INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING
Int. J. Numer. Meth. Engng 2001; **51**:865–882

A geometrical-based contact algorithm using a barrier method

G. Kloosterman[1,*,†], R. M. J. van Damme[2], A. H. van den Boogaard[3] and J. Huétink[3]

[1] *Department of Mechanical Engineering, Netherlands Institute for Metals Research, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*
[2] *Department of Mathematical Sciences, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*
[3] *Department of Mechanical Engineering, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*

SUMMARY

Most methods employed in the numerical solution of contact problems in finite element simulations rely on equality-based optimization methods. Typically, a gap function which is non-differentiable at the point of contact is used in these kind of approaches. The gap function can be seen as the Macaulay bracket of some distance function, where the latter is differentiable at the point of contact. In this article, we propose to use the distance function directly instead of using the gap function. This will give rise to a formulation involving inequality constraints. This approach eliminates the artificially introduced non-differentiability. To this end we propose a barrier algorithm as the method of choice to solve the problem. The method originates in optimization literature, where convergence proofs for the method are available. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: barrier method; frictionless contact; augmentation; large deformations; finite elements

1. INTRODUCTION

In order to be able to model the motion of several bodies in space using a finite element discretisation, a contact model is required. Simple as this statement sounds, contact models are among the hardest non-linear problems that are encountered in numerical mechanics.

Part of the complexity arises from the fact that we have to deal with variational inequalities, as demonstrated in Reference [1]. In this article we will limit ourselves by considering the elastic deformation problem only. In this case, we can view the problem as an inequality constrained optimisation problem [2]. The usual approach is to reformulate the problem as an equality constrained problem, and then choose a suitable method from the optimisation literature to solve it. The most simple approach is to use a penalty method as it is employed

*Correspondence to: G. Kloosterman, Department of Mechanical Engineering, Netherlands Institute for Metals Research, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
†E-mail: G.Kloosterman@wb.utwente.nl

Received 3 February 2000
Revised 29 November 2000

in the work of Papadopoulos *et al.* [3] and Chenot and Fourment [4], with the mathematical background discussed in a well-written book by Luenberger [5].

The advantage of penalty methods is the conceptual simplicity they offer to solve the problem. A major disadvantage is though that they are only exact in the limiting case, i.e. when the penalty parameter tends to infinity. However, increasing the penalty parameter leads to progressive ill-conditioning of the resulting Newton system, and thus high-accuracy solutions cannot hoped to be achieved with this approach.

A well-known method to overcome these problems for equality constrained problems is the method of augmented Lagrangians. For an overview on these type of problems we refer to the monograph of Bertsekas [6]. Applications to contact and friction algorithms can be found in the works of Laursen and Simo [7, 8] and Zavarise *et al.* [9, 10].

Though augmented Lagrangian methods based on equality constraints offer a more stable and accurate solution process, they still attempt to solve a problem posed under non-differentiable constraints. This can result in numerical difficulties using a Newton-based approach. Despite advances in methods employing Non-smooth Newton's method, as in the work of Qi and Sun [11], or using Bouligand-differentiable Newton method as employed by Pang [12], it is our opinion that it is desirable to start from the inequality directly. This approach will prevent the introduction of the non-differentiability caused by the Macaulay bracket. The necessity of this approach was already recognized in Reference [13]. Methods that employ these kind of strategies are usually barrier methods, crudely mentioned in Reference [5], be it only the classical barrier formulation, whereas we will discuss a more general class of barrier functions. Possible selections for the barrier function are given in Reference [2].

A penalty barrier approach, with a proof of global convergence under mild assumptions is given in the work of Breitfeld and Shanno [14, 15]. The approach was slightly modified by Franz *et al.* [16], which is the method we have employed. Further research can be found in the work of Conn *et al.* [17].

In this article, we consider deformation processes determined through minimisation of a potential, which are typically elastic problems. We will be considering a neo-Hookean geometrically nonlinear elastic model as one possible choice of this potential. For elastic/plastic problems the method will work as well, but in that case the method needs to be introduced in the variational setting. This would complicate the introduction of the algorithm unnecessarily. The generalization can however be made equivalently to the generalization of the Lagrange multiplier scheme.

The organization of this article is as follows: In Section 2, we define the geometric contact problem and present the inequality formulation. In Section 3, we introduce the barrier algorithm and show how it can be applied to the contact problem. In Section 4, some two-dimensional results are presented. Concluding remarks are given in Section 5.

2. THE GEOMETRIC CONTACT PROBLEM

This section will be spent on the derivation of the usual contact problem as an inequality constrained optimization problem. The approach which is usually taken, is to convert this inequality formulation into an equality formulation. This however introduces a non-differentiability into the formulation, which can cause unstable behaviour. Therefore, we retain the inequality

formulation and instead focus on a new type of algorithm for use in contact simulations in the following sections.

First in Section 2.1, we present the continuum formulation for the problem. The formulation we derive is based on hyper-elastic materials. However, this is only to simplify the discussion. The approach is in fact valid for any motion that can be described using a potential function.

After deriving the continuum formulation, it is discretized. This discretization is necessary to obtain a formulation that can effectively be solved using optimization-based techniques. The discretization is given in Section 2.2. To give a short comparison with the usual approach, we discuss in Section 2.3 how the equality formulation is obtained.

2.1. The continuum formulation

We consider two bodies, denoted by Ω^A and Ω^B, which are open subsets of $\mathbb{R}^2$ or $\mathbb{R}^3$. Their boundaries are denoted as $\Gamma^A = \partial\Omega^A$ and $\Gamma^B = \partial\Omega^B$. The union of both bodies is given by $\Omega = \Omega^A \cup \Omega^B$. A material point in the reference configuration of either body is denoted by $\mathbf{X} \in \Omega$. A point in the current configuration is given by $\mathbf{x} = \boldsymbol{\varphi}(\mathbf{X}, t)$, where the inverse Jacobian of $\boldsymbol{\varphi}$ is assumed to exist for all t, where t denotes time. $\boldsymbol{\varphi}$ is interpreted as the deformation of the bodies Ω_A and Ω_B.

As was already mentioned, we will be considering hyper-elastic materials only. In this case, the constitutive equation for the material is defined in terms of a stored-energy function W (also called the strain energy density). This energy function depends on the deformation function $\boldsymbol{\varphi}$ through the local deformation gradient $\mathbf{F}$ only, i.e. $W(\mathbf{X}, \mathbf{F})$. For the local deformation gradient we have

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \nabla \boldsymbol{\varphi}(\mathbf{X}, t) \tag{1}$$

where the gradient operator is in terms of the spatial co-ordinates only. Integrating the strain energy density over both bodies gives us the total energy at a given time. This gives rise to the energy functional $U(\boldsymbol{\varphi}, t)$:

$$U(\boldsymbol{\varphi}, t) = \int_{\Omega} W(\mathbf{X}, \nabla \boldsymbol{\varphi}(\mathbf{X}, t))\, \mathrm{d}\Omega \tag{2}$$

Apart from the restriction that $\boldsymbol{\varphi}$ be a bijective map, there is an additional constraint: The bodies Ω_A and Ω_B cannot simultaneously occupy the same region in space. This condition is commonly referred to as the impenetrability constraint.

To incorporate the impenetrability constraint into our contact formulation let us introduce a signed distance functional $d_B(\boldsymbol{\varphi}, \mathbf{X}, t)$. This functional expresses the signed distance from a point $\mathbf{X} \in \Omega$ to the boundary Γ_B of Ω_B for the given deformation $\boldsymbol{\varphi}$ at time t.

To simplify notation, let Γ_B^t denote the boundary Γ_B at time t, thus $\Gamma_B^t = \boldsymbol{\varphi}(\Gamma_B, t)$. Furthermore, let $\mathbf{x} = \boldsymbol{\varphi}(\mathbf{X}, t)$, and let $\bar{\mathbf{x}} \in \Gamma_B^t$ be such that $\|\mathbf{x} - \bar{\mathbf{x}}\| \leqslant \|\mathbf{x} - \mathbf{y}\|$ for any other $\mathbf{y} \in \Gamma_B^t$. The point $\bar{\mathbf{x}}$ can be interpreted as the projection of $\mathbf{x}$ on Γ_B^t. Finally, let $\mathbf{n}_B(\mathbf{y})$ denote the outward normal at $\mathbf{y} \in \Gamma_B^t$. An illustration of the concepts is given in Figure 1. The signed distance functional can now be defined as

$$d_B(\boldsymbol{\varphi}, \mathbf{X}, t) = (\mathbf{x} - \bar{\mathbf{x}}) \cdot \mathbf{n}_B(\bar{\mathbf{x}}) \tag{3}$$

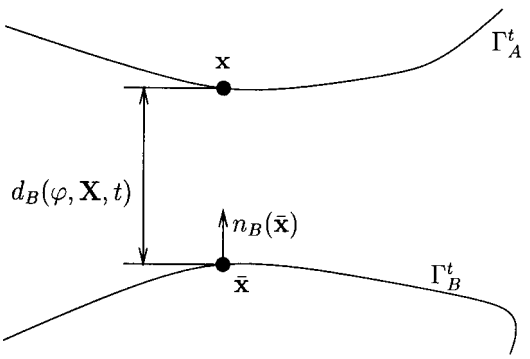

Figure 1. Illustration of the distance function.

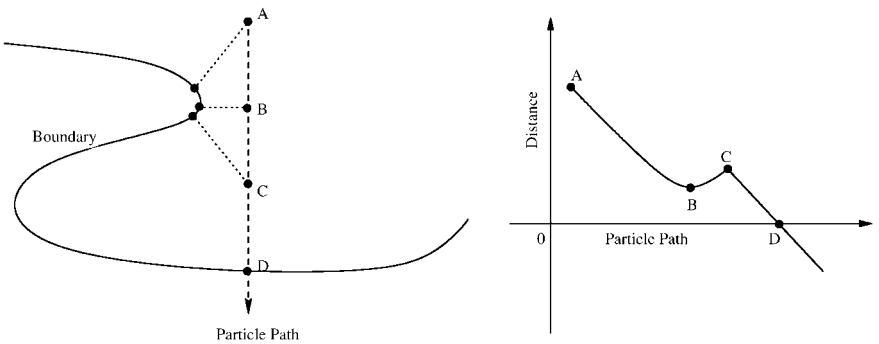

Figure 2. Example of a non-differentiable distance function.

Since $\bar{\mathbf{x}}$ is in fact a projection of $\mathbf{x}$ on Γ_B^t, we know that $\mathbf{x}-\bar{\mathbf{x}} = \delta \cdot \mathbf{n}_B(\bar{\mathbf{x}})$, δ being some number. As a result, we can conclude that if $\mathbf{X}$ lies within Ω_B the functional will take negative values. Whereas if $\mathbf{X}$ lies outside of Ω_B, the functional will take positive values.

We wish to stress that the projection point $\bar{\mathbf{x}}$ may not be unique. If this is the case, then the distance functional is locally non-differentiable. One such possible case is illustrated in Figure 2, where at point C the projection is non-unique. In most cases however, this type of discontinuity is not influential. As long as the boundaries are not too strongly curved with respect to the distance of the point to the boundary. In that case the projection remains well defined.

Using (3) we can write the impenetrability constraint as

$$d_B(\boldsymbol{\varphi}, \mathbf{X}, t) \geqslant 0 \quad \text{for all } \mathbf{X} \in \Omega_A \tag{4}$$

This is quite an elaborate constraint, since all points $\mathbf{X} \in \Omega_A$ must be checked. However, using the fact that the impenetrability constraint must hold for each time t, we can simplify it to

$$d_B(\boldsymbol{\varphi}, \mathbf{X}, t) \geqslant 0 \quad \text{for all } \mathbf{X} \in \Gamma_A \tag{5}$$

We now arrive at the continuum formulation for the contact problem by combining (2) and (5):

$$\begin{aligned} \min_{\boldsymbol{\varphi}} \quad & U(\boldsymbol{\varphi}, t) \\ \text{s.t.} \quad & d_B(\boldsymbol{\varphi}, \mathbf{X}, t) \geqslant 0 \quad \text{for all } \mathbf{X} \in \Gamma_A \end{aligned} \tag{6}$$

The formulation as it is given in (6) did not require an assumption on the sign of the compressive stresses. Nor did we require the well-known complementarity conditions. Instead, we consider the impenetrability constraint as a purely geometrical constraint. This view point gives us the mentioned properties as a result, instead of requiring them as input for obtaining the formulation. It is this formulation that will be used as the basis for our contact calculations.

2.2. *The discretized formulation*

Let V denote the space of functions $\boldsymbol{\varphi}$ which are candidate solutions for (6). The space V is too large to effectively find the minimum $\boldsymbol{\varphi}^*$ of (6). Instead, we limit ourselves to finite-dimensional subspaces V^h of V, which provide approximations to $\boldsymbol{\varphi}^*$. The problem stated in terms of the finite-dimensional subspace V^h of V is called a discretization of (6) and is the topic of this section.

A typical subspace choice for V^h is the finite element subspace, which is the space we will use. A typical member $\boldsymbol{\varphi}^h$ of this space takes the form

$$\boldsymbol{\varphi}^h(\mathbf{X}, t) = \sum_{i=1}^{N_{\text{nodes}}} \mathbf{x}_i(t) N_i(\mathbf{X}, t) \tag{7}$$

where $\mathbf{x}_i(t)$ denote the nodal displacements for node i at time t, N_{nodes} is the number of nodes and N_i is the composition of element shape functions for node i (see Reference [18]).

Since we will be looking at the problem for given points in time t, we drop the time dependency from here onwards. Let us further consider the vector x as defined by stacking the individual $\mathbf{x}_i$:

$$x = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_{\text{nodes}}}] \tag{8}$$

It is clear that each function $\boldsymbol{\varphi}^h$ is completely characterised through a given x. As a result the minimum over V^h can be written as

$$\begin{aligned} \min_{x} \quad & E(x) \\ \text{s.t.} \quad & d_i(x) \geqslant 0 \quad i \in I \end{aligned} \tag{9}$$

where I is the set of nodes that potentially come into contact. In the above equation we have written $E(x) = U(\boldsymbol{\varphi}^h)$, where x is the equivalent characterisation of $\boldsymbol{\varphi}^h$. The geometrical constraint $d_i(x)$, is now to be satisfied point wise. The discretization introduces an additional non-differentiability due to the non-smoothness of the boundaries, which is illustrated in Figure 3.

The discontinuity in the derivative is essentially the same as in Figure 2, with the difference that the curvature changes from finite to infinite at a node. Thus, a boundary node can never be near enough for the distance to be small with respect to the radius of curvature. This makes this kind of discontinuity problematic upon contact, whereas the previous one was not.

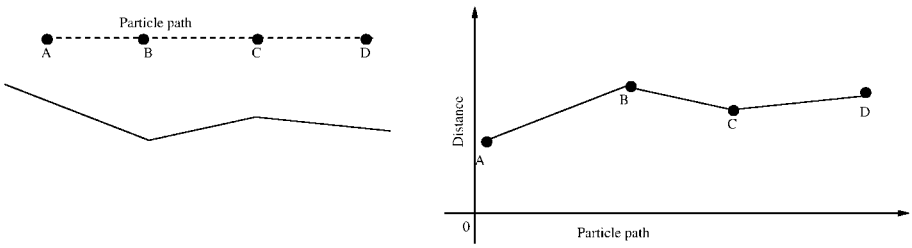

Figure 3. Example of non-differentiability of discrete boundary.

2.3. Solving the discretized formulation

As a next step, the discretized problem (9) needs to be solved. In this section, we discuss what approach is usually taken. This approach consists of converting the inequality problem into an equality problem. We show that this approach introduces yet another non-differentiability. Something that should be avoided to have a stable solution procedure.

The usual approach in literature is to convert the inequality constraint into an equality constraint by using the Macaulay bracket, which is defined as

$$\langle x \rangle = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geqslant 0 \end{cases} \tag{10}$$

Applying the Macaulay bracket to the constraints, converts problem (9) into

$$\begin{aligned} \min_{x} \quad & E(x) \\ \text{s.t.} \quad & \langle -d_i(x) \rangle = 0 \quad i \in I \end{aligned} \tag{11}$$

This formulation can subsequently be solved using a penalty approach, by converting (11) to the following unconstrained form:

$$\min_{x} E(x) + \frac{p}{2} \sum_{i \in I} \langle -d_i(x) \rangle^2 \tag{12}$$

The above formulation is only exact for the penalty p tending to infinity, and is ill conditioned for large penalties. To overcome this problem an augmentation procedure can be used. However, this does not address the fundamental issue that the Macaulay bracket introduces yet another non-differentiability.

The above regularization causes the sometimes observed phenomenon of change of contact state: A node being outside does not perceive any object being there, due to $\langle -d_i(x) \rangle$ being 0. As a result overestimated nodal displacements occur leading to penetration. This in turn induces a large contact force, causing the node to be put outside of the body, restarting the cycle.

Every non-differentiability of the distance function can cause difficulties or even non-convergence of the problem. We have now identified three types:

(1) The trajectory of the node with respect to a smooth boundary changes abruptly its projection point (see Figure 2). This should not pose a real problem if the curvature of the boundary is not too high and we are near contact.

(2) The boundary is non-smooth. This can be dealt with by several techniques. A non-smooth boundary is typical for finite element meshes (see Figure 3).
(3) The inequality constraint has been converted to an equality constraint, causing a problem upon contact.

The non-differentiability caused by the Macaulay bracket is present at every node in contact. The non-differentiability even occurs at the exact point of contact, which is the point of interest. At this point it causes the distance function to be ill behaved. This type of behaviour can be prevented by starting from the inequality directly. A different type of algorithm to deal with inequalities is required, and one such possible algorithm is proposed in the next section.

3. THE BARRIER APPROACH

To solve the contact problem in its inequality formulation, we require an inequality-based algorithm. This is different to the usual approach followed in solving contact problems, which are based on equality-based algorithms. The algorithm we employ is a modified barrier method, which to our knowledge has not yet been used for these type of calculations.

In this section, we use part of the work of Breitfeld and Shanno [14] and of Franz *et al.* [16] on modified barrier methods. They have tested the method on small but complex non-linear optimization problems. Our interest lies in more large-scale non-linear problems.

In Section 3.1 the inequality constrained optimization problem is introduced. The penalty part of the algorithm is omitted, since we will not be dealing with equality constraints. Next, in Section 3.2 the numerical algorithm is presented, which we have used to solve the problem. Slight modifications to the distance function were made to be able to work with the possible change of contacting segment.

3.1. The modified barrier regularization

The problem we will consider is the inequality constrained optimization problem as it appeared in Section 2.2:

$$\begin{aligned} \min_{x} \quad & E(x) \\ \text{s.t.} \quad & d_i(x) \geqslant 0 \quad i \in I \end{aligned} \tag{13}$$

where we assume for the moment that the functions E and d_i are twice continuously differentiable. The cases for which d_i does not satisfy this property are remedied by some stabilisation technique which is discussed later.

We are going to solve the problem stated in (13) by converting it to a sequence of unconstrained problems. The result of each problem is used as a starting point for the next problem. A sequence of estimates to the solution is then hoped to converge to a point x^*, which satisfies the Karush–Kuhn–Tucker conditions:

$$\begin{aligned} \nabla E(x^*) - \sum_i \lambda_i^* \nabla d_i(x^*) &= 0 \\ \lambda_i^* d_i(x^*) &= 0 \\ \lambda_i^* &\geqslant 0 \\ d_i(x^*) &\geqslant 0 \end{aligned} \tag{14}$$

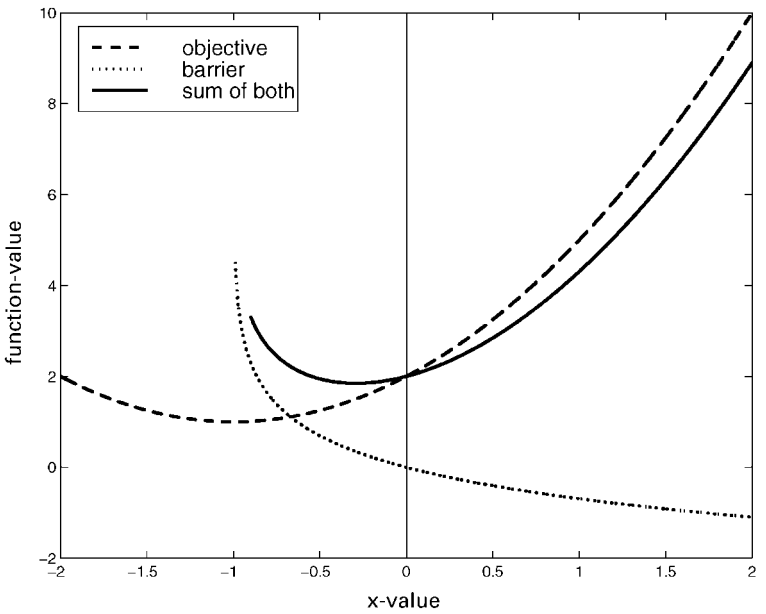

Figure 4. Illustration of the barrier concept.

where λ_i^* is the Lagrange multiplier associated with the constraint d_i. Traditional solution methods can be found in the work of Luenberger [5], Fiacco and McCormick [19] and Bertsekas [6].

Inequality constraints can be handled by converting problem (13) to the form

$$\min_x F(x,\mu,\lambda) = \min_x \left(E(x) - \mu \sum_{i\in I} \lambda_i \log\left(1 + \frac{d_i(x)}{\mu}\right)\right) \tag{15}$$

where $\lambda_i \geqslant 0$ are fixed estimates for the Lagrange multipliers, $\mu > 0$ is the barrier parameter and the log refers to the natural logarithm. The concept is roughly that as d_i approaches $-\mu$, the logarithm adds a significant penalty to the function F, whereas for d_i large, the function offers a negative contribution to the objective, and is as such 'preferable'.

For an illustration of the concepts, consider Figure 4. In it we have plotted as the objective $E(x) = (x+1)^2 + 1$. The constraint here is $x \geqslant 0$. The barrier term that is plotted is $-\lambda\mu \log(1 + x/\mu)$, with $\lambda = \mu = 1$. Notice that the minimum of the sum is an approximation to the correct minimum, which lies at 0. Decreasing μ improves the approximation. Choosing the correct λ will place the minimum of the sum at precisely the point 0.

The approach that is followed, is roughly equivalent to the augmented Lagrangian approach for equality constrained problems. A value for μ is chosen as well as fixed values for the λ_i's. The minimisation problem for F is then solved with these parameters. Next using the obtained point x, we improve our estimate for the Lagrange multipliers, μ is decreased, and we start again. Polyak [20] has shown that under some mild conditions this approach converges linearly to a solution of (13).

To improve the speed of convergence, we scale the constraints, the necessity discussed in Reference [14]. This is achieved by introducing a scaling factor s_i for each of the constraints. The value of the constraint is then scaled to be $d_i(x)/s_i$. Combining this with the unconstrained subproblem as in (15) results in

$$\begin{aligned} F(x,\mu,\lambda,s) &= E(x) - \mu \sum_{i\in I} \lambda_i \log\left(1 + \frac{1}{\mu}\frac{d_i(x)}{s_i}\right) \\ &= E(x) - \mu \sum_{i\in I} \lambda_i \log\left(\frac{1}{s_i}\left(s_i + \frac{d_i(x)}{\mu}\right)\right) \\ &= E(x) - \mu \sum_{i\in I} \lambda_i \log\left(s_i + \frac{d_i(x)}{\mu}\right) + \mu \sum_{i\in I} \lambda_i \log(s_i) \end{aligned} \tag{16}$$

Notice that the last term in (16) contains only terms in μ, λ and s. Thus, if we take the minimum with respect to x for set values of these parameters, the last term does not influence the value of the optimal x. Consequently, minimizing (16) is equal to

$$\min_x F(x,\mu,\lambda,s) = \min_x \left(E(x) - \mu \sum_{i\in I} \lambda_i \log\left(s_i + \frac{d_i(x)}{\mu}\right)\right) \tag{17}$$

An additional limitation that may occur, is that the function $F(x,\mu,\lambda,s)$ is not defined for all possible values for $d_i(x)$. This is due to requiring that $d_i(x) > -s_i\mu$, since otherwise the logarithm is not defined. To overcome this limitation we introduce a quadratic extrapolation for the logarithm, from somewhere in the admissible region. This extrapolation location is defined through the parameter β, which is some relative extrapolation point. If $d_i(x) \geqslant -\beta s_i\mu$, we use the logarithm. If $d_i(x) < -\beta s_i\mu$, we use the quadratic extrapolation. The parameter β should lie within the range $(0,1)$. The problem as it is defined in (17) becomes

$$F(x,\mu,\lambda,s,\beta) = E(x) - \mu \sum_{i\in I} \lambda_i \Phi(d_i(x),\mu,s_i,\beta) \tag{18}$$

In the previous formula Φ is the combination of the logarithmic term and the quadratic extrapolation term. The function does not actually depend on the ith constraint. It is defined as

$$\Phi(d_i(x),\mu,s_i,\beta) = \begin{cases} \log\left(s_i + \dfrac{d_i(x)}{\mu}\right) & \text{if } d_i(x) \geqslant -\beta\mu s_i \\ \frac{1}{2}a_i d_i(x)^2 + b_i d_i(x) + c_i & \text{if } d_i(x) < -\beta\mu s_i \end{cases} \tag{19}$$

The λ_i is initialized to 1. Their correct value is computed through application of an augmentation scheme. Using this augmentation scheme and an updating strategy for the barrier parameter μ, it was shown in Reference [14] that the algorithm is globally convergent to a Karush–Kuhn–Tucker point of (13).

In the next section, we give a complete presentation of the algorithm that is used to solve (13).

3.2. The algorithm

We now present a precise description of the barrier algorithm.

Step 0. **Start:** First we need to initialize all the parameters.

- Choose $x^{(0)} \in \mathbb{R}^N$. This is the starting point for the computation. In an incremental computation, $x^{(0)}$ is best chosen as the result of the previous increment.
- Choose $\tau > 0$, the outer loop termination criterion. This termination criterion is approximately equivalent to the unbalance termination criterion. $1 \cdot 0e^{-3}$ is a reasonable choice.
- Select a sequence of barrier parameters $\{\mu^{(k)}\}_{k \geqslant 0}$, which is descending. We start from $0 \cdot 1$. In Reference [14] $\mu^{(k+1)}$ is chosen as $0 \cdot 1\mu^{(k)}$.
- Select a sequence $\{\varepsilon^{(k)}\}_{k \geqslant 0}$, the inner loop termination criterion. A possible selection is to choose this sequence equal to the μ sequence.
- Select $\lambda_i^{(0)} = 1$, for $i \in I$.
- Choose $0 \leqslant \beta \leqslant \beta_u < 1$, the relative extrapolation point. We took $\beta = 0 \cdot 9$.
- Choose the scaling terms $s_i^{(0)} = \min\{\max\{1, -d_i(x^{(0)})\}, s_u\}$ for $i \in I$, where s_u is some upper bound for the scaling terms. $s_u = 1000$ is a possible choice.
- Compute the extrapolation coefficients for the ith barrier function $a_i^{(0)}, b_i^{(0)}$ and $c_i^{(0)}$ by (30). These coefficients are completely determined by demanding Φ to be twice continuously differentiable.
- Set $k = 0$.

Step 1. **Unconstrained minimization:** In this step we find an approximation $x^{(k+1)}$ of a local minimizer of the current regularized problem:

$$\begin{aligned} &\min_x F\left(x, \mu^{(k)}, \lambda^{(k)}, s^{(k)}, \beta\right) \\ &\quad = \min_x \left(E(x) - \mu^{(k)} \sum_{i \in I} \lambda_i^{(k)} \Phi\left(d_i(x), \mu^{(k)}, s_i^{(k)}, \beta\right) \right) \end{aligned} \tag{20}$$

where the barrier function Φ is defined as

$$\Phi(d_i(x), \mu, s_i, \beta) = \begin{cases} \log\left(s_i + \dfrac{d_i(x)}{\mu}\right) & \text{if } d_i(x) \geqslant -\beta\mu s_i \\ \frac{1}{2}a_i d_i(x)^2 + b_i d_i(x) + c_i & \text{if } d_i(x) < -\beta\mu s_i \end{cases} \tag{21}$$

Solving the unconstrained problem is the inner problem. This is the problem that actually involves taking Newton steps. To solve (20) we consider the first-order condition for it:

$$\nabla_x F\left(x, \mu^{(k)}, \lambda^{(k)}, s^{(k)}, \beta\right) = 0 \tag{22}$$

A. Set $y^{(0)} = x^{(k)}$, and $j = 0$.

B. Check the convergence:

$$w^{(j)} = \frac{\|\nabla_x F\left(y^{(j)}, \mu^{(k)}, \lambda^{(k)}, s^{(k)}, \beta\right)\|}{1 + \|y^{(j)}\|} \tag{23}$$

If $w^{(j)} < \varepsilon^{(k)}$, set $x^{(k+1)} = y^{(j)}$ and stop.

C. Solve the linearization of the first-order condition (22) at $y^{(j)}$:

$$\delta y^{(j)} = -\left[\nabla_x^2 F\left(y^{(j)}, \mu^{(k)}, \lambda^{(k)}, s^{(k)}, \beta\right)\right]^{-1} \nabla_x F\left(y^{(j)}, \mu^{(k)}, \lambda^{(k)}, s^{(k)}, \beta\right) \tag{24}$$

D. Take $y^{(j+1)} = y^{(j)} + \alpha \delta y$, where α is chosen such that F decreases. If we are sufficiently close to the minimum $\alpha = 1$ is a good choice, and corresponds to taking full Newton steps.

E. Return to step B.

Step 2: **Check convergence:** We now test whether the convergence criterion is met. First we define $v_1^{(k+1)}$ and $v_2^{(k+1)}$ according to

$$\begin{aligned} v_1^{(k+1)} &= \frac{\left\|\nabla E\left(x^{(k+1)}\right) - \sum_{i \in I} \lambda_i^{(k)} \nabla d_i\left(x^{(k+1)}\right)\right\|}{1 + \left\|x^{(k+1)}\right\|} \\ v_2^{(k+1)} &= \frac{\left|\sum_{i \in I} \lambda_i^k d_i\left(x^{(k+1)}\right)\right|}{1 + \left\|x^{(k+1)}\right\|} \end{aligned} \tag{25}$$

If $v_j^{(k+1)} < \tau$ for both $j = 1, 2$ then stop. The parameters measure the satisfaction of the Kuhn–Tucker conditions. Instead of these parameters, it is also possible to use the unbalance criterion. When the problem is properly scaled however, these two criteria return comparable results.

Step 3: **Update parameters:** We now update the scaling parameters, by scaling with respect to the new point

$$s_i^{(k+1)} = \min\left\{\max\left\{1, -d_i\left(x^{(k+1)}\right)\right\}, s_u\right\} \tag{26}$$

and the Lagrange multipliers are updated according to

$$\lambda_i^{(k+1)} = \mu^{(k)} \lambda_i^{(k)} \Phi_i'\left(d_i\left(x^{(k+1)}\right), \mu^{(k)}, s_i^{(k)}, \beta\right) \tag{27}$$

where Φ' denotes the derivative of Φ with respect to the first variable only.
The motivation for choosing λ in this manner comes from (22). If we expand this equation we find

$$\nabla E\left(x^{(k+1)}\right) - \sum_{i \in I} \mu^{(k)} \lambda_i^{(k)} \Phi_i'\left(d_i(x^{(k+1)}), \mu^{(k)}, s_i^{(k)}, \beta\right) \nabla d_i\left(x^{(k+1)}\right) = 0 \tag{28}$$

Thus using the solution of (22), we get that selecting $\lambda_i^{(k+1)}$ in the proposed way we have

$$\nabla E\left(x^{(k+1)}\right) - \sum_{i \in I} \lambda_i^{(k+1)} \nabla d_i\left(x^{(k+1)}\right) = 0 \tag{29}$$

In this we can recognize part of the conventional variational form of the equilibrium equations including contact. It is however not completely the same, since the λ_i are held constant.

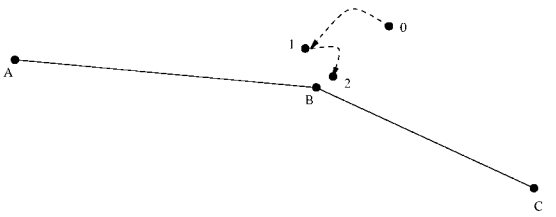

Figure 5. Illustration of the segment update.

Step 4: **Update coefficients:** What remains to be done is to update the extrapolation coefficients used in the barrier functions Φ_i.

$$a_i^{(k)} = \frac{-1}{\left(s_i^{(k)}\mu^{(k)}(1-\beta)\right)^2}$$

$$b_i^{(k)} = \frac{1-2\beta}{s_i^{(k)}\mu^{(k)}(1-\beta)^2} \tag{30}$$

$$c_i^{(k)} = \frac{\beta(2-3\beta)}{2(1-\beta)^2} + \log\left(s_i^{(k)}(1-\beta)\right)$$

These coefficients are fully determined by enforcing Φ to be twice continuously differentiable.

Step 5. **Continuation:** Set $k = k+1$, and return to Step 1.

To achieve fast convergence some fine tuning is required for the barrier sequence μ. Regardless of the fine tuning, Breitfeld and Shanno [14] showed that the algorithm has at least a convergent subsequence for any such sequence. In this proof it was however assumed that functions E and d_i where twice differentiable functions. This is for the discretized finite element method not generally the case.

To overcome this obstacle we employ the following approach:

- In Step 0, we find the nearest boundary segment or facet for the constrained node.
- In Step 1, the active boundary segment is not assumed to change. As a result all functions which are used in Step 1 are twice continuously differentiable. Hence, the Newton process will converge to an optimal point $x^{(k+1)}$.
- In Step 2, we again use the same functions.
- In Step 3, before the update of the parameter s_i, we update the nearest segment, to a possible new nearest segment. This update is not done after $k > k_{\text{update}}$ at which point we assume that the correct contact segment was already found.

The effects of this approach are illustrated in Figure 5. During the first outer iteration, a node starts at location 0. It is then nearest the segment BC. This segment is extended to a line, and the node is constrained not to cross this line. The result of the first outer iteration places the point at location 1. The point is now closest to the segment AB, which is in turn extended to form the constraining line. Apart from changing the contact segment, we also decrease μ. This has as an effect that the geometric error should decrease. Thus, the gap

should become smaller if the point is to come into contact. The result of the second outer iteration, places the point at location 2, where the point is again closer to BC. Even though we change contact segment continuously, we also decrease μ. This decrease drives the node closer to the curve, which as a result should converge to a solution.

In this way we eliminated the non-differentiability of the first and second type as it was given in Section 2.2 in the unconstrained optimization loop, where it might hinder the convergence of the Newton process. Non-differentiabilities of the third type have been avoided by using an inequality-based approach.

As a last remark we note that in the intermediate iterations the constraints may be violated. However, the approximate solution is usually strictly feasible.

4. NUMERICAL EXAMPLES

In this section some results will be given for the method as it was discussed in the previous sections. We have tested the procedure on two potentially difficult contact problems:

- A corner contact problem, which is difficult due to multiple possible contacting segments.
- A wedge problem, which is difficult when highly inclined angles are used.

Both of the above examples consist of two elastically deforming bodies in contact. We are expecting large strains, thus we have modelled the elastic behaviour with a neo-Hookean material model. This type of material model is furthermore objective, and this ensures that we can model the large deformation and/or rotational effects which may appear in the contact calculation.

The energy density function W, by which the neo-Hookean model we use is described, is given by

$$W(\mathbf{X}, \boldsymbol{\varphi}) = \kappa[(J(\mathbf{X}, \boldsymbol{\varphi}) - 1) - \log J(\mathbf{X}, \boldsymbol{\varphi})] + \frac{G}{2}[\mathrm{tr}(\bar{\mathbf{B}}(\mathbf{X}, \varphi)) - 3] \tag{31}$$

where κ and G are the Lamé parameters, J is the determinant of the deformation gradient and $\bar{\mathbf{B}}$ is the isochoric left Cauchy–Green stretch tensor, more explicitly.

More background on this type of hyper-elastic formulations can be found in a book by Simo and Hughes [21]. The choice of the form of the volumetric part and deviatoric part of the problem are dependent on the material one attempts to model. Instead of the traditional use in rubber elasticity in an incompressible formulation, we have chosen to add a compressibile term to the formulation.

We first define the deformation gradient

$$\mathbf{F}(\mathbf{X}, \boldsymbol{\varphi}) = \nabla \boldsymbol{\varphi}(\mathbf{X}, t_0) \tag{32}$$

where we differentiate with respect to the spatial co-ordinates in $\boldsymbol{\varphi}$ only. The determinant of F is given by

$$J(\mathbf{X}, \boldsymbol{\varphi}) = \det(\mathbf{F}(\mathbf{X}, \boldsymbol{\varphi})) \tag{33}$$

The left Cauchy–Green stretch tensor is defined as

$$\mathbf{B}(\mathbf{X}, \boldsymbol{\varphi}) = \mathbf{F}(\mathbf{X}, \boldsymbol{\varphi}) \cdot \mathbf{F}(\mathbf{X}, \boldsymbol{\varphi})^{\mathrm{T}} \tag{34}$$

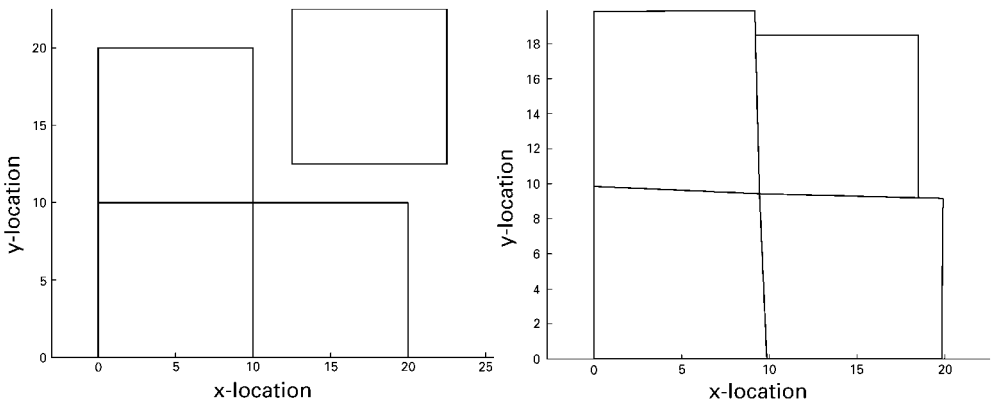

Figure 6. The corner problem.

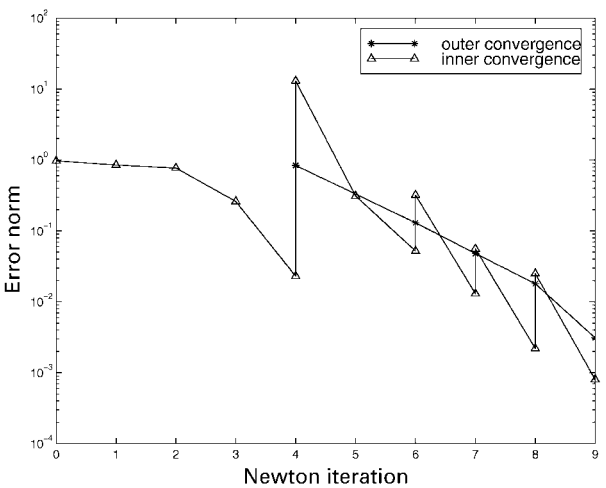

Figure 7. Result after displacement of the corner problem.

The isochoric left Cauchy–Green stretch tensor is

$$\bar{\mathbf{B}}(\mathbf{X},\boldsymbol{\varphi}) = J(\mathbf{X},\boldsymbol{\varphi})^{-2/3}\mathbf{B}(\mathbf{X},\boldsymbol{\varphi}) \tag{35}$$

The total energy is given by integrating over the total domain, i.e.

$$U(\boldsymbol{\varphi}) = \int_{\Omega} W(\mathbf{X},\boldsymbol{\varphi})\,\mathrm{d}\Omega \tag{36}$$

In Section 4.1, we present the corner contact problem, and in Section 4.2 we discuss the wedge problem. The results are obtained using a plane strain state. A support is assumed by prescribing zero displacements on the nodes, thus they are fixed. The material constants are defined by giving definitions for the modulus of elasticity E and poisson ratio v. From this,

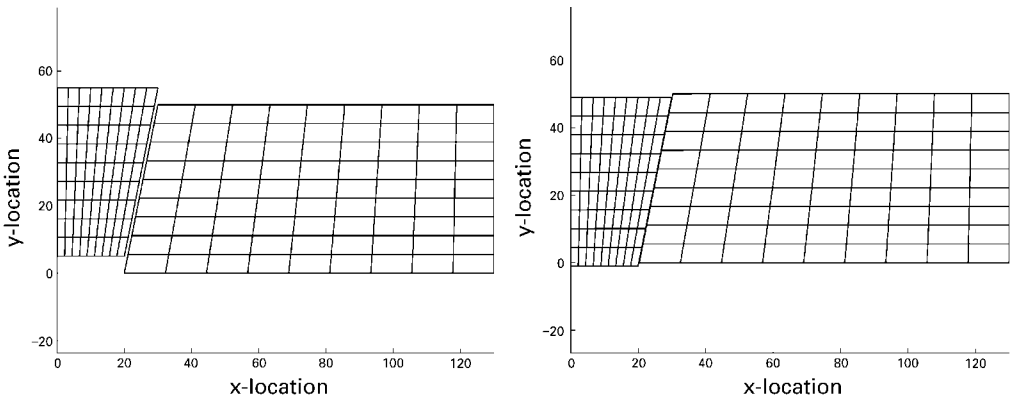

Figure 8. The wedge problem.

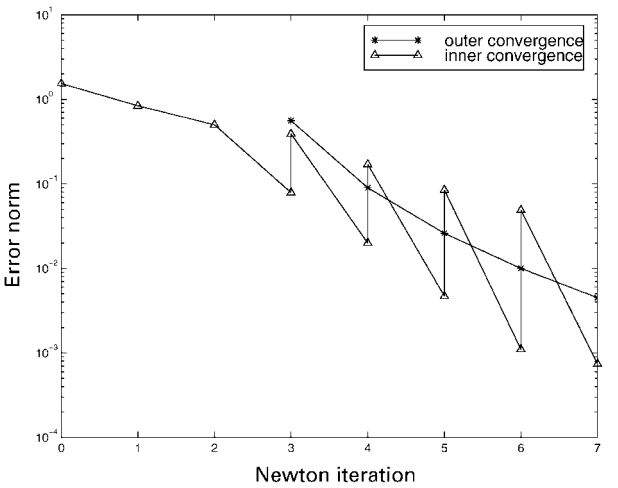

Figure 9. Pushing in of the wedge, 1 and 2 mm.

we obtain κ and G, by the formulae

$$\kappa = \frac{vE}{(1+v)(1-2v)}$$
$$G = \frac{E}{2(1+v)} \tag{37}$$

In fact, since the potential is linear in E, we can scale by its magnitude. By doing that we also scale the Lagrange multipliers. Thus, the result of the problem is independent of E. It is however dependent on the value of v, which we have chosen as 0.31.

Table I. Convergence results for wedge problem.

k	j	$w^{(j)}$	$v^{(k)}$
1	1	3.81e−01	
	2	2.79e−01	
	3	4.14e−02	
	4	4.41e−03	
	5	1.86e−05	
	6	5.52e−10	
	7	4.46e−14	4.36e−2
2	1	5.38e−03	
	2	2.29e−05	
	3	2.26e−10	
	4	4.93e−14	5.81e−3
3	1	7.76e−04	
	2	6.38e−07	
	3	9.61e−13	1.20e−3
4	1	3.03e−04	
	2	4.07e−07	
	3	3.56e−13	2.93e−4
5	1	1.71e−04	
	2	3.08e−07	
	3	9.21e−13	6.91e−5
6	1	1.50e−04	
	2	1.33e−06	
	3	1.16e−10	
	4	5.14e−14	1.26e−5
7	1	9.92e−04	
	2	1.84e−04	
	3	4.61e−06	
	4	3.04e−09	
	5	5.42e−14	1.21e−6

4.1. The corner contact problem

In this section, we consider the problem of a block of material being pushed into a corner. The mesh is very course, see also Reference [22]. The problem is considered difficult, since it is hard to decide for an algorithm what the correct contact segments are.

The setup of the problem is shown on the left in Figure 6. The foundation is assumed to deform elastically and is supported on the bottom and on the left. The block in the upper-right corner is then in one increment pressed into the corner, by prescribing downward displacements on the top and leftward displacements on the right nodes. The result of the prescribed displacements can be seen on the right in Figure 6.

The convergence results are plotted in Figure 7, where the inner convergence shows the progress within each subproblem loop, i.e. for set values of μ and λ. The outer convergence shows the progress of the total convergence. Just as with augmented Lagrangian schemes, the total scheme converges linearly. Each inner problem converges quadratically, when sufficiently close to the optimum for that subproblem. The zigzag pattern is due to fact that when we update the Lagrange multipliers and the value of μ, the error in our initial estimate increases.

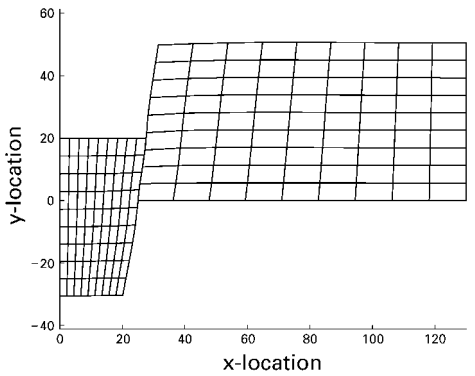

Figure 10. Wedge problem after a displacement of 30 mm.

4.2. The wedge problem

In this section, we consider the problem of a wedge being driven downwards into a cavity. It is an example of an elastic body which is pressed against another elastic body under a very steep angle. The situation is symmetric, and therefore we only show half the problem. In Figure 8 the setup is shown from the side. On the left we have the small block, and on the right we have an elastic body which is supported on the bottom and on the right-hand side.

The small wedge is now pushed into the hole by prescribing displacements on the top nodes. The result of which can be seen on the right in Figure 8.

We only show the second increment, further pressing down of the wedge does not alter the convergence behaviour of the algorithm. To check convergence, we scale the objective and the constraints. It turns out that the relative unbalance criterion is in this case of approximately the same magnitude as the criterion which we brought forward in the previous section. The convergence of this error norm is plotted in Figure 9 versus the number of Newton iterations. From iteration 3 onwards we can update the Lagrange multipliers at every Newton step taken. The line drawn through the stars shows linear convergence, as was predicted in Reference [20]. We first need several Newton iterations to obtain a good prediction, after which each step shows quadratic convergence. The method thus turns out to be quite stable and accurate, given the amount of iterations required. The numbers are shown in Table I. To generate the numbers for this table we let the inner loop finish up to full accuracy. This of course is totally unnecessary, since the error in the outer loop is dominating the one in the inner loop. The quadratic convergence of the inner loop and the linear convergence of the outer loop are evident from this table. In the table k denotes the outer iteration number and j the inner iteration number for that outer iteration. The column with $w^{(j)}$ are as in (23). The outer convergence as in (25) is given in the column labelled $v^{(k)}$.

For an illustration of a result further in the process, consider Figure 10. In this figure the result is displayed after prescribing a downward displacement of 30mm on the top of the wedge.

5. CONCLUSIONS

A contact algorithm based on solving an inequality constrained optimization problem was presented. The modified barrier algorithm presented can be regarded as an augmented

Lagrangian-like procedure, but now involving inequality constraints directly. Working from the inequality prevents introduction of non-differentiabilities that can radically slow down, or even destroy the convergence of finite element computations.

We have given two examples, which although of an academic nature are good test cases. Firstly, we allow both bodies to deform, and secondly the contact state of both bodies can be considered difficult. In both cases, the algorithm required quite a low number of Newton steps to converge per increment. Our increments were relatively large (20 per cent of the element size).

The method is backed up by convergence proofs for sufficiently smooth problems. Summarizing, the modified barrier approach offers an viable alternative to penalty and augmented Lagrangian procedures.

REFERENCES

1. Kikuchi N, Oden JT. *Contact Problems in Elasticity*: *A Study of Variational Inequalities and Finite Element Methods*. SIAM: Philadelphia, PA, 1988.
2. Christensen PW, Klarbring A, Pang JS, Strömberg N. Formulation and comparison of algorithms for frictional contact problems. *International Journal for Numerical Methods in Engineering* 1998; **42**:145–173.
3. Papadopoulos P, Jones RE, Solberg JM. A novel finite element formulation for frictionless contact problems. *International Journal for Numerical Methods in Engineering* 1995; **38**:2603–2617.
4. Chenot JL, Fourment L, Mocellin K. Numerical formulations and algorithms for solving contact problems in metal forming simulation. *International Journal for Numerical Methods in Engineering* 1999; **46**:1435–1462.
5. Luenberger DG. *Introduction to Linear and Nonlinear Programming*. Addison-Wesley: Reading, MA, 1973.
6. Bertsekas DP. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press: New York, 1982.
7. Laursen TA, Simo JC. Algorithmic symmetrization of coulomb frictional problems using augmented lagrangians. *Computer Methods in Applied Mechanics and Engineering* 1992; **108**:133–146.
8. Laursen TA, Simo JC. A continuum-based finite element formulation for the implicit solution of multibody large deformation frictional contact problems. *International Journal for Numerical Methods in Engineering* 1993; **36**:3451–3485.
9. Zavarise G, Wriggers P, Schrefler BA. On augmented lagrangian algorithms for thermomechanical contact problems with friction. *International Journal for Numerical Methods in Engineering* 1995; **38**:2929–2949.
10. Zavarise G, Wriggers P. A superlinear convergent augmented lagrangian procedure for contact problems. *Engineering Computations* 1999; **16**:88–119.
11. Qi L, Sun J. A nonsmooth version of newton's method. *Mathematical Programming* 1993; **58**:353–367.
12. Pang JS. Newton's method for *b*-differentiable equations. *Mathematics of Operations Research* 1990; **15**: 311–341.
13. Zavarise G, Wriggers P, Schrefler BA. A method for solving contact problems. *International Journal for Numerical Methods in Engineering* 1998; **42**:473–498.
14. Breitfeld MG, Shanno DF. A globally convergent penalty-barrier algorithm for nonlinear programming and its computational performance. *Technical Report RRR* 12-94, RUTCOR, April 1994.
15. Breitfeld MG, Shanno DF. Computational experience with penalty-barrier methods for nonlinear programming. *Annals of Operations Research* 1996; **62**:439–464.
16. Franz J, Liepelt M, Schittkowski K. Penalty-barrier-methods for nonlinear optimization: implementation and computational results. *Technical Report*, Department of Mathematics, University of Bayreuth, Germany, 1995.
17. Conn AR, Gould N, Toint PhL. A globally convergent lagrangian barrier algorithm for optimisation with general inequality constraints and simple bounds. *Mathematics of Computations* 1997; **66**:261–288.
18. Hughes TJR. *The Finite Element Method*: *Linear Static and Dynamic Analysis*. Prentice-Hall: Englewood Cliffs, NJ, 1987.
19. Fiacco AV, McCormick GP. *Nonlinear Programming*: *Sequential Unconstrained Minimization Techniques*. Wiley: New York, 1968.
20. Polyak R. Modified barrier functions (theory and methods). *Mathematical Programming* 1992; **54**:177–222.
21. Simo JC, Hughes TJR. *Computational Inelasticity*. Springer: New York, 1998.
22. Schreppers GJMA, Brekelmans WAM, Sauren AAHJ. A finite element formulation of the large sliding contact. *International Journal for Numerical Methods in Engineering* 1992; **35**:133–143.

