{total}} \left\{ \mathbf{K}^e \mathbf{u}^e - \mathbf{r}^e \right\} + \sum_{s}^{S_{total}} \left\{ \mu \mathbf{c}^s (\mathbf{c}^s)^t \mathbf{u} + \mathbf{c}_s \lambda_k^s \right\} = \mathbf{0} \qquad (109.a)$$

The equation for updating λ can be written as:

$$\lambda_{k+1}^{s} = \lambda_{k}^{s} + \mu(\mathbf{c}^{s})^{t}\mathbf{u} \tag{109.b}$$

Several features are worth repeating about equations (109). First, the summation extends over all of the elements representing the discretization of the bodies in contact. Second, the contact stiffness contribution is a rank one update of the global stiffness matrix over each segment in contact. Third, the method is inherently iterative because the Lagrange parameter is not known before the solution is attempted. Fourth, the interpolation across a contact segment is done in an "average" sense, rather than "point-wise." Finally, note that even though the solution is iterative due to the nature of equation (109), it is also not known in advance which segments are in contact: each loading configuration requires control logic based on the current values of the displacements. Newton's method is used to solve equations (109). The solution procedure is outlined as Algorithm 1.

In practice, each contact segment has information containing the current interpolation functions $\mathbf{c}$, current Lagrange parameter λ, and penalty value μ. Generally, the penalty value varies slowly from its original value to something about ten to a hundred times larger as the solution converges. This has the effect of tightening the constraint as the answer becomes better known while keeping the solution reasonably well conditioned at the early stages.

The fact that the augmented Lagrangian method is a combination of both the penalty and Lagrangian methods has already been shown. Depending on ones viewpoint, it is relatively straightforward to extend or simplify Algorithm 1 to incorporate these cases. For the penalty method, the Lagrange parameters can obviously be ignored, leading to a system of equations analogous to equation (109):

$$\sum_{e}^{E_{total}} (\mathbf{K}^{e}\mathbf{u} - \mathbf{R}^{e}) + \sum_{s}^{S_{total}} \mu \mathbf{c}^{s}(\mathbf{c}^{s})^{t}\mathbf{u} = \mathbf{0} \tag{110}$$

A similar procedure holds for the perturbed Lagrangian method, since equation (36) expresses a value for the Lagrange parameter across the contact segment. Here, the step in Algorithm 1 which updates the Lagrange parameter may be skipped since this item is considered constant across a given load configuration.

Step (II.D) in Algorithm 1 states that the *exact* value of λ does not have to be found. The user can select the desired precision. Furthermore, this precision is, unlike that of the penalty, classical Lagrangian or perturbed Lagrangian methods, independent of the penalty value. This penalty value, by the results of section 2.4, governs the *rate of convergence* of the augmented Lagrangian method. It is generally expected that results to machine precision can be

Static Augmented Lagrangian Method

I. Set $\mathbf{u}_0 = \mathbf{0}$, $\boldsymbol{\lambda}_0 = \mathbf{0}$

II. For k = 1, 2, 3, ... until convergence do:

A. Build the system $\mathbf{K}_{system}$ and $\mathbf{R}_{system}$

B. For s = 1 to S_{total}

1. Determine $\mathbf{c}^s$, $g^s = (\mathbf{c}^s)^t \mathbf{u}_k^s$

2. Check to see if the segment is in contact:

If $\left[g^s > \frac{\lambda^s}{\mu} \right]$

then the segment is **not** in contact

else

a. Calculate the stiffness contribution: $\mathbf{K}^s_{contact} = \mu \mathbf{c}^s (\mathbf{c}^s)^t \mathbf{u}_k$

b. Calculate the residual contribution: $\mathbf{R}^s_{contact} = -(\lambda_k^s)^t \mathbf{c}^s$

c. Add the contact $\mathbf{K}^s_{contact}$, $\mathbf{R}^s_{contact}$ to system $\mathbf{K}_{system}$, $\mathbf{R}_{system}$ to get $\mathbf{K}_{total}$, $\mathbf{R}_{total}$

C. Solve the system for the displacement increment:

$\mathbf{K}_{total} \ \Delta \mathbf{u} = \mathbf{R}_{total} - \mathbf{K}_{total} \ \mathbf{u}_k$

D. Check for convergence:

if $\| \mathbf{K}_{total} \ \mathbf{u}_k - \mathbf{R}_{total} \| < tolerance$

then announce convergence and **stop**.

E. Update the displacement field: $\mathbf{u}_{k+1} = \mathbf{u}_k + \Delta \mathbf{u}_{k+1}$

F. For s = 1 to S_{total}:

Update the Lagrange parameters: $\lambda^s_{k+1} = \lambda^s_k + \mu (\mathbf{c}^s)^t \mathbf{u}_{k+1}$

G. Next k, go back to step II.

Algorithm 1: Description of the static augmented Lagrangian method.

obtained after only a few iterations.

Since the algorithm depends on determining the state of a contact segment in step (II.B.2), it should be noted that the convergence properties discussed in section 2.4 might not always apply. This is because, as segments change state from "not in contact" to "in contact" or vice-versa, the tangent matrix becomes discontinuous. For problems involving practical engineering applications, these discontinuities last only during the initial iterations and the solution quickly converges.

The augmented Lagrangian method can easily be extended to handle problems involving dynamic contact. Algorithm 2 presents a solution of the problem

Dynamic Augmented Lagrangian Method

I. Initialize ${}^0\mathbf{u}$, ${}^0\dot{\mathbf{u}}$, ${}^0\ddot{\mathbf{u}}$.

II. For $t = \Delta t, 2\Delta t, 3\Delta t, \ldots$ until T_{final} do:

A. Set $\mathbf{u}_0$, $\boldsymbol{\lambda}_0 = \mathbf{0}$

B. For k = 1, 2, 3, ... until convergence do:

1. Build the system ${}^t\mathbf{K}_{system}$, ${}^t\mathbf{M}_{system}$ and ${}^t\mathbf{R}_{system}$

2. For s = 1 to S_{total} do:

a. Determine $\mathbf{c}^s$, $g^s = (\mathbf{c}^s)^t\mathbf{u}_k^s$

b. Check to see if the segment is in contact: If $\left(g^s > \frac{\lambda^s}{\mu}\right)$

then the segment is **not** in contact else

i. Calculate the stiffness contribution: $\mathbf{K}^s_{contact} = \mu \mathbf{c}^s (\mathbf{c}^s)^t\mathbf{u}_k$

ii. Calculate the residual contribution: $\mathbf{R}^s_{contact} = -(\lambda_k^s)^t\mathbf{c}^s$

iii. Add the contact $\mathbf{K}^s_{contact}$, $\mathbf{R}^s_{contact}$ to system ${}^t\mathbf{K}_{system}$, ${}^t\mathbf{R}_{system}$ to get K_{total} , R_{total}

3. Build ${}^t\mathbf{K}_{dynamic} = {}^t\mathbf{K}_{system} + a_0{}^t\mathbf{M}_{system}$

4. Calculate the residual:

$${}^t\mathbf{R}_{dynamic} = {}^t\mathbf{R}_{total} - {}^t\mathbf{K}_{total}{}^t\mathbf{u} - {}^t\mathbf{M}_{system}\left\{a_0\left[\mathbf{u}_k - {}^t\mathbf{u}\right] - a_2{}^t\dot{\mathbf{u}} - a_3{}^t\ddot{\mathbf{u}}\right\}$$

5. Solve the system for the displacement increment: ${}^t\mathbf{K}_{dynamic}\ \Delta\mathbf{u} = {}^t\mathbf{R}_{dynamic}$

6. Update the current configuration: $\mathbf{u}_{k+1} = \mathbf{u}_k + \Delta\mathbf{u}$

7. Check for convergence:

if $\left(\left\| {}^t\mathbf{R}_{total} - {}^t\mathbf{K}_{total}\mathbf{u}_{k+1} - {}^t\mathbf{M}_{system}{}^t\ddot{\mathbf{u}} \right\| < tolerance\right)$ then

a. announce convergence in k: ${}^{t+1}\mathbf{u} = \mathbf{u}_{k+1}$

b. Update the velocity and acceleration fields:

$${}^{t+1}\ddot{\mathbf{u}} = a_0\left({}^{t+1}\mathbf{u} - {}^t\mathbf{u}\right) - a_2{}^t\dot{\mathbf{u}} - a_3{}^t\ddot{\mathbf{u}}$$

$${}^{t+1}\dot{\mathbf{u}} = {}^t\dot{\mathbf{u}} + a_6{}^t\dot{\mathbf{u}} + a_7{}^{t+1}\ddot{\mathbf{u}}$$

c. Go back to step II.

9. For s = 1 to S_{total}:

Update the Lagrange parameters: $\lambda^s_{k+1} = \lambda^s_k + \mu(\mathbf{c}^s)^t\mathbf{u}_{k+1}$

10. Go back to to step II.B

Algorithm 2: Description of the dynamic augmented Lagrangian method.

The parameters in Algorithm 2 are defined by:

$$\alpha \geq 0.50 \quad \beta \geq 0.25\left\{0.5 + \alpha\right\}^2 \tag{111}$$

$$a_0 = \frac{1}{\beta \Delta t^2} \quad a_1 = \frac{\alpha}{\beta \Delta t} \quad a_2 = \frac{1}{\beta \Delta t} \quad a_3 = \frac{1}{2\beta} - 1 \quad a_4 = \frac{\alpha}{\beta} - 1$$

$$a_5 = \Delta \frac{t}{2}\left\{\frac{\alpha}{\beta} - 2\right\} a_6 = \Delta t\left\{1 - \alpha\right\} \quad a_7 = \alpha \Delta T$$

in the time domain by the Newmark method. Note that for linear elastic problems, steps which involve forming $\mathbf{K}_{system}$ and $\mathbf{R}_{system}$ could be moved out of the loops in both Algorithms 1 and 2.

Below, the two spring problem is again considered for a suddenly applied load. A comparison with the exact results in Figure 10 shows excellent agreement.

$$\mathbf{K}_{system} = \begin{bmatrix} 2k + \mu & -(k + \mu) \\ -(k + \mu) & k + \mu \end{bmatrix} \quad \mathbf{M}_{system} = \begin{bmatrix} \frac{2}{3}m & \frac{1}{6}m \\ \frac{1}{6}m & \frac{1}{3}m \end{bmatrix} \tag{112}$$

This system leads to the following collection of matrices:

$$\mathbf{K}_{dynamic} = \begin{bmatrix} 2k + \mu & -(k + \mu) \\ -(k + \mu) & k + \mu \end{bmatrix} + a_0 \begin{bmatrix} \frac{2}{3}m & \frac{1}{6}m \\ \frac{1}{6}m & \frac{1}{3}m \end{bmatrix} \tag{113}$$

$$\mathbf{R}_{dynamic} = \begin{bmatrix} 0 \\ -.1 \end{bmatrix} - \begin{bmatrix} 2k + \mu & -(k + \mu) \\ -(k + \mu) & k + \mu \end{bmatrix} \begin{bmatrix} {}^t u_1 \\ {}^t u_2 \end{bmatrix} - \lambda_k \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \tag{114}$$

$$\begin{bmatrix} \frac{2}{3}m & \frac{1}{6}m \\ \frac{1}{6}m & \frac{1}{3}m \end{bmatrix} \left\{ a_0 \left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \begin{bmatrix} {}^t u_1 \\ {}^t u_2 \end{bmatrix} \right) - a_2 \begin{bmatrix} {}^t \dot{u}_1 \\ {}^t \dot{u}_2 \end{bmatrix} - a_3 \begin{bmatrix} {}^t \ddot{u}_1 \\ {}^t \ddot{u}_2 \end{bmatrix} \right\}$$

The initial conditions are assumed to be:

$$\begin{bmatrix} {}^0 u_1 \\ {}^0 u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} {}^0 \dot{u}_1 \\ {}^0 \dot{u}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \begin{bmatrix} {}^0 \ddot{u}_1 \\ {}^0 \ddot{u}_2 \end{bmatrix} = \begin{bmatrix} \frac{.6}{7} \\ \frac{-2.4}{7} \end{bmatrix} \tag{115}$$

Furthermore, the system constants are taken to be:

$$\alpha = \frac{1}{2} \quad \beta = \frac{1}{4} \quad k = 1 \quad m = 1 \quad q = -.1 \tag{116}$$

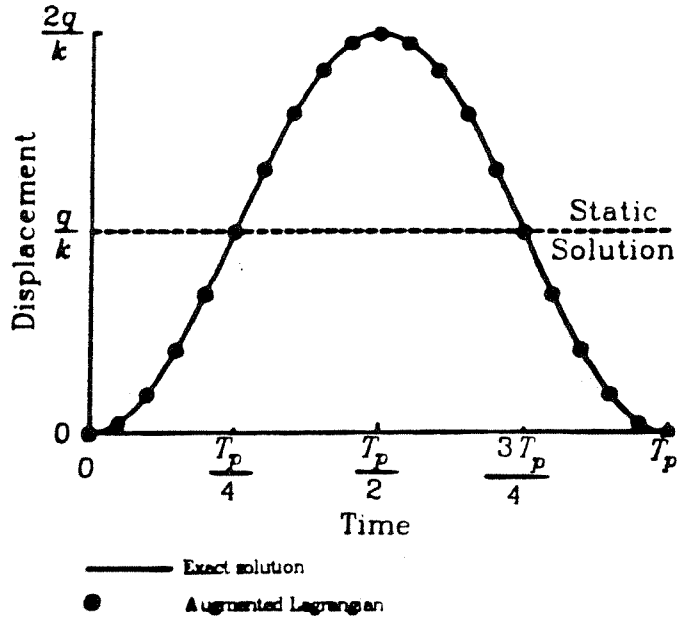

Figure 10: Comparison of exact and augmented Lagrangian results.

4. Examples

This section contains examples of the augmented Lagrangian solution method for both the static and dynamic cases. The examples presented here are compared to work done by previous researchers. In particular, the augmented Lagrangian method is compared to the results of references [6] and [7] for the perturbed Lagrangian approach, reference [15] for the dynamic case and reference [14] for the Hertz solutions.

The augmented Lagrangian method was applied to the system of a rigid punch into an elastic foundation, illustrated in Figure 11. The finite element mesh is successively refined in order to get an idea on the convergence characteristics of the system. Using information from reference [7], the following parameters were used:

$$\mathbf{E}_{punch} = 10^8 \quad \nu_{punch} = 0.0 \quad \mathbf{E}_{foundation} = 10^5 \quad \nu_{foundation} = 0.3 \qquad (117.a)$$

$$\mathbf{L}_{punch} = 20 \quad \mathbf{L}_{foundation} = 58 \quad \mathbf{D}_{foundation} = 22 \qquad (117.b)$$

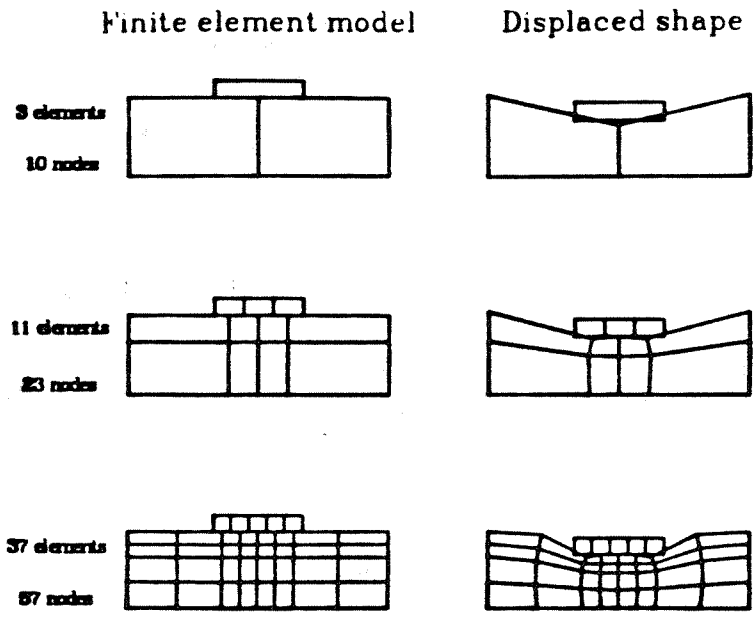

Figure 11: Rigid punch into an elastic foundation.

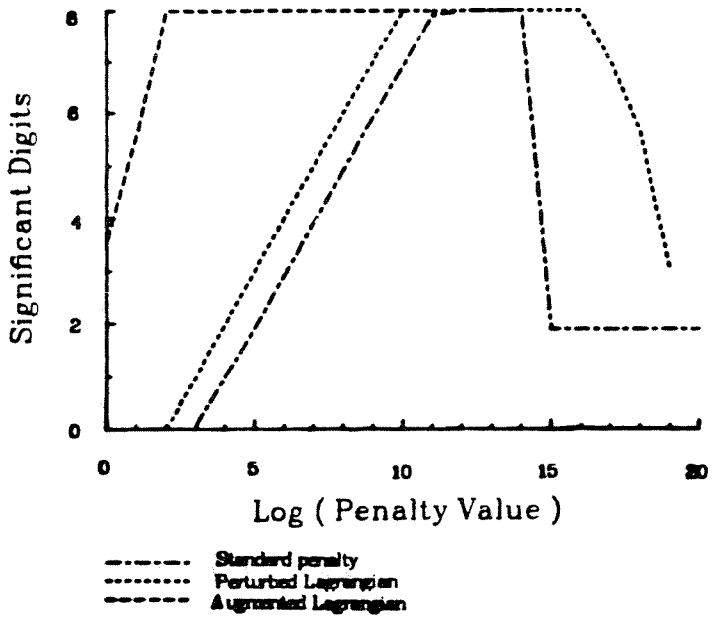

Figure 12: Results after 1 iteration for the 3 element mesh.

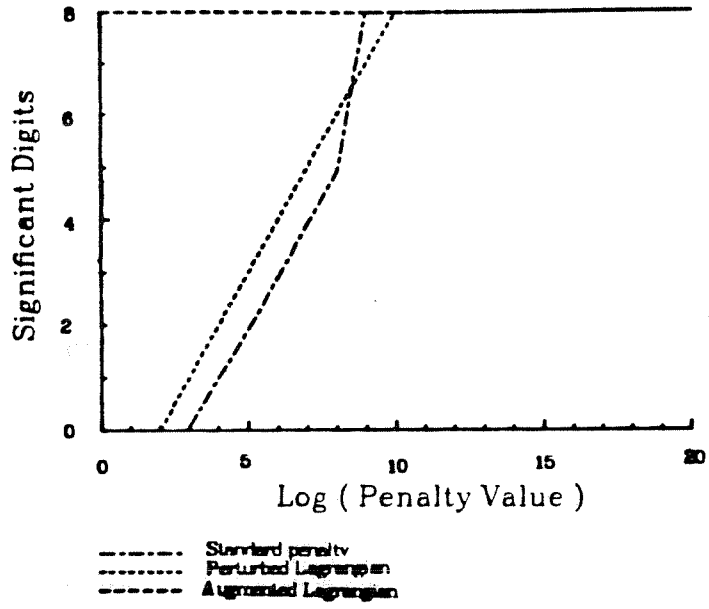

Figure 13: Comparison of converged results for the 3 element mesh.

Figures 12 and 13 display the results of applying the penalty, perturbed Lagrangian and augmented Lagrangian methods to the three block problem. Figure 12 shows the results after 1 iteration of each of the methods. Note that the augmented Lagrangian method gives accurate results over a much wider range of penalty parameter values. The converged results are shown in Figure 13. In this case, the augmented Lagrangian approach always yields the correct results, while the other methods only approximate the solution for smaller penalty values. Table 2 displays the comparison on the number of iterations for the three methods, using a penalty value of $\mu = 10^7$. Observe that even though the solutions are improved over the standard penalty and perturbed Lagrangian methods, the values in Table 2 show that little extra work is required.

Method	Mesh 1	Mesh 2	Mesh 3
penalty	3	6	5
perturbed Lagrangian	2	4	6
augmented Lagrangian	2	4	6

Table 2: Iteration results for the rigid punch problem.

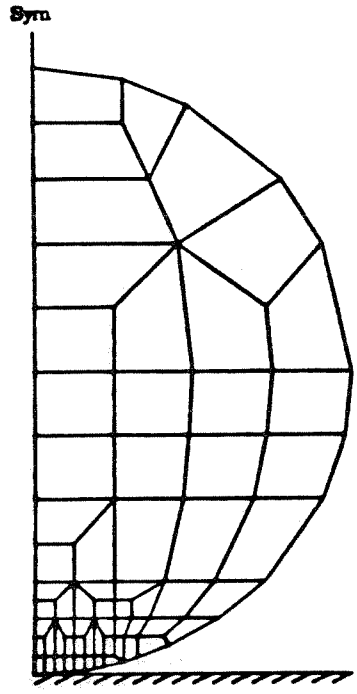

Figure 14: Finite element mesh for the Hertz contact problem.

In order to further illustrate the augmented Lagrangian method, the finite element system in Figure 14 was studied. This mesh represents a sphere in contact with a rigid wall. The parameters for this system are:

$$\mathbf{E} = 10^8 \quad \nu = 0.3 \quad \rho = 0.01 \quad \mathbf{R} = 8.0 \tag{118}$$

The study of a static analysis involved three different load values applied to the top of the sphere. Results of the loading, compared to the classical Hertz solution are illustrated in Figure 15. An iteration count comparing the three methods can be found in Table 3. Each of these method used a penalty value of $\mu = 10^7$. A dynamic analysis was also carried out, assuming that the sphere was given an initial velocity of 0.1 downward. The results displayed in Figure 16 represent the maximum impact values. In this dynamic analysis, only the augmented Lagrangian method obtained reasonable results.

Method	iterations
penalty	6
perturbed Lagrangian	5
augmented Lagrangian	6

Table 3: Iteration results for the Hertz contact problem.

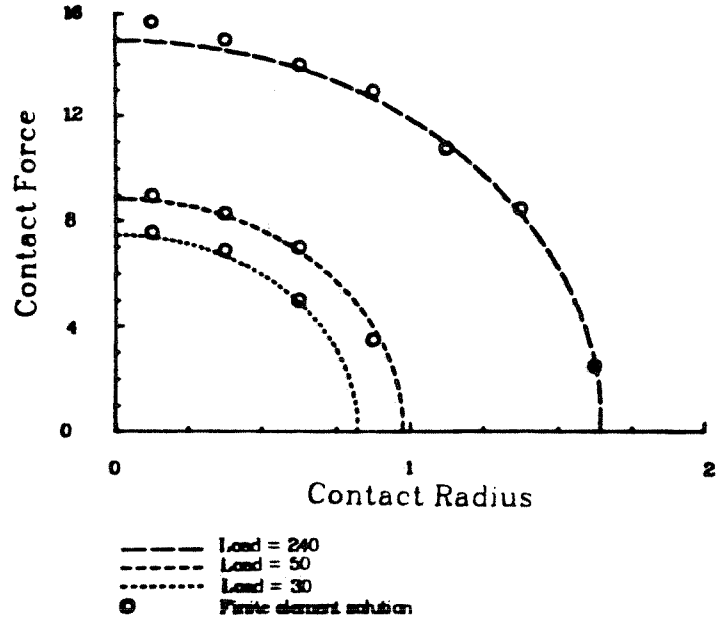

Figure 15: Static results for the Hertz contact problem.

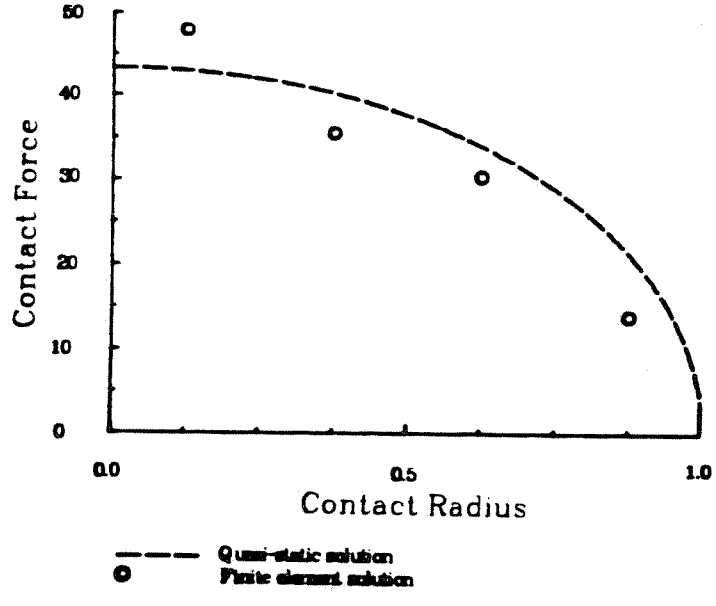

Figure 16: Dynamic results for the Hertz contact problem.

5. Recommendations for further work

This section contains a list of possible extensions to the solution method and its associated algorithms. In particular, proposals are made which would improve the rate of convergence of the algorithm, make the slide line computations more efficient and extend the algorithm to cover new problem domains.

Using equation (93), a second order updating scheme could be added to the augmented Lagrangian method. By comparing equations (83) and (84), the following scheme could be developed.

$$-\left[\nabla^2 d(\lambda_k)\right]^{-1} = \nabla^2 p(\mathbf{u}) + \mu \mathbf{I} \tag{119}$$

Thus, the iteration corresponding to equation (93) is obviously:

$$\lambda_{k+1} = \lambda_k + \left[\nabla^2 p(\mathbf{u}) + \mu \mathbf{I}\right] \nabla d(\lambda_k) \tag{120}$$

As the computation proceeds, values are found for $p(\mathbf{u})$ and $\nabla p(\mathbf{u}) = \lambda_k + \mu \mathbf{u}$. These values could then be used to generate an approximation to $\nabla^2 p(\mathbf{u})$. The need to use second derivatives is eliminated entirely if one chooses to solve the contact problem by some quasi-Newton methods such as the DFP or BFGS. See reference [4] for more details on these methods.

A further enhancement might be to extend the slide line contact logic. As was mentioned in Section 3, the logic is not well defined when multiple segments are in contact with a single surface. For four node isoparametric elements, the extension is straightforward. However, an extension to cover three dimensional problems would entail the use of a clipping algorithm to determine if segments were in contact. Essentially, the algorithm is cubic in the number of nodes for this case, but if linear elements are used, a proper implementation could reduce the required computation significantly. The extension to cover higher order elements is less clear. It is believed, however, that by using techniques from constructive solid geometry and exploiting the convex hull properties of the finite element mesh, a reasonably efficient method could be developed.

A third area which needs further work is the development of algorithms which include friction on the contact surface. In particular, methods which develop consistent tangent matrices need to be devised. The addition of frictional contact to the algorithms presented here would greatly enhance their generality.

Finally, the contact algorithms should not be limited to small deformation kinematics. All of the solution efforts that have been described here need to be evaluated for situations in which motions on the contact surface are finite.

6. Summary and Conclusions

An augmented Lagrangian method for the solution finite element contact problems offers several advantages over the standard penalty, classical Lagrangian or perturbed Lagrangian methods. By using the augmented Lagrangian method, the user has direct control over the accuracy of a given problem, since it is the number of iterations that control the tolerance of the solutions. The value of the penalty parameter controls the rate of convergence. Expressions have been presented, based on error considerations, which may be used as a guide to selecting a reasonable penalty parameter.

The algorithm for the augmented Lagrangian method can be simplified to incorporate both the standard penalty and perturbed Lagrangian Methods. Furthermore, this method is applicable to both static and dynamic problems. The results presented in section 4 show that the method compares very well to work done by previous researchers. Convergence is quite fast, even in cases where the other methods fail.

References

1 S. K. Chan and I. S. Tuba, "A Finite Element Method for Contact Problems of Solid Bodies: I. Theory and Validation," *Int. J. Mech. Sci.*, Vol. **13**, pp. 627-639, 1971.

2 David G. Luenberger, *Linear and Nonlinear Programming*, Second Edition, Addison-Wesley, Reading, Mass., 1984.

3 Dimitri P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

4 Peter Wriggers and Bahram Nour-Omid, "Solution Methods for Contact Problems," *Report UCB/SESM 84/09*, Dept. of Civil Engng., Univ. Calif., Berkeley, 1984.

5 Carlos A. Felippa, "Error Analysis of Penalty Function Techniques for Constraint Definition in Linear Algebraic Systems," *Int. J. Num. Meth. Engng.*, Vol. **11**, pp 709-728, 1977.

6 Juan C. Simo, Peter Wriggers and Robert L. Taylor, "A Perturbed Lagrangian Formulation for the Finite Element Solution of Contact Problems," *Report UCB/SESM 84/14*, Dept. of Civil Engng., Univ. Calif., Berkeley, 1984.

7 Peter Wriggers, Juan C. Simo and Robert L. Taylor, "Penalty and Augmented Lagrangian Formulations for Contact Problems," *Proceeding of the NUMETA '85 Conference*, pp 97-106, 1985.

8 Carlos A. Felippa, "Iterative Procedures for Improving Penalty Function Solutions of Algebraic Systems," *Int. J. Num. Meth. Engng., Vol.* **12**, *pp. 821-836, 1978.*

9 O. C. Zienkiewicz, *The Finite Element Method*, 3rd Edition, McGraw-Hill, London, 1977.

10 J. D. Foley and A. Van Dam, *Fundamentals of Interactive Computer Graphics*, Addison-Wesley, Reading Mass., 1982.

11 David F. Rogers and J. Alan Adams, *Mathematical Elements for Computer Graphics*, McGraw-Hill, New York, 1976.

12 Donald A. Pierre and Michael J. Lowe, *Mathematical Programming via Augmented Lagrangians*, Addison-Wesley, Reading Mass., 1975.

13 Michel Fortin and Roland Glowinski, *Augmented Lagrangian Methods: Applications to the Solution of Boundary-value Problems*, North-Holland, Amsterdam, 1983.

14 Werner Goldsmith, *Impact: The Theory and Physical Behavior of Colliding Solids*, Edward Arnold, London, 1960.

15 Thomas J. R. Hughes, Robert L. Taylor, Jerome L. Sackman, Alain Curnier and Worsak Kanoknukulchai, "A Finite Element Method for a Class of Contact-Impact Problems",

Comp. Meth. Appl. Mech. Engng., Vol. **8**, pp. 249-276, 1976.

16 S. Valliappan, I. K. Lee and P. Boonlualohr, "Non-linear Analysis of Contact Problems," in ***Numerical Methods in Coupled Systems***, R. W. Lewis, P. Bettess, E. Hinton editors, John Wiley, New York, 1984.

17 G. F. Carey and J. T. Oden, ***Finite Elements, Volume II***, Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

18 J. T. Oden, "Exterior Penalty Methods for Contact Problems in Elasticity," in ***Europe-US Workshop: Nonlinear Finite Element Analysis in Structural Mechanics***, (ed., Bathe, Stein and Wunderlich), Springer-Verlag, Berlin, 1980.

19 K. J. Bathe, ***Finite Element Procedures in Engineering Analysis***, Prentice-Hall, Englewood Cliffs, New Jersey, 1982.

20 G. Strang, ***Linear Algebra and Its Applications***, Academic Press, New York, 1980.

21 A. Francavilla and O. C. Zienkiewicz, "A note on the Numerical Computation of Elastic Contact Problems," ***Int. J. Num. Meth. Engng.***, Vol. **9**, pp 913-924, 1975.

22 N. Kirkuchi and J. T. Oden, "Contact Problems in Elastostatics," in ***Finite Elements: Special Problems in Solid Mechanics, Vol. IV***, Oden and Carey editors, Prentice Hall, Englewood Cliffs, New Jersey, 1984.